UNIVERSITY OF CALIFORNIA, SAN DIEGO

Systems Biology Approaches to Discerning Striated Muscle Pathologies

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in

Bioinformatics and Systems Biology

by

Kavitha Mukund

Committee in charge:

Professor Shankar Subramaniam, Chair
Professor Richard L. Lieber, Co-Chair
Professor Vineet Bafna
Professor Shyni Varghese
Professor Sheng Zhong

2016

The Dissertation of Kavitha Mukund is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Co-Chair

_____

Chair

University of California, San Diego

2016

DEDICATION

To everyone who has inspired and encouraged me;

Especially to Atti.

.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ACh	Acetylcholine
nAChR	Acetylcholine receptors
AQM	Acute quadriplegic myopathy
ATP	Adenosine triphosphate
ALS	Amyotrophic lateral sclerosis
BMD	Becker muscular dystrophy
BH	Benjamini Hochberg correction
BoNT-A	Botulinum neurotoxin A
CP	Cerebral palsy
CFS	Chronic fatigue syndrome
DM	Dermatomyositis
DGA	Differential gene activity
DHPR	Dihydropyridine receptors
DCM	Dilated cardiomyopathy
DGIdb	Drug-gene interaction database
DMD	Duchenne muscular dystrophy
EDMD	Emery dreifuss muscular dystrophy
ECC	Excitation contraction coupling
eQTL	Expression quantitative trait loci
ECM	Extracellular matrix

FFA	Free fatty acids
FMA	Functional module activity
GCRMA	Gene chip robust multi array average
GEO	Gene expression omnibus
GTeX	Genotype-Tissue Expression
HIBM	Hereditary inclusion body myositis
IDCM	Idiopathic dilated cardiomyopathy
IBM	Inclusion body myopathies
IGF	Insulin-like growth factors
ICM	Ischemic dilated cardiomyopathy
KLF	Kruppel-like factors
LGMD	Limb girdle muscular dystrophy
MeSH	Medical subject headings
MT	Metallothioneins
MELAS	Mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes
MRF	Myogenic regulatory factors
MRF	Myogenic regulatory factors
NMJ	Neuromuscular junction
OMIM	Online mendelian inheritance in Man
PM	Polymyositis
PEO	Progressive external opthalmoplegia

PMA	Protein module activity
PPIN	Protein-protein interaction network
ROS	Reactive oxygen species
qPCR	Real time quantitative polymerase chain reaction
RYR	Ryanodine receptors
SR	Sarcoplasmic reticulum
SMC	Signed module centrality
SLC	Solute carriers
SP	Spastic paraplegia
SP	Specificity proteins
SD	Standard deviation
TA	Tibialis anterior
TOM	Topological overlap measure
TF	Transcription factors
TCA	Tricarboxylic acid
WGCNA	Weighted gene co-expression network analysis
Y2H	Yeast 2 hybrid

LIST OF SUPPLEMENTARY FILES

Zip of supplementary tables

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to acknowledge all co-authors of papers published, submitted and under preparation for submission.

The content of chapter 2 is a modified presentation of the manuscript published in Muscle and Nerve, titled “Systems analysis of transcriptional data provides insights into muscle's biological response to botulinum toxin” by Mukund K, Mathewson M, Minamoto V, Ward SR, Subramaniam S, Lieber RL. The dissertation author was the primary author for this material.

The content of chapter 3 is a modified presentation of the manuscript submitted for publication to BMC medical genomics, titled “Botulinum Neurotoxin-A Effects on Skeletal Muscle from Gene Co-Expression Networks: A Time Course Study” by Mukund K, Ward SR, Lieber RL, Subramaniam S. The dissertation author was the primary author for this material.

The content of chapter 4 is a modified presentation of the manuscript published in BMC research notes, titled “Dysregulated mechanisms underlying Duchenne muscular dystrophy from co-expression network preservation analysis” by Mukund K, Subramaniam S. The dissertation author was the primary author for this material.

The content of chapter 5 is a modified presentation of material being prepared for submission currently titled “Functional relationships amongst diseases affecting skeletal muscle: a network theoretic approach” by Mukund K, Subramaniam S. The dissertation author is the primary author for this material.

The content of chapter 6 is a modified presentation of material being prepared for submission, currently titled “Mechanisms underlying ischemic and idiopathic dilated cardiomyopathy utilizing signed differential co-expression network.” by Mukund K, Subramaniam S. The dissertation author is the primary author for this material.

None of this would have been possible without the constant guidance and support of my advisor Dr. Shankar Subramaniam. It has been an amazing experience to work in a lab with a brilliant PI who was patient, encouraging, and open to new ideas. Dr. Subramaniam has been integral to my growth as a human being, and a scientist. Thank you.

I would like to thank my co-advisor Dr. Richard Lieber, for his valuable inputs and guidance through the course of my graduate studies. I would also like to thank all my committee members- Drs. Bafna, Varghese and Zhong for their invaluable inputs during each phase of my doctoral study.

I would like to all the current and past members of the Subramaniam lab for their help and meaningful discussions over the course of my graduate research. I would also like to thank Carol Kling, assistant to Dr. Subramaniam, for all her help during the course of my PhD.

Finally, the gratitude I feel for my family and friends in being my support system cannot be expressed in a few words. I am ever grateful to my parents- Suma and Mukund, and my brother Bharath, for being so wonderful and supportive in every aspect of my life. I can’t thank my father enough for instilling in me the love and passion for

science, and my mother for never doubting the abilities of her daughter, allowing her to dream and encouraging her to follow them.

A special thank you is due to my son Srijan, for being such a caring, loving young man, for teaching me to be a good mother and a better person. Lastly and most importantly, I thank my husband Atti, for believing in me, at times when I didn't believe in myself. For being my cheer leader, my best friend and worst critic. But for you, this journey would not have been possible! Thank you.

VITA

2002-2006	Visvesvaraya Technological University, India Bachelor of Engineering, Electrical Engineering
2008-2009	Arizona State University, Tempe Master of Science, Biomedical Informatics
2011- 2016	University of California, San Diego Doctor of Philosophy, Bioinformatics and Systems Biology *Graduate Researcher, Subramaniam Lab*

TEACHING EXPERIENCE

University of California, San Diego

- Teaching Assistant, Department of Bioengineering
  Systems Biology and Bioengineering I: Biological Components. Dr. Kun Zhang. Fall 2012

- Teaching Assistant, Department of Chemistry and Biochemistry
  Applied Bioinformatics. Dr. Alexander Hoffmann. Winter 2013.

PUBLICATIONS

Peer-Reviewed Research Articles

[Under preparation] **Mukund K**, Subramaniam S. Mechanisms underlying ischemic and idiopathic dilated cardiomyopathy utilizing signed differential co-expression network. 2016.

[Under preparation] **Mukund K**, Subramaniam S. Functional relationships amongst diseases affecting skeletal muscle - a network theoretic approach. 2016.

[Under preparation] Chapman MA, **Mukund K**, Subramaniam S, Brenner D, Lieber RL. Three distinct skeletal muscle cell types produce collagen for different niches during fibrosis. 2016.

[Submitted] **Mukund K**, Ward SR, Lieber RL, Subramaniam S. Botulinum Neurotoxin-A Effects on Skeletal Muscle from Gene Co-Expression Networks: A Time Course Study. 2016

Mukund K, Subramaniam S. Dysregulated mechanisms underlying Duchenne muscular dystrophy from co-expression network preservation analysis. BMC research notes. 2015 May 3;8(1):182.

Mukund K, Mathewson M, Minamoto V, Ward SR, Subramaniam S, Lieber RL. Systems analysis of transcriptional data provides insights into muscle's biological response to botulinum toxin. Muscle & nerve. 2014 Nov 1; 50(5):744-58.

Cohen T, Whitfield GK, Schvaneveldt RW, **Mukund K**, Rindflesch T. EpiphaNet: an interactive tool to support biomedical discoveries. Journal of biomedical discovery and collaboration. 2010;5:21.

Conference Proceedings

Mukund K, Ward SR, Lieber RL, Subramaniam S. Systems approaches to uncovering mechanisms underlying skeletal muscle pathophysiology. March 2014. Conference talk presented at the semi-annual CMM Symposium, La Jolla, California.

Mukund K, Subramaniam S. A differential co-expression networks approach to uncovering mechanisms underlying skeletal muscle pathophysiology. May 2014. Poster presented at the RECOMB 2014, Pittsburgh, PA.

ABSTRACT OF THE DISSERTATION

Systems Biology Approaches to Discerning Striated Muscle Pathologies

by

Kavitha Mukund

Doctor of Philosophy in Bioinformatics and Systems Biology

University of California, San Diego, 2016

Professor Shankar Subramaniam, Chair

Professor Richard L. Lieber, Co-Chair

The human muscular system represents nearly 75% of the body mass and encompasses two major muscle forms- striated and smooth. Striated muscle, composed broadly of myofibers, accompanying membrane systems, cytoskeletal networks together with the metabolic and regulatory machinery, have revealed complexities in composition, structure and function. A disruption to any component within this complex

system of interactions lead to disorders of the muscle, typically characterized by muscle fiber loss, reduced motor output and in some cases death. Advent of high-throughput technologies coupled with elegant approaches to deciphering data using bioinformatics and systems biology, are providing new venues for detailed exploration of mammalian muscle.

This dissertation describes the use of publicly available high-throughput data, in conjunction with co-expression network methodologies developed for a comprehensive, interpretable systems-level perspective on mechanisms underlying associated muscle pathologies. This study begins with the exploration of the temporal transcriptional response of skeletal muscle to Botulinum Neurotoxin-A (Botox ®) over a 1-year period, in the framework of muscle physiology. Next, utilizing co-expression network analysis, putative markers associated with recovery of muscle trophicity are identified, furthermore providing an unbiased validation of the response documented earlier. These studies represent the first attempt at categorically assessing the whole-transcriptomic changes associated with BoNT-A treatment in muscle.

The latter half of this research focuses on discerning patho-mechanisms of human diseases affecting muscle. Particularly, co-expression network statistics are leveraged to identify dysregulated pathways and biomarkers of disease progression, underlying duchenne muscular dystrophy. Next, a quantitative framework integrating transcriptional, protein interaction, and drug-target data is developed to extract functional similarities and mechanisms amongst 20 diseases affecting the muscle. Lastly, an approach to differential co-expression analysis using signed and weighted co-

expression networks is described. This approach is subsequently utilized to assess and identify differential mechanisms underlying ischemic and idiopathic dilated cardiomyopathy. The analysis and results from the aforementioned studies have enabled a deeper understanding of the complex interactions underlying muscle pathologies; providing opportunities for drug development and personalized medicine.

CHAPTER 1 - Introduction

1.1 Introduction

The human muscular system represents nearly 75% of the body mass and encompasses two major muscle forms- striated and smooth. Striated muscle, composed broadly of myofibers, accompanying membrane systems, cytoskeletal networks together with the metabolic and regulatory machinery, have revealed complexities in composition, structure and function. Precisely coordinated activity of each of these components is essential for normal functioning with factors intrinsic (such as genetic, epigenetic, and developmental) and environmental (such as hormonal, immune) for shaping the destiny of muscular health and associated motor activity. A disruption to any component within this complex system of interactions lead to disorders of the muscle, typically characterized by muscle fiber loss, reduced motor output and in some cases death [1].

The striated muscular system is composed of two major muscle types- skeletal and cardiac. While the cardiac (heart) muscle functionally represents a set of self-stimulating, non-fatiguing muscle cells with an intermediate energy requirement, skeletal muscle tissue represents a set of innervated muscle cells that exhibit fatigue with high energy requirement. Prior to exploring the extant research in the –omics era, for deciphering the complex system of interactions in their pathophysiology, we first review the structure and function of muscle.

1.2 Structure and function of muscle

A cursory glance of the complex interactions in striated muscle indicate the degree to which striated muscle is designed to accomplish the task of generating contraction, force and movement. This muscle type is characterized by the presence of repeating functional units called sarcomeres. The sarcomeres visually manifest as a series of bands along the muscle fibers, leading to the striated appearance under a microscope [2]. Since we focus mainly on the pathophysiology of skeletal muscle in the course of this dissertation, we provide a basic overview of its structure and function in the following sections.

Skeletal muscle, a type of striated muscle produces force by interaction of two primary proteins within the sarcomere, actin and myosin (Figure 1.1). Sarcomeres are joined end to end, in a series, to form myofibrils; with tight bundles of myofibrils forming the multinucleated and long myofibers. The plasma membrane (sarcolemma) of the myofibers is surrounded by satellite cells and the basal lamina. The many nuclei of the muscle fiber are located at the periphery of the cell, just under the sarcolemma. The regenerative capacity of muscle is attributed to this normally quiescent population of intrinsic stem (or satellite) cells [3]. The sarcolemma projects long, finger-like processes called T- tubules that ring around every sarcomere and interact with the sarcoplasmic reticulum (SR, which serves as Ca^{2+} store within muscle). About 80% of the sarcoplasm is occupied by myofibrils surrounded by mitochondria. Bundles of muscle fibers form the fascicles; with bundles of fascicles forming the tissue [2].

Each of the above mentioned structures within the muscle tissue are encapsulated by the extracellular matrix (ECM) [4]. The epimysium is a dense connective tissue layer encapsulating the entire muscle. The perimysium derives from the epimysium and surrounds the fascicles. The endomysium, a delicate layer of reticular fibers surrounds each myofiber. An extensive network of capillaries and nerves, flexible enough to adjust to contraction-relaxation changes, utilize the connective tissue to reach individual myofibers [5]. Varying composition of the major protein isoforms (actin and myosin) within skeletal muscle gives rise to slow and fast muscle fibers that cater to varying metabolic and contractile needs [6].

Functionally, the muscle is a specialized tissue for contraction, where energy from the hydrolysis of adenosine triphosphate (ATP) is transformed into mechanical energy [7]. Upon arrival of an electric impulse a sequence of events are set in motion at the neuromuscular junction (NMJ), beginning with the opening of voltage-gated calcium channels, releasing Ca^{2+} ions into the presynaptic cytosol. This triggers the release of Acetylcholine (ACh), a Ca^{2+} regulated neurotransmitter from the synaptic vesicles into the postsynaptic cleft at the NMJ. The depolarizing effect upon binding of ACh at the NMJ, spreads along the post-synaptic membrane and activates voltage-sensitive sodium channels in muscle. This in turn triggers the transmission of neural excitation past the NMJ terminating at a specialized set of voltage sensors within the muscle's T-tubules, called the dihydropyridine receptors (DHPR) [8]. The DHPR are mechanically coupled to ryanodine receptors (RYR) in the SR, which release Ca^{2+} from the SR upon arrival of the impulse. Binding of the released Ca^{2+} to troponin-C within

the sarcomere brings about a conformational change in the troponin-tropomyosin complex. This results in the exposure of myosin binding sites on the actin filaments [9]. Myosin heads bind and crawl along the length of the actin filament bringing about hydrolysis of ATP and contraction of the sarcomere and subsequently of the entire muscle [10].

1.3 Systems biology

Physiological and biochemical studies have contributed enormously to the understanding of mechanisms underlying muscle function and pathology [11,12]. Additionally, epidemiological studies have gleaned further insight into individual muscle pathologies leading to better strategies for clinical diagnosis and therapy e.g. [13,14]. High-throughput studies of the current era have augmented our understanding of muscle pathophysiology with goals for personalized medicine e.g. [15,16]. Recent advancements in high throughput technologies, and the availability of public repositories hosting the high throughput genomic, proteomic, metabolomic, epigenetic and related data such as NCBI [17] and Ensemble [18], have led to improved understanding of mechanisms and function underlying various muscle pathologies.

Systems biology is a field of research that is aimed towards integrating various forms of biological data for gaining an overall picture of the complex dynamics underlying a biological system (Figure 1.2). In contrast to the traditional reductionist approach to studying only an aspect of a biological system, systems biology aims to gain a more holistic understanding of the system by integrating the multi-scale

interactions that occur between molecular-, cellular-, organ- level entities and environmental factors [19,20]. The past decade has been extremely fruitful in the development of new technologies to analyze biological systems at various levels of magnification. Ever increasing magnitude of data has necessitated extensive use of extant mathematical and computational modelling techniques, in addition to development of new techniques providing enhanced biological insights [21].

A computational approach extensively adopted by systems biology utilizes concepts from graph theory and data based network models to model biological interaction. This involves the idea of visualizing the biological system as a highly interconnected "network" comprising of several interacting groups of components (e.g. genes, protein, metabolites, drugs, diseases) e.g [22,23]. Systems biology is being increasingly used in skeletal and cardiac muscle research to aid in biomarker discovery and reveal mechanistic details on its pathophysiology [24–30].

During the course of this thesis, we rely heavily on concepts from co-expression network theory to decipher muscle pathophysiology. In this following section, we present an introduction to the basic concepts of network theory and in extension to co-expression network theory.

1.4 Network theory

Network theory offers a quantifiable approach to understanding complex, dynamic, biological systems. Networks are graphically represented as a collection of nodes and edges [31]. Depending on the context of study, nodes define various

molecular/biological entities (e.g genes, proteins, metabolites, drugs, diseases) linked by edges that describe the relationship between nodes. For example, Sáez et al [32] present an interesting use of network theory as a framework to extract information from muscle biopsy images in diagnosis of muscular dystrophies and neurogenic atrophies where each myofiber within an image serves as nodes and fiber contacts as links.

Each of the network elements – nodes and edges have associated properties. Directionality, sign and weight are attributes most commonly associated with network edges. Networks that exhibit edge directionality (node A influences node B) are called directed networks, while undirected networks exhibit no edge directionality (node A is associated with node B and vice versa). A network can either be unweighted – where edges carry a weight of 1 or 0 (1 if link present 0 otherwise) or weighted- where the edges are associated with a strength. Weighted networks can either be signed (edge weights in the interval [-1, 1]) or unsigned (edge weight in the interval [0, 1]). Nodes, likewise are associated with degree- the most elementary characteristic, defined as the sum of weights of the edges incident upon a node (entering (in-Degree) + exiting (out-Degree)); nodes with high degree (hubs) often have been found to potential biomarkers. Degree centrality takes discrete values for unweighted networks and continuous values for weighted networks.

Most biological systems are described to be scale-free, that is there exist only few highly connected hubs within each network [33,34]. Likewise, high node/local clustering co-efficient indicates the ability of nodes to cluster together suggesting a strong relationship between groups of nodes. Modularity detection algorithms have been

extensively used to identify modules- or subnetworks of highly connected nodes that are in close graphical (topological) proximity, suggesting functional association. Modularity has been suggested as a fundamental property of many complex systems [35].

Network topology combined with data from epigenetics, proteomics, functional enrichment, and/or clinical studies have revealed the dynamics underlying multiple disease pathologies. For example, Goh et al [36] extracted a “diseaseome”, a network of disease nodes connected by links if disease genes were shared between them (from OMIM). Wu et al. [37] proposed a computational network CIPHER- that integrated human protein–protein interactions (PPIN), disease phenotype similarities, and known gene–phenotype associations to capture the complex relationships between diseases; more recently, Zhou et al [38] used a large-scale biomedical literature database to construct a symptom-based human disease network to investigate the connection between clinical manifestations of diseases and their underlying molecular interactions. Suthram et al [39], extended co-expression networks further by integrating protein interaction to provide a quantitative framework for analysis of disease-related mRNA expression across a wide array of unrelated diseases.

Co-expression networks are a specialized class of undirected, mostly weighted networks where the edge weights are calculated as the co-expression (such as correlation, cosine similarity) between the nodes represented in the network. In the following section we provide a brief introduction to an approach of co-expression networks analysis, which is used extensively as part of this dissertation.

1.4.1 Co-expression networks

Co-expression network theory using transcriptional data is based on the premise that co-regulated genes are strongly co-expressed. Several approaches to constructing and analyzing co-expression networks have been proposed. One such heavily utilized approach is called weighted gene co-expression network analysis (WGCNA) [40], which we utilize in the course of this thesis. WGCNA has been repeatedly shown to extract meaningful gene associations, revealing interesting aspects of pathologies using transcriptional and other forms of high throughput data e.g [41–43]. Figure 1.3 summarizes the major steps involved in analysis using WGCNA. Briefly, in WGCNA, co-expression network generation is based on the scale free network topology - that is, WGCNA constructs the network for several thresholds (soft thresholds, β) and selects the threshold which leads to a network with scale-free topology [44]. The edge weight in this network reflects how significant the co-expression relationship is between two nodes and is calculated as the correlation raised to soft power β. Gene modules are ascertained using agglomerative (hierarchical) clustering. In contrast to traditional distance measures, WGCNA utilizes a measure of network similarity called topological overlap measure [35]. Genes that share a large number of neighbors tend to be more strongly connected, with maximal similarity between two genes equaling 1 and a minimal similarity of 0. Modules are defined as branches of the hierarchical tree determined using the dynamic branch cutting approach [45]. Network parameters are computed similar to other weighted networks.

1.5 Introduction to the dissertation

Striated muscle is a versatile tissue, with science devoting many several decades to its understanding. Advent of genomic technologies have allowed for a more comprehensive evaluation of the pathophysiology of the muscle under various perturbed states. Elucidating functional mechanisms underlying various muscle pathologies utilizing genomic data while employing techniques from bioinformatics and systems biology is the focus of this research.

Chapter 2 is a modified presentation of the manuscript published in Muscle and Nerve, explores the transcriptional response of skeletal muscle to Botulinum Neurotoxin-A (BoNT-A, Botox ®), in the framework of muscle physiology [46]. This work provided the first document model for global transcriptional changes occurring upon neurotransmitter blockade with BoNT-A in mammalian skeletal muscle. Overall, gene expression changes correlated with the clinically accepted BoNT-A time course and suggested that the direct action of BoNT-A in skeletal muscle is relatively rapid.

Chapter 3, is a modified presentation of the manuscript submitted for publication to BMC Medical Genomics, provided a co-expression network approach to assessing changes associated with cross-sectional temporal data obtained in Chapter 1. Grouping of network modules revealed a hierarchical functional response to changes occurring during the course of BoNT-A-induced paralysis with an early metabolic response and later response affecting ECM. Two highly ranked genes Dclk1 and Ostalpha were identified to be potentially associated in the recovery of muscle from BoNT-A induced

atrophy. Additionally, our results provide an unbiased, data driven validation of the response documented in our previous work (Chapter 2).

Chapter 4 is a modified presentation of the manuscript published in BMC research notes, which utilizes network statistics associated with co-expression networks to identify and rank differentially co-expressed gene modules associated with duchenne muscular dystrophy [47] . Duchenne muscular dystrophy (DMD) is an X-linked recessive disorder with its primary insult on skeletal muscle. Severe muscle wasting, chronic inflammation and fibrosis characterize dystrophic muscle. This work illustrated the use of network “preservation” statistics in identifying dysregulated pathways underlying DMD. Out analysis identified highly specific interactions between known markers of disease to be differential, in addition to identification of putative markers likely associated with the progression of DMD.

Chapter 5 is a modified presentation of the manuscript being prepared for submission, focused on utilizing a systems approach to elucidation of muscle pathologies. Majority of the genomic/genetic studies in skeletal muscle research have focused extensively on identifying biomarkers associated with individual diseases. Our work provided the first attempt at exploring the interactions between 20 diseases affecting human skeletal muscle utilizing an integrated network theoretic approach incorporating publicly available transcriptomic, protein interaction and drug target data. In contrast to current studies, this form of analysis allowed for a synergistic identification of functional similarities and mechanisms likely shared among muscle diseases, which may or may not share clinical similarities.

Chapter 6 is a modified presentation of the manuscript being prepared for submission, focused on defining and utilizing a signed differential co-expression networks approach to understanding the mechanisms underlying ischemic and idiopathic dilated cardiomyopathy. Our analysis confirmed the functional mechanisms underlying the pathogenesis of heart failure. Our results also suggested a differential regulation of the targets of SP/KLF family of transcriptional factors between idiopathic and ischemic dilated cardiomyopathy. This work provides the first attempt at utilizing signed and weighted differential co-expression network approach to discerning disease pathology.

Chapter 7 summarizes the major results and their significance in the context of muscle systems biology.

1.6 Summary

Epidemiological, clinical, physiological and biochemical studies of the past several decades have provided invaluable insights into the working of mammalian striated muscle. The advent of high-throughput technologies coupled with elegant approaches for deciphering data using bioinformatics and systems approach are providing new windows of opportunities for detailed exploration of muscle. In particular, use of network theory is providing unique opportunities for integrating different sources of data such as genetic, epigenetic, proteomic and drug. This form of analysis has lent itself to a better understanding of the complex system of interactions

underlying muscle pathologies; providing opportunities for accelerated drug development and personalized treatments to improve patient outcome.

1.7 Acknowledgements

The author would like to acknowledge Encyclopedia Britannica, Inc., as the source for Figure 1.1 utilized in this chapter.

1.8 Figures

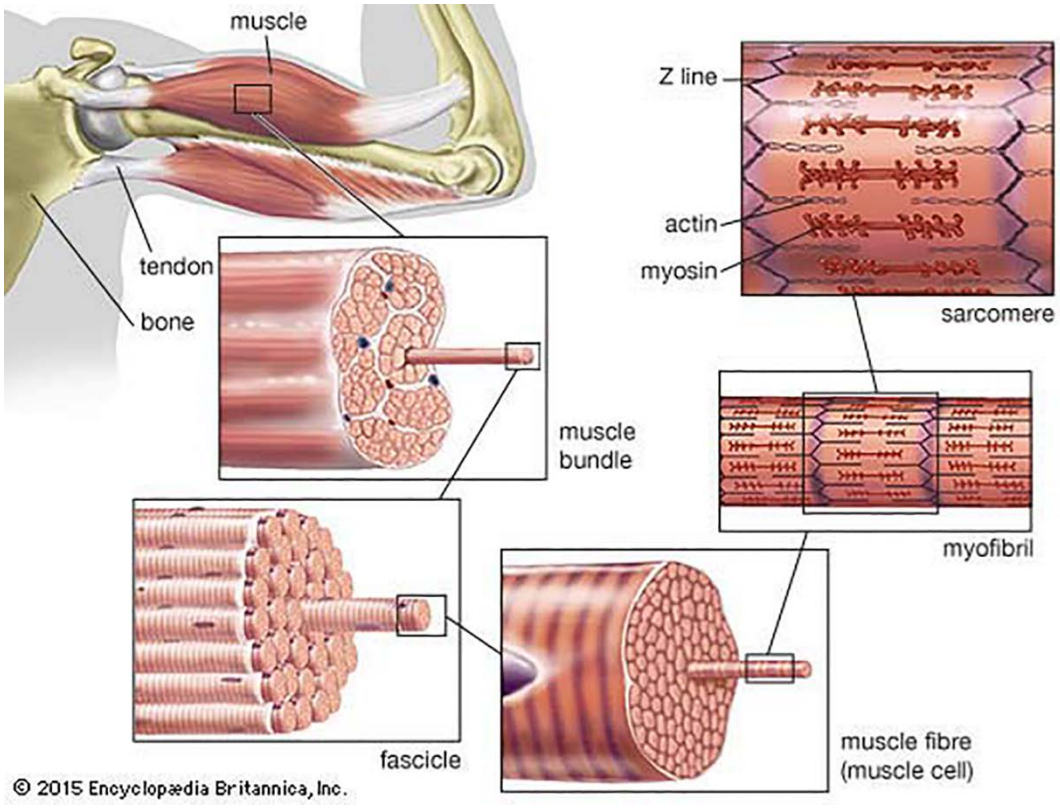

Figure 1.1: Skeletal muscle physiology
This image shows the structural aspects of the skeletal muscle at various levels of magnification as discussed in section 1.1 of the text. (© Encyclopaedia Britannica, Inc)

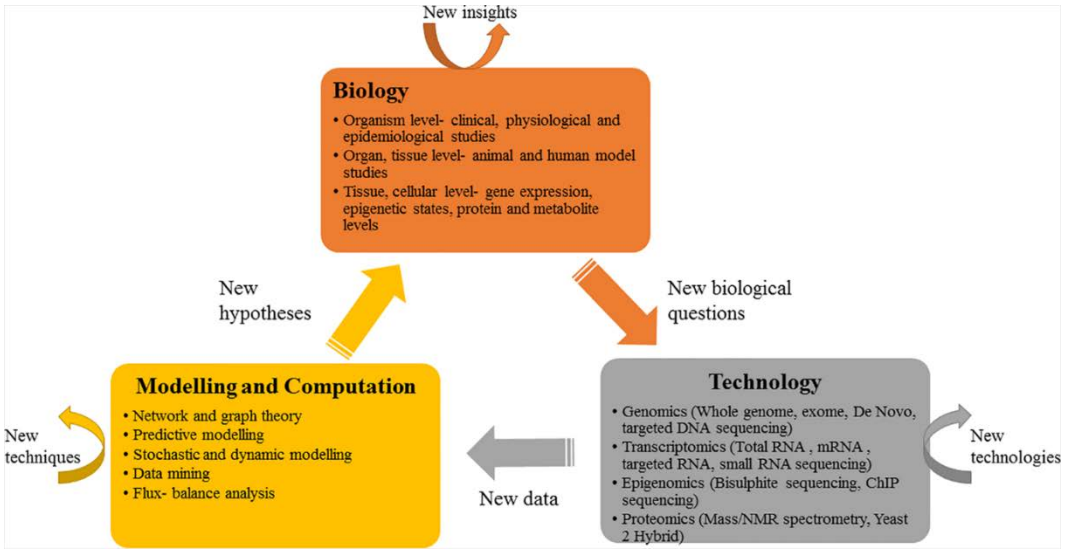

Figure 1.2: An overview of systems biology
This figure captures the essence of the different aspects underlying current biological research in the context of increased availability of new high throughput technologies and subsequently data and the use of bioinformatics and systems approaches to gaining new insights.

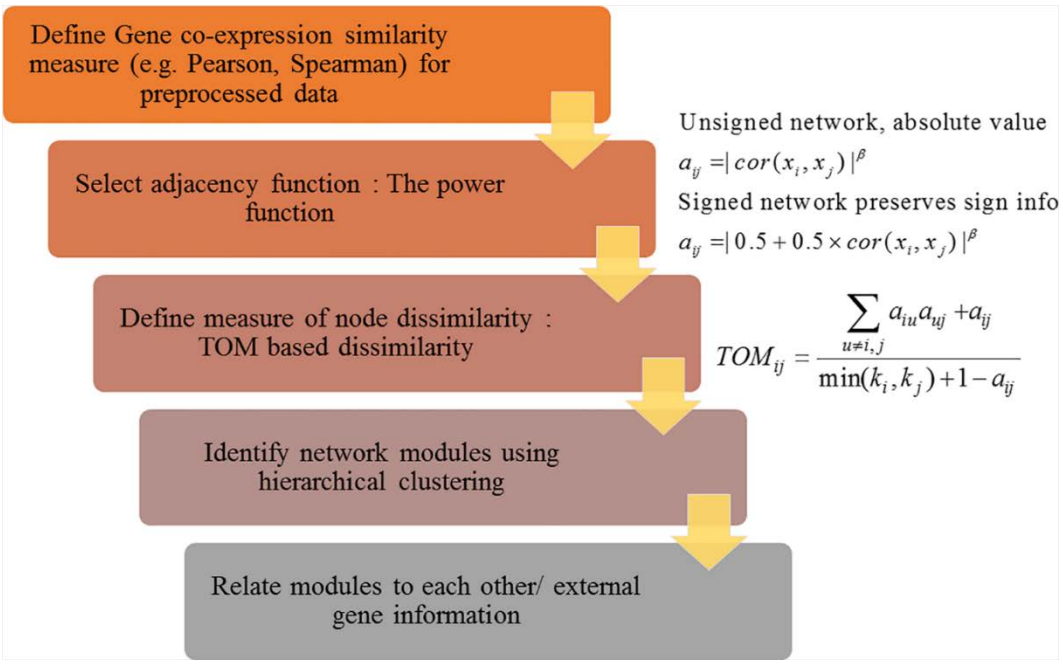

Figure 1.3: Workflow of weighted gene co-expression network analysis (WGCNA) The main steps involved in generating a co-expression network using WGCNA are presented here.

1.9 References

1. Bourne G. The Structure and Function of Muscle V4: Pharmacology and Disease. Elsevier; 2014.

2. Kierszenbaum AL, Tres L. Histology and cell biology: an introduction to pathology. Elsevier Health Sciences; 2015.

3. Mauro A. Satellite cell of skeletal muscle fibers. J. Biophys. Biochem. Cytol. 1961;9:493–5.

4. Gillies AR, Lieber RL. Structure and function of the skeletal muscle extracellular matrix. Muscle Nerve. 2011;44:318–31.

5. Fuxe K, SEDVALL G. The distribution of adrenergic nerve fibres to the blood vessels in skeletal muscle. Acta Physiol. Scand. 1965;64:75–86.

6. Schiaffino S, Reggiani C. Fiber types in mammalian skeletal muscles. Physiol. Rev. 2011;91:1447–531.

7. Ebashi S. Excitation-contraction coupling. Annu. Rev. Physiol. 1976;38:293–313.

8. Rios E, Brum G. Involvement of dihydropyridine receptors in excitation–contraction coupling in skeletal muscle. 1987;

9. Gordon A, Homsher E, Regnier M. Regulation of contraction in striated muscle. Physiol. Rev. 2000;80:853–924.

10. Rayment I, Holden HM, Whittaker M, Yohn CB, Lorenz M, Holmes KC, Milligan RA. Structure of the actin-myosin complex and its implications for muscle contraction. Science. 1993;261:58–65.

11. Jones DA, Round JM. Skeletal muscle in health and disease: a textbook of muscle physiology. Manchester University Press; 1990.

12. Rędowicz MJ, Moraczewska J. Insight into muscle physiology through understanding mechanisms of muscle pathology. J. Muscle Res. Cell Motil. 2015;1–3.

13. Codd M, Sugrue D, Gersh B, Melton L. Epidemiology of idiopathic dilated and hypertrophic cardiomyopathy. A population-based study in Olmsted County, Minnesota, 1975-1984. Circulation. 1989;80:564–72.

14. Mah JK, Korngut L, Dykeman J, Day L, Pringsheim T, Jette N. A systematic review and meta-analysis on the epidemiology of Duchenne and Becker muscular dystrophy. Neuromuscul. Disord. 2014;24:482–91.

15. Lim LE, Rando TA. Technology Insight: therapy for Duchenne muscular dystrophy—an opportunity for personalized medicine? Nat. Clin. Pract. Neurol. 2008;4:149–58.

16. Ginn SL, Alexander IE, Edelstein ML, Abedi MR, Wixon J. Gene therapy clinical trials worldwide to 2012–an update. J. Gene Med. 2013;15:65–77.

17. Coordinators NR, Acland A, Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Bryant SH, Canese K, Church DM. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2014;42:D7.

18. Cook CE, Bergman MT, Finn RD, Cochrane G, Birney E, Apweiler R. The European Bioinformatics Institute in 2016: Data growth and integration. Nucleic Acids Res. 2016;44:D20–6.

19. Kitano H. Systems biology: a brief overview. Science. 2002;295:1662–4.

20. Hood L, Heath JR, Phelps ME, Lin B. Systems biology and new technologies enable predictive and preventative medicine. Science. 2004;306:640–3.

21. Kitano H. Computational systems biology. Nature. 2002;420:206–10.

22. Hu G, Agarwal P. Human disease-drug network based on genomic expression profiles. PloS One. 2009;4:e6536.

23. Vidal M, Cusick ME, Barabasi A-L. Interactome networks and human disease. Cell. 2011;144:986–98.

24. Chien KR. Genomic circuits and the integrative biology of cardiac diseases. Nature. 2000;407:227–32.

25. Laaksonen R, Katajamaa M, Paiva H, Sysi-Aho M, Saarinen L, Junni P, Lütjohann D, Smet J, Van Coster R, Seppänen-Laakso T, Lehtimäki T. A systems biology strategy reveals biological pathways and plasma biomarker candidates for potentially toxic statin-induced changes in muscle. PloS One. 2006;1:e97.

26. Hudson NJ, Reverter A, Wang Y, Greenwood PL, Dalrymple BP. Inferring the transcriptional landscape of bovine skeletal muscle by integrating co-expression networks. PloS One. 2009;4:e7249.

27. Dewey FE, Perez MV, Wheeler MT, Watt C, Spin J, Langfelder P, Horvath S, Hannenhalli S, Cappola TP, Ashley EA. Gene coexpression network topology of cardiac development, hypertrophy, and failure. Circ. Cardiovasc. Genet. 2011;4:26–35.

28. Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, Curnow SJ, Rodriguez DA, Cascante M, O'Neill L, Egginton S. A systems biology approach

identifies molecular networks defining skeletal muscle abnormalities in chronic obstructive pulmonary disease. PLoS Comput Biol. 2011; 7(9): e1002129.

29. Azuaje FJ, Dewey FE, Brutsaert DL, Devaux Y, Ashley EA, Wagner DR. Systems-based approaches to cardiovascular biomarker discovery. Circ. Cardiovasc. Genet. 2012;5:360–7.

30. Blandin G, Marchand S, Charton K, Danièle N, Gicquel E, Boucheteil J-B, Bentaib A, Barrault L, Stockholm D, Bartoli M, Richard I. A human skeletal muscle interactome centered on proteins involved in muscular dystrophies: LGMD interactome. Skelet Muscle. 2013;3(1):1.

31. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 2004;5:101–13.

32. Sáez A, Rivas E, Montero-Sánchez A, Paradas C, Acha B, Pascual A, Serrano C, Escudero LM. Quantifiable diagnosis of muscular dystrophies and neurogenic atrophies through network analysis. BMC Med. 2013;11(1):1.

33. Barabási A-L, Albert R. Emergence of scaling in random networks. science. 1999;286:509–12.

34. Barabási A-L. Scale-free networks: a decade and beyond. science. 2009;325:412.

35. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási A-L. Hierarchical organization of modularity in metabolic networks. science. 2002;297:1551–5.

36. Goh K-I, Cusick ME, Valle D, Childs B, Vidal M, Barabási A-L. The human disease network. Proc. Natl. Acad. Sci. 2007;104:8685–90.

37. Wu X, Jiang R, Zhang MQ, Li S. Network-based global inference of human disease genes. Mol. Syst. Biol. 2008;4:189.

38. Zhou X, Menche J, Barabási A-L, Sharma A. Human symptoms–disease network. Nat Commun [Internet]. 2014;5. Available from: http://dx.doi.org/10.1038/ncomms5212

39. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. PLoS Comput. Biol. 2010;6:e1000662.

40. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 2005;4:1128.

41. Miller JA, Horvath S, Geschwind DH. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc. Natl. Acad. Sci. 2010;107:12698–703.

42. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011;474:380–4.

43. Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, De Nardo D, Gohel TD, Emde M, Schmidleithner L, Ganesan H. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. Immunity. 2014;40:274–88.

44. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

45. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics. 2008;24:719–20.

46. Mukund K, Mathewson M, Minamoto V, Ward SR, Subramaniam S, Lieber RL. Systems Analysis of Transcriptional Data Provides Insights Into Muscle's Biological Response to Botulinum Toxin. Muscle Nerve. 2014;50:744–58.

47. Mukund K, Subramaniam S. Dysregulated mechanisms underlying Duchenne muscular dystrophy from co-expression network preservation analysis. BMC Res. Notes. 2015;8:182.

CHAPTER 2- Systems analysis of transcriptional data provides insights into muscle's biological response to botulinum toxin

2.1 Abstract

Introduction: This study provides global transcriptomic profiling and analysis of botulinum toxin A (BoNT-A) treated muscle over a one-year period. Methods: Microarray analysis was performed on rat tibialis anterior muscles from 4 groups (n= 4/group) at 1, 4, 12, and 52 weeks after BoNT-A injection compared with saline-injected rats at 12 weeks. Results: Dramatic transcriptional adaptation occurred at 1 week with a paradoxical increase in expression of slow and immature isoforms, activation of genes in competing pathways of repair and atrophy, impaired mitochondrial biogenesis, and increased metal ion imbalance. Adaptations of the basal lamina and fibrillar extracellular matrix (ECM) occurred by 4 weeks. The muscle transcriptome returned to its unperturbed state 12 weeks after injection. Conclusion: Acute transcriptional adaptations resemble denervated muscle with some subtle differences but resolved more quickly compared to denervation. Overall gene expression, across time, correlates with the generally accepted BoNT-A time course and suggests that the direct action of BoNT-A in skeletal muscle is relatively rapid.

2.2 Introduction

Skeletal muscle contraction is controlled by impulses received from the central nervous system via the neuromuscular junction (NMJ). In cases where skeletal muscle function is impaired due to altered activity of nerve impulses, for instance, in movement disorders such as cerebral palsy (CP), it can be advantageous to suppress muscle contraction by reducing NMJ activity. Signal reduction can be achieved by physically

decoupling the muscle and nerve by selective dorsal rhizotomy [1] or through the use of chemical agents such as neurotoxins [2]. One such neurotoxin in common clinical use is Botulinum toxin A (BoNT-A), which has applications ranging from decreasing spasticity, tics, and tremors, to managing pain and controlling glandular secretions [3].

BoNT-A is one of 7 serotypes produced by *Clostridium botulinum* that functions to reversibly paralyze muscle by affecting the NMJ. BoNT-A reduces presynaptic ACh release by specifically cleaving a SNARE protein, *SNAP25*, required for its exocytosis. BoNT-A induced neuromuscular block causes physical and physiological changes to the NMJ and skeletal muscle fiber [4, 5]. Previously published experimental studies have reported that muscle reinnervation via neuronal sprouting begins immediately upon injection, with control slowly reverting back to the parent terminal over time [6, 7]. It has also been observed that, during this time, skeletal muscle is characterized by reduced fiber size, paresis and atrophy [8] until it gradually regains functionality. When used as a therapeutic agent in disorders such as CP, BoNT-A is administered by intramuscular injection repeatedly over extended periods of time. While clinical experience demonstrates that the injection effects last between 3-6 months [9], a cohesive temporal picture of there is not yet a clear understanding of underlying muscle functional and transcriptional regulation. Although the effects of BoNT-A treatments in skeletal muscle have been studied extensively experimentally [4, 7, 10], to the best of our knowledge, only a single genomic study was published that focused on certain genes associated with BoNT-A action in skeletal muscle [11]. In contrast, we now report a complete systems analysis of the BoNT-A treated skeletal muscle transcriptome over a

period of 1 year, with the goal of understanding the underlying biological response to BoNT-A and the relationship between transcriptional and functional changes associated with its reversible paralysis. We analyze our results in the context of “physiological families” of skeletal muscle as recently published [12, 13]. The primary goals for this study were therefore two-fold: 1) to create a documented model for global transcriptional changes that occur with neurotransmitter blockade using BoNT-A in skeletal muscle, and 2) to gain insights into the biological basis for adaptation and recovery of muscle after BoNT-A treatment.

2.3 Materials and Methods

2.3.1 Animals

All procedures were performed with the approval of the University of California, San Diego Institutional Animal Care and Use Committee. Mature male Harlan Sprague Dawley rats (age 3 months, 399 ± 3.05 g) were given a single 100 μl injection in the tibialis anterior (TA) muscle containing with either saline or saline with 6U/kg BoNT-A (BOTOX® (onabotulinum toxin A), Allergan, Irvine, CA, USA). At 1, 4, 12, and 52 weeks after injection, rats were sacrificed by intracardiac pentobarbital sodium (0.5 ml of 390 mg/mL solution) injection. Maximum isometric contraction strength was measured on all rats prior to sacrifice, as described previously [14]. After sacrifice, bilateral TA muscles were excised, weighed, and snap-frozen in isopentane cooled by liquid nitrogen (-159°C). All samples were stored at -80°C for further analysis.

2.3.2 Hydroxyproline assay

A modified version of hydroxyproline assay [15] was used to determine collagen content. Briefly, muscles were hydrolyzed at 110˚C overnight in hydrochloric acid, then methyl red was added, and samples were pH adjusted. Chloramine T and p-diaminobenzaldehyde were added sequentially to samples, which were then incubated for 30 min at 60˚C. A standard curve was determined, and samples were read at 550 nm and 558 nm.

2.3.3 RNA preparation

Samples were prepared for 5 groups (n=4/group) that include tissue from TAs of BoNT-A injected rats at 1, 4, 12, and 52 weeks after injection. Control tissue was obtained from the contralateral TA of saline-injected rats sacrificed at 12 weeks. RNA was extracted using a Trizol (Invitrogen, Carlsbad, CA, USA) and RNeasy (Qiagen, Valencia, CA, USA) method. Briefly, 30 mg of frozen tissue was mixed with 0.5 ml of Trizol and homogenized at 4°C in a Bullet Blender (Next Advance, Inc., Averill Park, NY, USA). The homogenate was mixed with 100 µl of chloroform, and samples were incubated for 2 minutes at room temperature and spun at 4°C for 15 minutes. The aqueous portion was removed and mixed with equal amounts of 70% EtOH. The solution was then washed through an RNeasy spin column, incubated for 15 minutes with RNAse-free DNAse (Qiagen, Inc., Valencia, CA, USA), washed 3 times, and eluted according to the manufacturer's instructions. Absorbance was measured at 260 nm to determine RNA concentration, and the 260/280 nm absorbance ratio was calculated to determine RNA purity. RNA was reverse-transcribed into cDNA using the

SuperScript First-Strand Synthesis System (Life Technologies, Grand Island, NY, USA).

2.3.4 Microarray Data Collection

Affymetrix microarrays (RG-230 2.0; Affymetrix, Santa Clara, CA) were used for microarray analysis of all samples. The UCSD Cancer Center Microarray Shared Resource (San Diego, CA) provided RNA processing and quality control using the GS FLX System (Roche Diagnostic Corporation, Basel, Switzerland).

2.3.5 Real time quantitative PCR

Real-time quantitative PCR (qPCR) was conducted to validate the expression of 8 genes (*Chrna1, Myl3, Sln, Myog, Aqp4, Runx1, Scd1, Atp1b4*) utilizing cDNA prepared from RNA samples used for microarray analysis. We also quantified the expression of MuSK (Muscle specific tyrosine kinase receptor) through qPCR as it was undetectable at any time point on our gene chip. RNA was reverse-transcribed into cDNA using the SuperScript First-Strand Synthesis System (Life Technologies, Grand Island, NY, USA). Samples were diluted 1:100, and qPCR was performed using KAPA SYBR FAST Master Mix (Kapa Biosystems, Woburn, MA, USA) and the Eppendorf Mastercycler system (Eppendorf, Hamburg, Germany). Primers for *Chrna1, Sln, Myl3, Myog, Aqp4, Runx1,* and *GAPDH* were designed in Oligo 6.8 (Molecular Biology Insights, Cascade, CO, USA; Allele Biotechnology, San Diego, CA, USA) while those for *Scd1, MuSK,* and *Atp1b4* were ordered premade from Integrated DNA Technologies, Coralville, IA, USA. Primer sequences for these genes are listed in Table 2.1.

A temperature gradient was used to determine the optimal reaction temperature for each primer based on the DNA melting temperature curve and single product production on an agarose gel. Samples were run in triplicate using the following protocol: samples were heated to 95°C for 2 minutes, then run through 40 cycles of heating at 95°C for 15 seconds, cooling to 55°C for 15 seconds, and heating for 20 seconds to the optimal primer temperature determined by the temperature gradient described above. The triplicate results of each gene from qPCR data were normalized with respect to the housekeeping gene *GAPDH*. Fold change was computed in accordance with a previous publication [16].

2.3.6 Microarray data preprocessing

Expression data were preprocessed using packages available through R [17] and Bioconductor [18]. Gene Chip Robust Multiarray Average (GCRMA) was employed for normalizing expression using the "gcrma" function available through the GCRMA package **[19].** All raw .CEL files along with GCRMA-normalized data is available through Gene expression omnibus (GEO) [20] accession GSE52350. Outlier samples were those with average inter-sample correlation <2 standard deviations (SDs) below mean. A single array at 1 week (3.4 SDs below mean) was removed. Annotation files for RG 230 2.0 (GPL 1355) were downloaded from GEO. Multiple probes were accounted for using the "collapseRows" function in R's WGCNA library [21]. All probes with missing ENTREZ gene identifiers were excluded from this study. Based on this processing, we obtained a final reduced dataset containing log2 based normalized expression values of 13,751 genes across 19 samples.

2.3.7 Differential analysis through pairwise comparisons

Pairwise comparison between every time point BoNT-An injected vs. saline was performed using the Cyber-T [22] Bayes regularized analysis for two-sample unpaired data, with a confidence interval of 8. This study utilized control tissue from saline-injected rats 12 weeks after injection for all pairwise comparisons, in contrast to using age-matched controls. (Rats from this time were considered adult animals, representative of rats from the other time points in the study.) Previous studies showed that skeletal muscle glucose uptake [23] and muscle protein expression [24] changes little among rats until they reach more than 18 months of age. Since all our rats were within this age group, we considered it acceptable to perform pairwise comparison using a control from a single time point. Fold change for each gene was computed as the difference in mean log based expression between treated and control samples. Genes with a log2 based fold of >1 and a Benjamini Hochberg (BH) *P*-value of < 0.05 were identified as being significantly differentially expressed as listed in Table S2.1 online.

2.3.8 Enrichment analysis

DAVID [25] was used to identify enrichment of genes (categories: GO_BP_FAT and KEGG_PATHWAYS) in Table S2.2 online.

2.4 Results

2.4.1. Differential gene expression over time

Gene expression changed dramatically during the experimental time period. Table 2.2 summarizes the number of genes that were identified as being differentially

regulated at each time. Consistent with previous studies, pairwise analysis revealed that muscle is transcriptionally hyperactive, with dramatic transcriptional changes at 1 week (compared to 4, 12, and 52 weeks). Visual analysis of differentially regulated genes suggests that the bulk of regulation occurs at 1 week, with a large fraction of genes (1,718/1,989) being exclusively and significantly regulated at this time (Figure 2.1). As expected, the genes regulated at 1 week cover a wide spectrum of functions, such as stabilizing the NMJ, sarcomeric contraction, and muscle metabolism. Of the 113 genes regulated exclusively at 4 weeks, most were associated with extracellular matrix (ECM) and collagen fibril organization (Table S2.2). No genes were regulated significantly across the entire course of the study.

2.4.2 Systems analysis of differential expression in skeletal muscle

In contrast to using traditional ontology enrichment to analyze transcriptional regulation, we systematically categorized and analyzed differentially expressed genes in the novel framework of "physiological networks" specifically identified in skeletal muscle from 2 previous studies [12, 13]. Yu *et al.* [13] identified families of genes based on 4 major functions occurring in skeletal muscle: mechanical, metabolic, excitation-contraction coupling, and signaling, while Smith *et al* [12] characterized the physiology of the muscle into 8 distinct "networks" required for its functioning. Taken together, these models have identified gene networks that are crucial for normal skeletal muscle function and homeostasis.

Utilizing these models to guide our analyses, we derived a systems view of the regulation underlying skeletal muscle after BoNT-A treatment. Based on our data set,

transcriptional activity of muscle can be grouped into 7 networks: 1) neuromuscular junction, 2) excitation-contraction coupling system, 3) muscle contraction, 4) energy metabolism and mitochondrial biogenesis, 5) extracellular matrix, 6) oxidative stress, and 7) muscle atrophy and recovery (Figure 2.2, Table S2.3). Each of these networks can be considered in their physiological context in light of the genes measured.

1. Neuromuscular junction (NMJ)

Expression changes at the NMJ are illustrated graphically in Figure 2.3. Consistent with previous experimental studies, BoNT-A injection leads to rapid disruption and repair of the NMJ. Genes encoding postsynaptic proteins were detected, including the adult nAChr subunits *Chrna1*, *Chrnd, and Chrne* as well as the developmental subunit *Chrng*, which is usually only expressed in humans prior to the 33rd week of gestation. The co-receptor for Agrin- *Lrp4*, *Emb*, and linker protein *Rapsn* were all upregulated. *Chrna1* and *Emb* were upregulated until 4 weeks. Two immature isoforms of Na^{2+} and K^{+} channels, *Scn5a* and *Kcnn3*, were upregulated significantly at 1 week. Genes selectively involved with the synaptic basal lamina including *Lama5, Col4a5, and Nid2* were upregulated only at 4 weeks.

2. Excitation-contraction coupling (ECC)

Genes involved in ECC and maintenance of calcium homeostasis such as ion pumps and ion channels were differentially regulated, especially at 1 week (Figure 2.4) along with *Cacnb1* (an L-type voltage gated Ca^{2+} channel) and *Fkbp1a* (an *Ryr1* binding protein). Sustained and significant upregulation of sarcoplasmic Ca^{2+} handler Sarcolipin (*Sln*) occurred up to 12 weeks. *Jph1* and genes required to

modulate cytosolic Ca^{2+} levels including *Pde4d, Calm3,* and *Camk2a* were downregulated at 1 week.

Aqp4 was the most strongly downregulated gene, solely at 1 week, along with several other K^+ ion channels such as *Kcnc1, Kcnab1, Kcnj11*, and ion pumps such as *Atp1b1 and Atp1b.*

3. Muscle contraction and activation

Muscle contraction requires coordinated effort between the contracting sarcomeres and cytoskeletal framework. There is a general downregulation of genes associated with activating sarcomere contraction in fast fibers, particularly 1 week after injection (Figure 2.4). These include tropomyosin (*Tpm3, Tpm2*), troponins (*Tnnc1, Tnnt1, and Tnni1*), tropomodulin (*Tmod1*), and genes that encode proteins associated with the sarcomeric contractile apparatus, such as myosin light chains (*Myl2, Myl3*), Myl kinases (*Mylk2*), *Mybpc2,* and myosin heavy chains (*Myh2, Myh7*). There was a significant downregulation of the M line structural proteins including myomesins (*Myom1 and Myom2*), Z-disk associated proteins such as *Actn3, Myot,* myozenins *(Myoz1, Myoz2),* and *Ldb3*. Cytoskeletal proteins *Ank1, Sgc, and LARGE* were downregulated, whereas cytoskeletal proteins required to increase sarcolemmal stability were upregulated (*Csrp3, Dysf, Dtna, Flnc, and Lmna*). We also observed strong upregulation of a muscle-specific calcium handling protein, *Ankrd1,* up to 4 weeks. Upregulation of certain cardiac isoforms such as *Actc1, Myl6b,* and *Tnnt2,* along with immature isoforms normally absent from adult muscle (*Myh3 and Myh8*) was also observed. Consistent with this observation was

the appearance of developmental myosin isoforms in 14% of the 1-week and 1-month injected muscles, but none in control muscle.

4. Energy metabolism and mitochondrial biogenesis

Genes involved in energy metabolism, specifically mitochondrial energy production from glycolysis and β-oxidation, were downregulated significantly 1 week after injection (Figure 2.5). The glucose transporter *Glut4/SLc2a4* and glycolysis intermediates and enzymes *Pgm5, Gys, Pygm, and Pfkfb1* were downregulated at 1 week. Other enzymes involved in generation of pyruvates acting in the cytosol such as *Gpi, Pfkm, Pgam2, Eno3, Pkm2, Pdk4, Ldha,* and *Ldhb* were downregulated at 1 week. Enzymes involved in each step of the TCA cycle, including *Cs, Aco2, Idh2, Idha, Idhb, Idhg, Dlst, Dld, Suclg1, Sdha, Sdhb, Sdhc, Sdhd, Fh1, Mdh1,* and *Mdh2* were downregulated. AMP deaminase (*Ampd3*), required for replenishing TCA cycle intermediates, was upregulated strongly.

Genes associated with β oxidation and lipid metabolism, such as fatty acid transporters *Cd36* and *Fabp3; Lipin-1* (required to break down triacylglycerol to free fatty acids (FFA); ATP-dependent enzymes required to convert FFA into long, medium, and short acyl-CoA esters (*Acadvl, Acsl6, and Acss1*) and their transporters (*Cpt2); Hadh, Echs1, Echdc1, and Echdc2* were all downregulated. Prolonged upregulation of stearoyl-Coenzyme A desaturase 1 (*Scd1*) was observed for most of the study. Genes of the immediate ATP replenishment system of muscle, *Ckmt2* and *Ak1*, were downregulated. Major energy/ATP availability sensor, AMPK α/β/γ

(*Prkaa1*, *Prkag3*), was upregulated. Targets of AMPK, the PPAR cofactors *Ppargc1a* and *Ppargc1b*, were downregulated.

Downregulation of solute carriers necessary for metabolism, such as several members of the mitochondrial phosphate transporter family (*Slc25*- *Slc25a23* in particular) and members of the monocarboxylate transporter subfamily (*Slc16*-*Slc16a3* in particular) was observed.

5. Changes to the extracellular matrix

There was a general upregulation of ECM genes, particularly at 4 weeks after injection. Genes encoding proteins of the basal lamina such as Fbn1, and its collagens (*Col4a1*, *Col4a2*, *Col8a1*) were upregulated (Figure 2.6). Genes associated with the fibrillar ECM, including *Col1a1*, *Col1a2*, *Co114a1*, *Col3a1*, *Col5a1*, *Col5a2*, *Col5a3*, and *Col6a3*, and other ECM-associated genes such as *Lum*, *Ctgf*, *Bgn*, and *Postn* were upregulated. Enzymes *Lox* and *Loxl1* that are involved in collagen crosslinking were also upregulated. *S100a4*, a biomarker correlated with proliferation of fibroblasts, was upregulated through 12 weeks. Increased collagen at the protein level in samples from 4 weeks was also detected using the hydroxyproline assay (Figure S2.1)

6. Oxidative stress response

The most striking change in expression of genes involved in oxidative stress was the global activation of chemoprotective and antioxidant genes, especially at 1 week, such as the isoforms of glutathione-S transferase (*Gst), Gstm1* and *Gstt2;*

Gpx3, Hmox, Nqo1, Aldh3a2, Txn1, and metallothioneins (*Mt1a, Mt2a*). Mitochondrial ROS scavenger *Sod2* was downregulated at 1 week.

7. Muscle atrophy and recovery

After BoNT-A injection, muscle appears to activate conflicting cellular programs, showing simultaneous signs of breakdown and repair. Upregulation of myogenic regulatory factors (MRFs) *Myod1, Myog,* and *Myf6* at 1 week with a concomitant and drastic increase in 2 potent regulators of cell proliferation, *Cdkn1a* and *Cdkn1c*, was observed. These, in conjunction with activated *Pcna*, serve as markers of satellite cell activation in skeletal muscle. Signaling pathways active in BoNT-A treated skeletal muscle are as follows (Figure 2.7).

TGF-β pathway- Several genes in the TGF-β pathway, including *Tgfb2*, *Fst, Myc, Ltbp1,* and early response factors downstream, *Junb* and *Fos,* were upregulated significantly up to 4 weeks. Small GTPases *RhoA* and *RhoC* downstream of the TGF-β pathway were upregulated at 1 week with the *Mstn* receptor *Acvr2b*, *Acvr1*, and a TF, *Atf4* downregulated at 1 week. Interestingly, inhibitors of *Tgfb1* such as *Smad7* and *Fkbp1a* were upregulated at 1week.

NF-κB signaling: Several genes including *Traf2, Nfkb2, Nfkbie,* ubiquitin ligases downstream of the *NF-κB* pathway, *Atrogin1*/ *Fbxo32 MuRF1*/*Trim63,* and *Casp3,* were upregulated at 1 week. Positive activators of the *NF-κB* pathway, such as *Ascc2* and *Litaf* were also upregulated at 1 week.

MAPK signaling: Several members of the MAPK family were upregulated at 1 week, including Mapk1, *Mapk3*, *Map3k1*, *Map3k14*, *Map4k4,* and *Mapk1ip1*,

as well as its downstream targets such as *Eif4e2*, which initiates protein translation and activation of *Myod1*.

Insulin signaling: Though we found no significant regulation of *Igf1* in our study, several genes of this pathway were upregulated at 1 week, including *Igf1r*, *Igfbp5*, *Shc*, and downstream *Pik3r4* and *Akt1*. Regulation of *Igfbp5* and *Igf2* is observed at 4 weeks. *Glut4*/*Slc2a4* and *Irs1* were downregulated at 1 week.

ID signaling pathway: Inhibitor of DNA binding proteins *Id1*, *Id2*, *Id3*, and *Id4* were upregulated up to 4 weeks. This pathway is believed to play a role in repairing muscle.

TP53/ Cell cycle control: Activation of genes that may play a role in satellite cell proliferation and activation of apoptosis, including *Cdkn1a*, *Gadd45a*, *Pcna*, and *Myc*, were upregulated at 1 week.

VEGF pathway: Genes involved in angiogenesis, including *Vegfa*, *Vegfb*, angiopoetins (*Angpt1*, Angpt2), *Nos3*, *Rtn4*, and *Nrp1* were downregulated at 1 week.

Transcriptional regulation of factors required for proteolytic degradation such as Ca^{2+} dependent calpains (*Capn2*, *Capn3*) and lysosomal cathepsins (*Ctsl1*) was seen. *Runx1*, a transcription factor that promotes anti-atrophic programs, was upregulated strongly up to 4 weeks.

2.4.3 Validation of regulated gene expression using qPCR

To validate the chip-based expression results, we performed qPCR on a subset of 8 relevant genes. Some have been highlighted previously as being active in skeletal

muscle during atrophy, specifically after BoNT-A treatment, including *Myog*, *Chrna1*, *Sln*, and *Myl3*. Genes that were shown more recently to be active in atrophy/muscle recovery from atrophy include *Runx1* [26]. Genes that were ranked highly in our analyses with a demonstrated role in skeletal muscle include *Scd1* and *Aqp4*, and 1 with no known role in skeletal muscle but highly ranked in our differential analysis was *Atp1b4*. The average fold change of all genes was normalized with respect to *GAPDH*. Average fold change was computed with reference to the saline-injected samples. The trends seen in qPCR were similar to those computed through microarray differential analysis (Figure 2.8). Quantitative correlation between relative gene expression levels from microarray data and qPCR (r^2) ranged from 0.74 to 0.99.

2.4.4 Correlation of gene expression with muscle function post injection

Because measurement of isometric contraction was made before and after BoNT-A injection on the same set of rats used for gene expression analysis, we are able to study the correlation over the time course of this experiment. We correlated isometric contraction strength post-injection with genes identified as significantly altered for the 1 and 4 week time periods (Table S2.2). At 1 week, 72 genes were positively correlated while 37 genes were negatively correlated ($P<0.05$) with contraction strength. These same 109 genes showed the opposite weak correlation (albeit non-significantly) on isometric force prior to injection (Figure 2.9, left), suggesting that expression levels might be functionally significant. Positively correlated genes were enriched for skeletal muscle contraction and include genes such as *Chrna1*, *Myl2*/*3*, *Tnni1*, *Tnnc1*, *Lama5*, *Scn5a*, *Myoz2* and *Tpm3*. Since a functional contractile apparatus is required for muscle

contraction, it is not surprising that increasing expression of related genes might improve contraction strength after BoNT-A injection. At 4 weeks, we observed 15 genes correlate positively ($P<0.05$) and 37 correlate negatively with isometric force ($P<0.05$; Figure 2.9, right) after injection. Negatively correlated genes were overrepresented for angiogenesis, cell death and ECM (such as *Lox*, *Col1a1*). While it has been shown previously that fibrosis and ECM remodeling may lead to abnormal muscle function [27], the link between angiogenesis and muscle force is less clear. These data may be indicative of still-injured muscle undergoing continued repair and regrowth at 4 weeks.

2.5 Discussion

This study is a high-throughput analysis of global expression changes in BoNT-A treated mammalian skeletal muscle over a period of 1 year (Figure 2.2). Albeit with some differences, the transcriptional regulation observed in chemodenervated muscle after BoNT-A is similar to that seen in denervation models [28] and neuromuscular diseases [29, 30], with suppression of metabolism and muscle contraction, activation of atrophic pathways, and increased oxidative stress. We discuss the observed regulation below, defining the “early” response, which is more complex, and the “late” response to neurotoxin.

2.5.1 Early response to botulinum toxin injection

1. Alterations in excitation-contraction coupling and sarcomeric contraction

Reduced availability of ACh due to BoNT-A impaired exocytosis causes increased expression of specific nAChRs, *Chrnd*, *Chrne*, embryonic *Chrng*, and

especially *Chrna*. As previously reported, overexpression of *Chrng*, ordinarily undetectable in adult skeletal muscle, is a compensatory mechanism to create greater current flow, as it is localized along the entire length of the fiber [31, 32]. Though our microarray did not capture *MuSK*, a crucial protein for nAChR clustering post BoNT-A [10] (upregulation confirmed through qPCR, Figure S2.2), *Lrp4* (a co-receptor for *Agrin* [33]) and *Rapsn (*required for clustering nAChRs on the postsynaptic membrane [34]) were upregulated significantly. The observed upregulation of *Lrp4* and *Emb* may serve as a "retrograde signal" to stabilize the NMJ, facilitating formation of terminal sprouts and induction of nAChRs [35, 36].

Imbalance in ion flux after BoNT-A was reflected in the regulation of adult isoforms of K^+ gating channels such as *Kcnc1*, *Kcnab1*, and inwardly rectifying *Kcnj11*. Interestingly, strong upregulation of 2 non-adult voltage-gated Na^{2+}/K^+ ion channels, *Scn5a* (a cardiac isoform recently implicated in the occurrence of fibrillation potentials in denervated muscle fibers [37]) and *Kcnn3* (a K^+ channel implicated in fibrillation and hyperpolarization of denervated muscles [38,39]), suggests hyper excitability of BoNT-A injected muscle, analogous to denervated models. Reduced electrical activity also implies absence of an active need to maintain the Na^+/K^+ gradient, reducing the utility of certain Na^+/K^+ ion pumps, such as *Atp1b1* and *Atp1b2* [40].

The transmission of neural excitation past the NMJ terminates at a specialized set of voltage sensors within the muscle T-system called the dihydropyridine receptors (DHPR), which are coupled mechanically to ryanodine

receptors (RYR) in the sarcoplasmic reticulum (SR). Downregulation of docking protein *Jph1* (which holds the T-system spatially close to the SR) suggests instability in the structural and spatial association between the SR and T-tubules at 1 week. Upregulation of *Cacnb1* of the DHPR, *Fkbp1a* [41] (essential for minimizing Ca^{2+} leakage) in conjunction with the protracted and sustained upregulation of *Sln* (which inhibits the uptake of Ca^{2+} back into the SR), suggests increased availability of cytoplasmic Ca^{2+}, in contrast to denervated muscle [42].

Genes that affect free Ca^{2+} dynamics, such as *Calm3*, calcineurin (*Ppp3cb*/*Ppp3ca*), and *Camk2a* were clearly regulated. *Aqp4*, a water channel expressed at the sarcolemma of fast-twitch skeletal muscle, whose expression is altered in dystrophy and atrophy, was the single most strongly downregulated gene. Although the exact physiological role of *Aqp4* in skeletal muscle has yet to be defined, recent research on $Aqp4^{-/-}$ mice suggest it has a role in regulating the osmotic balance of muscle, affecting Ca^{2+} handling [43]. Taken together, these suggest a lack of calcium homeostasis and impaired Ca^{2+} handling in BoNT-A treated muscle, especially at 1 week after injection. These results are consistent with muscle responding to decreased neural activity.

The transmitted action potential in normal skeletal muscle is ultimately converted to mechanical contraction through physical coupling of several proteins within the muscle [13]. As expected, we observe suppression of several adult sarcomeric proteins of fast muscle. Downregulation of *Myoz1*, combined with upregulation of *Csrp3* [28], suggests a shift in fiber composition at 1 week post

BoNT-A treatment. This “mixed” state of expression beyond 1 week is further compounded by upregulation of genes expressing cardiac immature fiber isoforms such as *Actc1*, *Myl6b,* and *Tnnt2*, with the largest increases occurring in the expression of *Myh3* and *Myh8*.

A disrupted state of the sarcolemma at 1 week is evidenced by downregulation of *Ank1* [44] (necessary to maintain integrity of network SR) along with upregulation of several other cytoskeletal proteins, *γFilamin*, *Sgc*, *Dmd*, *Dtna*, and *Dysf*. Overall, the observed activation of several mixed muscle isoforms points to activation of programs not seen in adult skeletal muscle, reinforcing the general idea that muscle injected with BoNT-A reverts to a more “immature” state in order to recover contractile function.

2. Reduced metabolism and impaired mitochondrial biogenesis

With BoNT-A induced paralysis, there is reduced requirement for ATP consumption. Akin to denervation models [28], we observed suppression of most genes involved in energy metabolism and production (via both oxidation of lipids and glycolysis), specifically at 1 week, which resolve by 4 weeks. As described in Results and Figure 2.5, there is clear downregulation of enzymes involved in energy production via glycolysis, except for hexokinase (*Hk6*).

It has been reported previously in atrophy with preferential loss of fast muscle fibers that there is dramatic upregulation of *Ampd3* [45] (replenishes TCA-cycle intermediate substrates). Dramatic and prolonged upregulation of *Ampd3* in our study not only suggests an impaired TCA cycle, but further supports the idea of

a shift in fiber composition. Another observation of BoNT-A treated muscle is prolonged upregulation of *Scd1* (also validated using qPCR). Deficiency in Scd1, a rate-limiting enzyme that catalyzes the synthesis of monounsaturated fatty acid, has been correlated with increased oxygen consumption and subsequent β-oxidation in skeletal muscle [46,47].Conversely, overexpression studies have shown decreased fatty acid oxidation, increased TAG synthesis, monosaturation of muscle fatty acids and impaired glucose uptake and insulin signaling pathway [48]. The fact that it is upregulated until 12 weeks after injection leads us to speculate that *Scd1* may play a significant role in reduced energy production (via β-oxidation) in addition to playing a role in reducing glucose uptake post BoNT-A injection.

Suppression of *Ckmt2* (outer mitochondrial membrane enzyme required for generating ATP from phosphocreatine and ADP) and *Ak1* (cytoplasmic enzyme that catalyzes generation of ATP from ADP) also point to reduced ATP turnover in injected muscle.

Upregulation of major energy/ATP availability sensors, AMPK α/β/γ (*Prkaa1*, *Prkag3*), point to reduced availability of ATP at 1 week. However, the targets of AMPK, the PPAR cofactors- *Ppargc1a* and *Ppargc1b*, known biomarkers of mitochondrial biogenesis in skeletal muscle, were downregulated suggesting possible stress-induced impairment of mitochondrial biogenesis [49-51].

3. Increased oxidative stress and metal ion imbalance

Denervation and immobilization studies have demonstrated repeatedly that 1 of the causes of atrophy is increased accumulation of reactive oxygen species

(ROS) and trace metals in skeletal muscle [52-54]. In that same vein, we observed transcriptional activation of several oxidative stress markers implicated in atrophy, such as the metallothioneins (Mt1a, Mt2a) [45, 55]. Most striking however, was the increase in the various isoforms of glutathione-S transferase including *Gstm1* and *Gstt2* as a compensatory response to increased production of ROS or oxidative stress [56]. Interestingly, however, the mitochondrial ROS scavenger superoxide dismutase 2 (*Sod2*) was downregulated and may have been confounded by mitochondrial dysfunction.

Imbalance of metal ion concentration has been reported previously in studies of immobilization and disuse [57]. Zinc ion homeostasis has been linked closely to a redox state of cells in various tissues [58]. We observe significant upregulation of zinc solute carriers (*Slc30a2* and *Slc30a4*), which are suggested to confer a cytoprotective effort by preventing cells from free Zn ion toxicity [59]. While the exact physiological role of Zn ion toxicity in chemodenervated muscle is not understood fully, we hypothesize that the observed upregulation of these transporters in conjunction with increased expression of metallothioneins suggests a metal ion imbalance that may contribute to BoNT-A induced atrophy of muscle.

4. Competing pathways contributing to concomitant atrophy and recovery of skeletal muscle

Atrophy and consequent muscle loss in skeletal muscle can occur through activation of the NF-κB signaling pathway (*Traf2, Nfkb2,* and *Nfkbie* and positive activators *Ascc2* and *Litaf*) along with activation of TGF-β pathway [60]. Loss of

muscle mass has been attributed to accelerated proteolytic degradation of the contractile apparatus through deployment of factors such as *Capn2, Ctsl1,* and *Casp3* [61] and eventual degradation of the fragmented actin-myosin complexes through the ubiquitin-proteasomal system. Similar to denervation studies, but in contrast to neuromuscular diseases [29, 30] activation of atrophy markers- *Atrogin1/MAFbx (Fbxo32)* and *Trim63 (MuRF1),* 2 muscle-specific ubiquitin ligases downstream of the NF-κB pathway is observed following BoNT-A treatment [11, 62].

Several studies have demonstrated the role of TGF-β signaling in atrophying skeletal muscle and the powerful role of TGF-β family growth factors such as myostatin (*Mstn*) in regulating muscle size [63]. Though differential regulation of *Mstn* was not observed in our study, follistatin (*Fst*) [64], a myostatin inhibitor was upregulated significantly. This, taken along with the repression of *Acvr2b*, a transmembrane activin receptor of *Mstn,* points to inhibition of the pro-atrophic action of *Mstn* in injected muscle. Upregulation of early response genes downstream of *Tgfb1,* including *Junb* and *Fos,* along with small GTPases *RhoA* and *RhoC* and its inhibitors (*Smad7* and *Fkbp1a),* further emphasizes the conflicting signaling of muscle treated with BoNT-A. In contrast to ATPases such as *Atp1b1* or *Atp1b2*, *Atp1b4* has been shown to localize to intracellular stores, predominantly the inner myonuclear membrane, in perinatal skeletal muscle of placental mammals and to regulate TGF- β signaling in skeletal muscle. Though no direct evidence of its localization patterns exists, we hypothesize that prolonged upregulation of *Atp1b4*

(validated via qPCR) may be contributing to similar functions in BoNT-A treated muscle.

Insulin-like growth factors (IGFs) and their role in upregulation of nAChRs, muscle growth [69, 70], and their metabolic effects, have been studied extensively. The observed regulation of IGF binding proteins, such as *Igfbp5 (*inhibits action of *Igf1* by sequestering it to ECM [67] and suppresses nerve sprouting [7]) is consistent with previous studies of BoNT-A treatment [11]. Though we found no significant regulation of *Igf1*, upregulation of its receptor, *Igf1r*, may compensate for the decreased availability of *Igf1*. Activation of *Igf1r* results in phosphorylation of insulin receptor substrates (*Irs1)* and regulation of several downstream players such as *Akt1*, *Pik3* (*Pik3r4*), and the energy/ATP availability sensor-AMPK α/β/γ. The observed spike in *Igf2* at 4 weeks correlates with studies that have shown a preferential spike in *Igf2* nearly 20 days post-denervation/nerve injury [65].

Although the exact role of myogenic regulatory factors (MRFs) in differentiated post-mitotic skeletal muscle is not understood fully [68], the observed upregulation of *MyoD*, *Myog*, and *Mrf4*/*Myf6* may reflect satellite cell activation [69]. These factors may be necessary for activating transcriptional programs required for recovery of muscle function, such as *Ankrd1*. Concomitant with activation of MRFs, there is upregulation of several cell-cycle control genes known to play a role in satellite cell proliferation, such as *Tp53*, *Pcna*, *Myc*, cyclin-dependent kinase inhibitors *Cdkn1a* (suggested to confer a protective, anti-apoptotic effect [70]), and *Cdkn1c*. *Gadd45a*, a marker for atrophy also involved in cell cycle

control has been identified repeatedly as being overexpressed in models of denervation/chemodenervation [10, 11, 28]. Recent reports have indicated that the pro-atrophy transcription factor *Atf4* may induce expression of *Gadd45a* in muscles subject to 3 distinct skeletal muscle stresses: fasting, immobilization, and denervation [71]. Its expression was shown to be necessary but not sufficient for expression of *Gadd45a*. However, we observe a conflicting program of regulation in our data with *Atf4* being downregulated, suggesting alternate roles for *Atf4* and regulation of *Gadd45a* in BoNT-A treated muscle at 1 week.

Further regulatory conflicts occur through upregulation at 1 week of 4 inhibitors of DNA binding genes (*Id1*, *Id2*, *Id3*, and *Id4*), which have been shown to inhibit muscle growth and differentiation [72]. Reduced contractile activity leads to a reduction in signaling that promotes muscle growth but inhibits complete fiber death (autophagy) triggered through pathways such as *Pik3*/*Akt* [73] and activation of runt transcription factor (*Runx1*). *Runx1* has been shown to sustain muscle under atrophic conditions by inducing expression of genes required for muscle growth and function (*Myh2*, *Scn5a*, *Rrad*, *Myh3*, and *Chrng*) and repressing atrophy genes (*Gadd45a* and *Aqp4*) [26].

2.5.2 Later response to botulinum toxin injection

Importantly, we found that, by 4 weeks, the transcriptional events leading to muscle atrophy and weakness were essentially completed. Although the functional properties of muscle are highly impaired at this time point, the transcriptional response is essentially complete and is in the process of recovering. This is seen clearly seen by

the fact that, of the 1,989 genes regulated after injection, only 231 were actually still changing after 4 weeks.

1. Recovering NMJ, sustained oxidative stress, and lack of Ca^{2+} homeostasis

Though the muscle slows down transcriptionally, the expression of certain pathways is still significant at 4 weeks. In contrast to other genes of the NMJ, *Emb* and *Chrna1* were still upregulated up to 12 weeks, with significant remodeling of the synaptic basal lamina (*Lama5*, *Col4a5*, and *Nid2* were all upregulated) at 4 weeks. *Nid2* is involved in synapse maintenance [74] and is associated selectively with the synaptic basal lamina at the NMJ. Taken together, these suggest reinnervation and continuing effort by muscle for stabilization of the NMJ at 4 weeks.

Upregulation of the SR calcium sequestering protein *Casq2* suggests an effort by muscle to maintain Ca^{2+} within the SR, possibly counteracting continued over-activation of *Sln*. Interestingly, we found strong upregulation of a calmodulin-like protein called *Calml3* beginning at 4 weeks. Though its exact function is not yet determined in skeletal muscle, it is known to compete with calmodulin in other tissues, further suggesting alterations in the Ca^{2+} handling properties of muscle [75]. Upregulation of certain transcription regulators of atrophy and growth such as *Myod1*, *Id1*, *Id3*, *Runx1*, *Gadd45a*, *Cdkn1a*, and *Cdkn1c* were observed until 4 weeks.

2. ECM remodeling and fibrosis

The most pronounced effect at 4 weeks was active remodeling of the ECM and possibly even fibrosis. ECM production is regulated in part by activation of several targets of TGF-β, including *Ctgf* and *Ltbp1*. Upregulation of early growth response (*Egr-1*), a zinc-finger transcription factor known to regulate collagen expression (particularly *Col1a2*) in response to TGF-β [76] and act downstream of multiple pro-fibrotic agents to regulate transcription, was observed. The persistent activation of these genes in conjunction with dramatic upregulation of ECM genes beyond 1 week (see Results and Figure S2.1) leads us to propose activation of similar fibrotic programs by 4 weeks after BoNT-A injection, resulting in fibrosis of injected tissue. This also emphasizes the possibility of a multi-faceted role of *Igfbp5* following BoNT-A injection [77-79].

Significant regulation of *Scd1*, *Sln*, *Cdkn1a*, *Cdkn1c,* and *S100a4* beyond 4 weeks suggested incomplete recovery of muscle even 4 weeks after treatment. Importantly, we observed no biologically significant changes, at least transcriptionally, in skeletal muscle treated with BoNT-A at 52 weeks after injection.

2.6 Conclusion

In conclusion, this analysis provides a global assessment of changes occurring in BoNT-A treated muscle over a period of 1 year. By utilizing previously described physiological networks of muscle, we provided a systems-level analysis that

categorically assesses expression changes after BoNT-A treatment. Dramatic transcriptional regulation in several of these networks was evident at week 1 leading to derangement of the ECM and fibrillar components by week 4. The shift toward expression of slow and immature isoforms emulating “immature” muscle possibly aids in muscle recovery. It should be noted that, while this is a transcriptional expression study, the data have clinical relevance. Specifically, they indicate that, at the molecular level, the effects of BoNT-A are relatively rapid, since most transcripts returned to control level within 4 weeks. This is consistent with the use of the term “reversible chemodenervation” often used with reference to the action of BoNT-A. It is also interesting to note that, in spite of the relatively fast transcriptional response, the structural and functional response lags somewhat. This is probably a function of the length of time required for a neuromuscular unit to recover from a period of denervation-induced atrophy and fibrosis. Transcriptional regulation associated with atrophy and fibrosis suggests the possibility of transient extracellular effects after BoNT-A injection. Though no long-term transcriptional abnormalities were observed, further studies are necessary to determine optimal intervals for BoNT-A treatment from both a biological and physiological point-of-view.

2.7 Acknowledgements

The content of chapter 2 is a modified presentation of the manuscript published in Muscle and Nerve, titled “Systems analysis of transcriptional data provides insights into muscle's biological response to botulinum toxin” by Mukund K, Mathewson M,

Minamoto V, Ward SR, Subramaniam S, and Lieber RL. Muscle & nerve. 2014 Nov 1; 50(5):744-58. The dissertation author was the primary author for this material. Funding for this study was provided by the Department of Veterans Affairs Grant RX000670 (RL), NIH grants R24HD050837 (RL), AR057013 (SRW), Allergan, Inc. (RL) and National Heart, Lung, and Blood Institute Grants HL087375-02 (SS), HL106579 (SS) and HL108735 (SS) and NSF Collaborative Grant STC-0939370 (SS)

2.8 Tables

Table 2.1: List of forward and reverse strand primer sequences that were utilized for validation of gene expression using qPCR.

Gene	**Forward Sequence**	**Reverse Sequence**
Chrna1	*TACTTGAATCCTTTCGCGCT*	*CTTAACCGCTGAGCCATCTC*
Sln	*TGGTGTGCACTCAGAAGTCC*	*TGAGGAGCACAGTGATCAGG*
Myl3	*AATCCTACCCAGGCAGAGGT*	*CATATGTGCCCGTGTCTTTG*
Myog	*ACCAGGAGCCCCACTTCTAT*	*TTACACACCTTACACGCCCA*
Aqp4	*GCATGTGATCGACATTGACC*	*GTGAAACAAGAAACCCGCAT*
Runx1	*TAACCCTGCCTGGGTGTAAG*	*GGACTCGGATCTTCTGCAAG*
GAPDH	*AGACAGCCGCATCTTCTTGT*	*TGATGGCAACAATGTCCACT*

Table 2.2: Summary of differentially regulated genes identified at each time (with respect to saline injected muscle, BH<0.05).

Time (in weeks)	**Differentially Expressed**	**Upregulated**	**Downregulated**
1	1989	1183	806
4	372	303	69
12	32	19	13
52	32	19	13

2.9 Figures

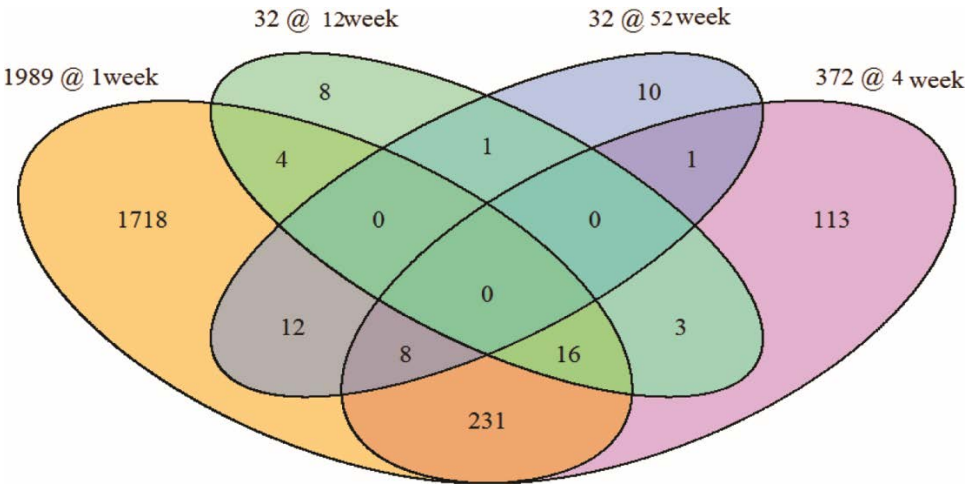

Figure 2.1: A 4-way Venn diagram depicting the distribution of differentially expressed gene across all pairwise comparisons
Additionally, the figure also provides counts of overlapping genes between multiple pairwise comparisons. The count within each shaded area represents the number of differentially expressed genes identified in common between time points. For example, 231 genes were differentially expressed at both 1 and 4 weeks.

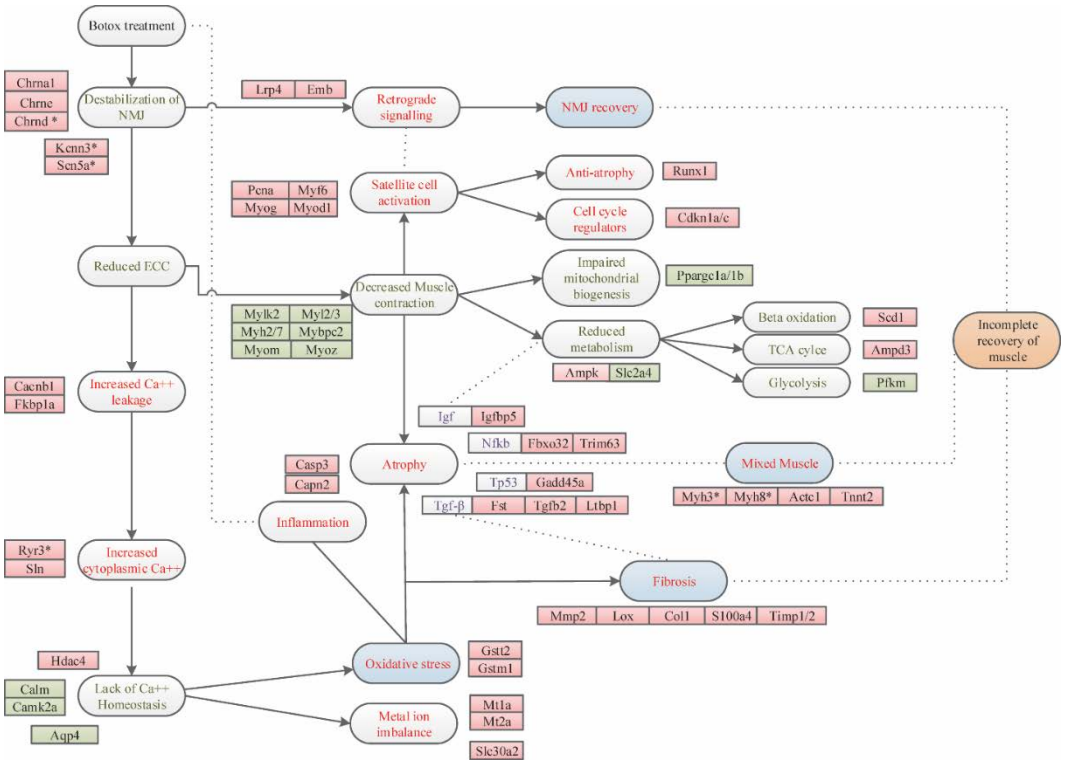

Figure 2.2: Overview of transcriptional changes occurring in adult skeletal muscle after BoNT-A treatment.
Functional changes represented in blue boxes show associated gene expression until 4 weeks after injection. Genes in yellow boxes exhibit upregulation, and blue boxes exhibit downregulation. Genes identified with an * represent immature muscle isoforms. Signaling pathways are indicated with red text. Dotted lines indicate association, and arrows indicate a cause-effect relationship.

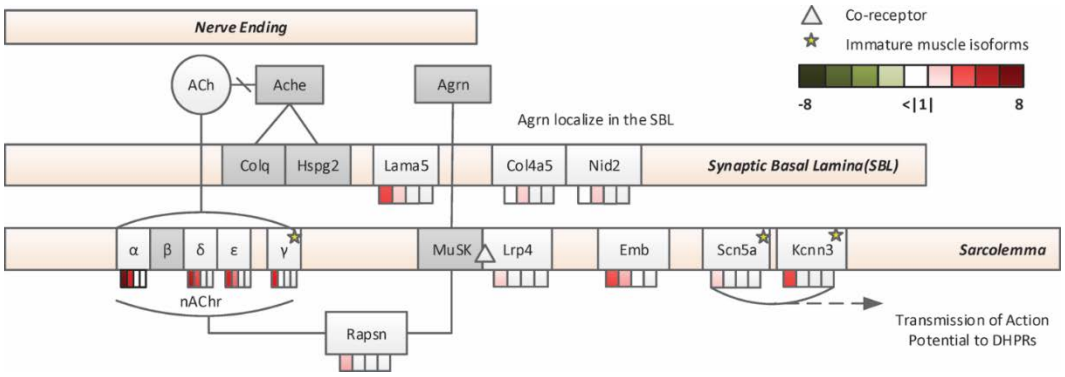

Figure 2.3: Expression of genes involved in the neuromuscular junction over time. Each box has 4 partitions representing the time points at which samples were obtained. Each box represents the fold changes observed and is colored according to the legend. Genes that are not represented in the microarray or identified as differentially expressed are presented in grey. Solid lines indicate either an interaction among genes or an association between genes.

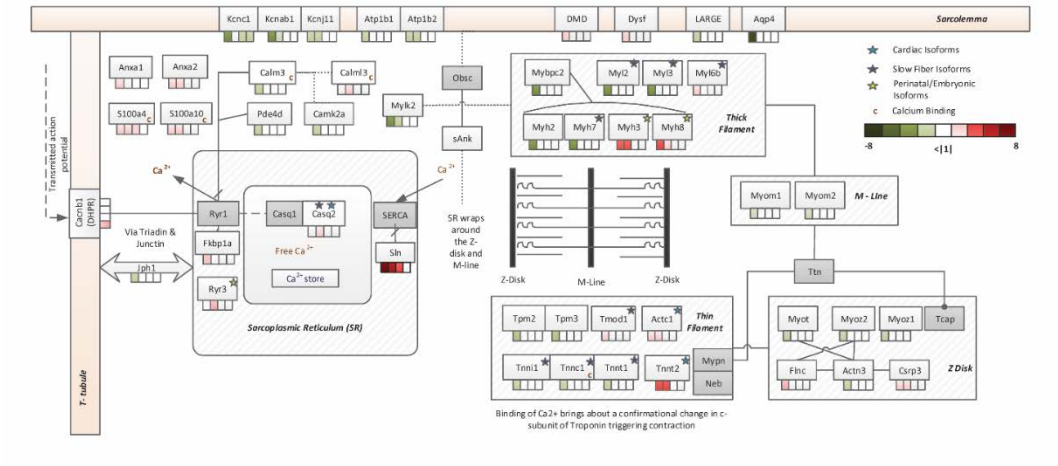

Figure 2.4: Expression of genes involved in excitation-contraction coupling and muscle contraction after BoNT-A injection.
Cellular localization as well as expression levels are depicted as described in Figure 2.3.

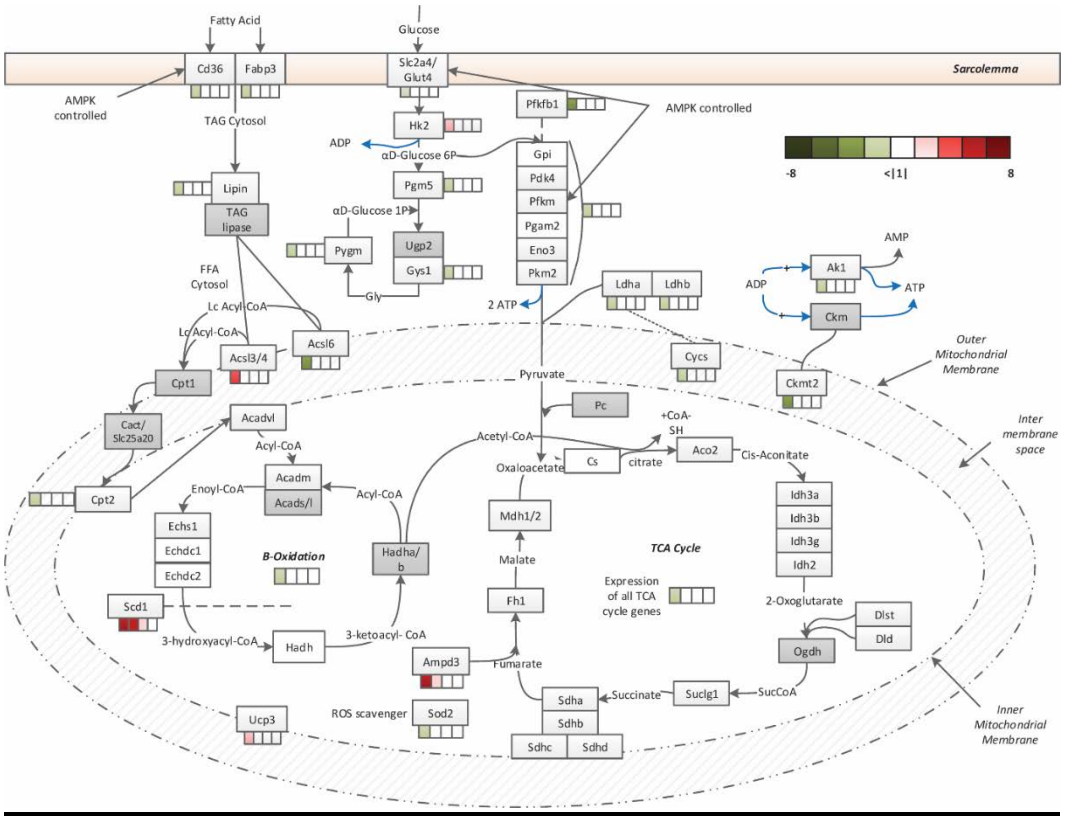

Figure 2.5: Expression of genes involved in mitochondrial metabolism after BoNT-A Cellular localization as well as expression levels are depicted as described in Figure 2.3.

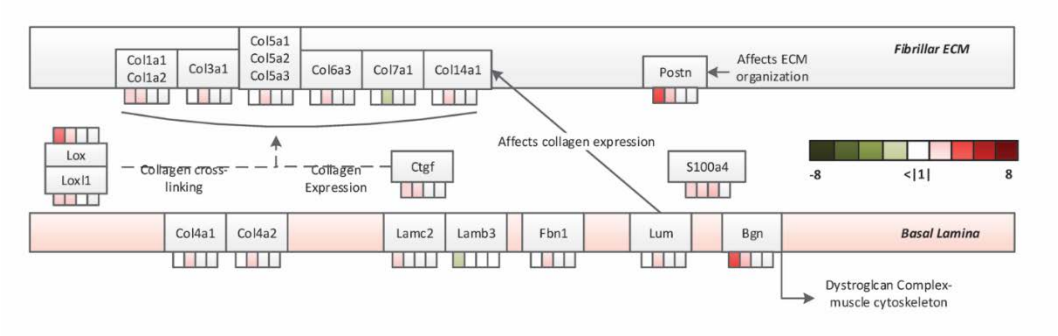

Figure 2.6: Expression of genes involved in the basal lamina and fibrillar ECM. Cellular localization as well as expression levels are depicted as described in Figure 2.3.

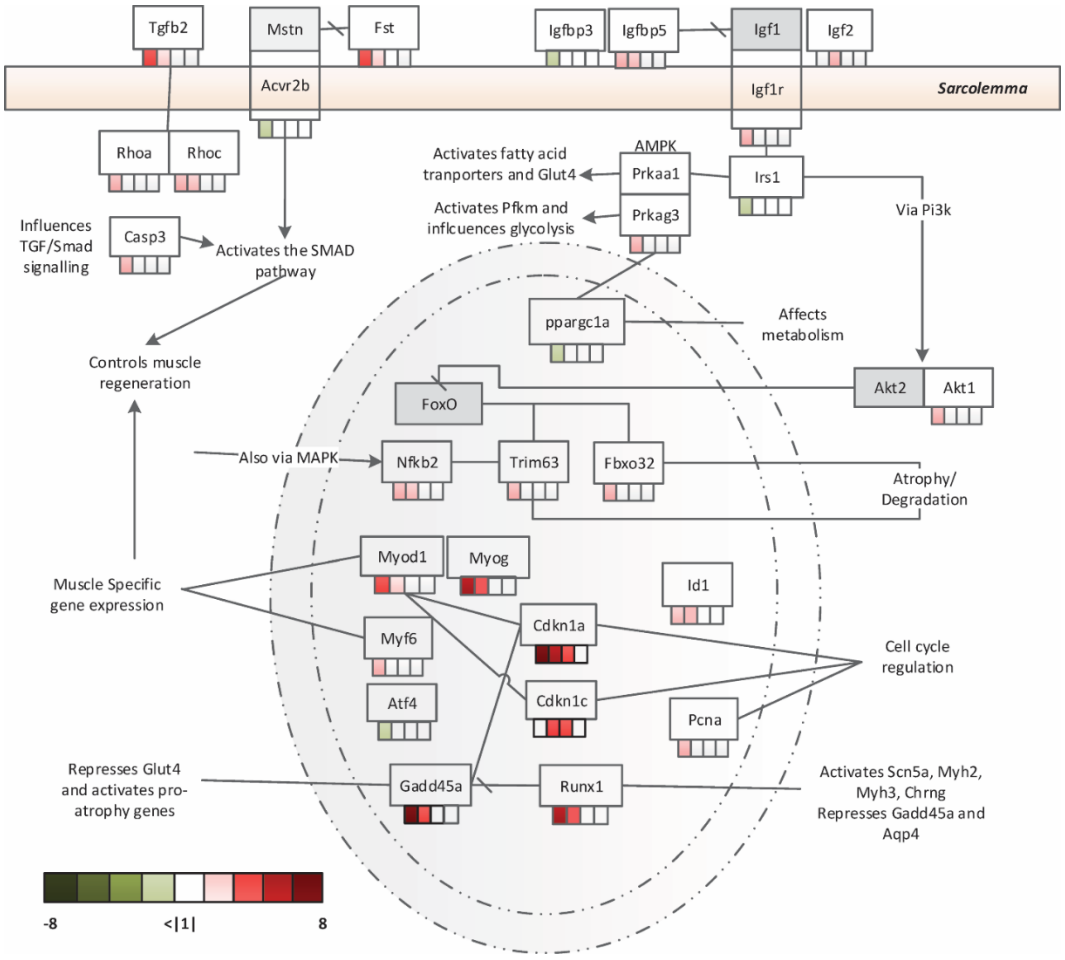

Figure 2.7: Illustration of a representative set of active transcription factors and signaling pathways involved in atrophy and muscle recovering from BoNT-A injection. A detailed list of genes is provided in the Table S2.3. Cellular localization as well as expression levels (fold changes) are depicted as described in Figure 2.3.

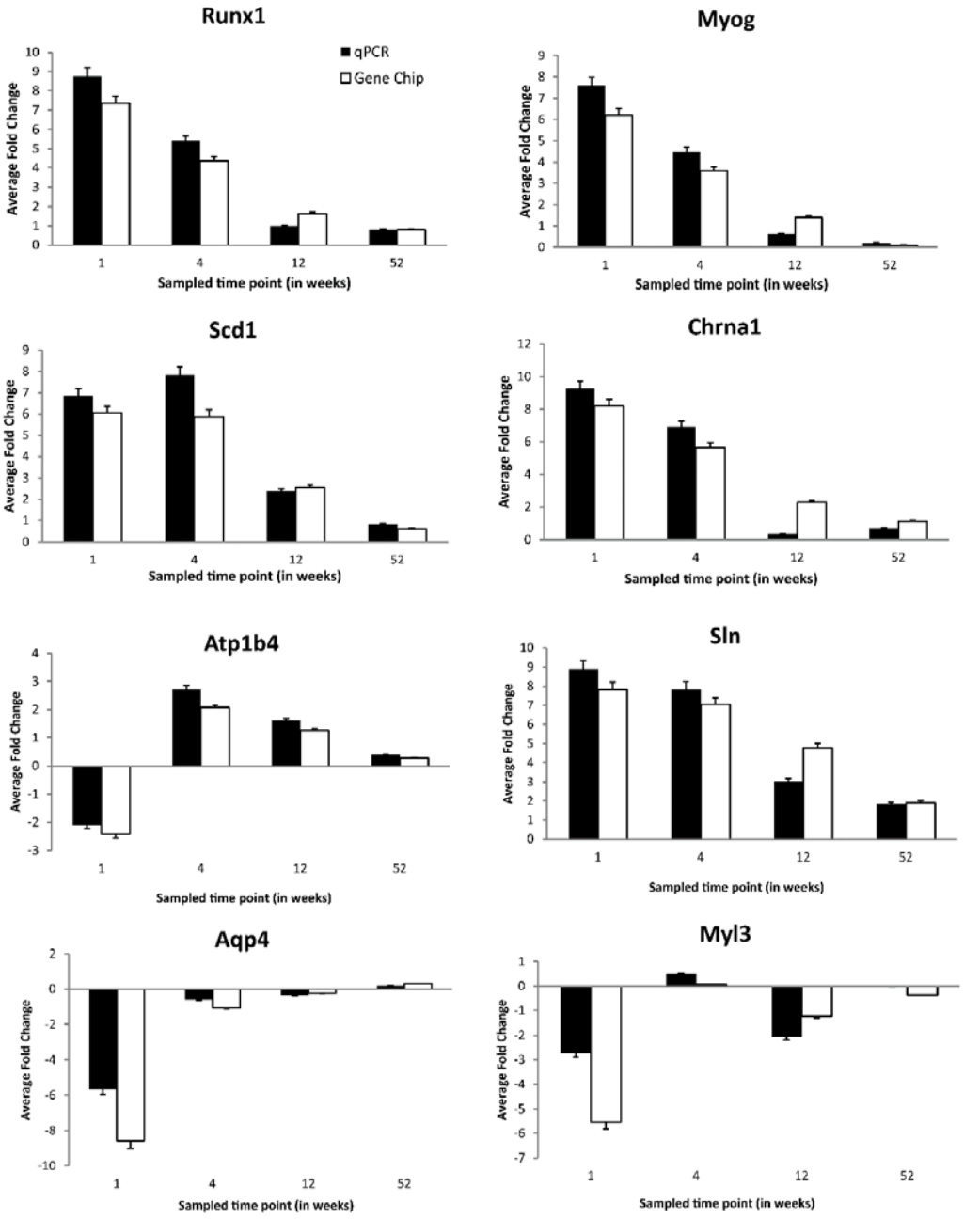

Figure 2.8: Fold changes observed based on the qPCR assay compared to the microarray data. Each plot provides a comparison between the genes's calculated average fold change (log2 based) with respect to control using qPCR and microarray analysis computed for each time point.

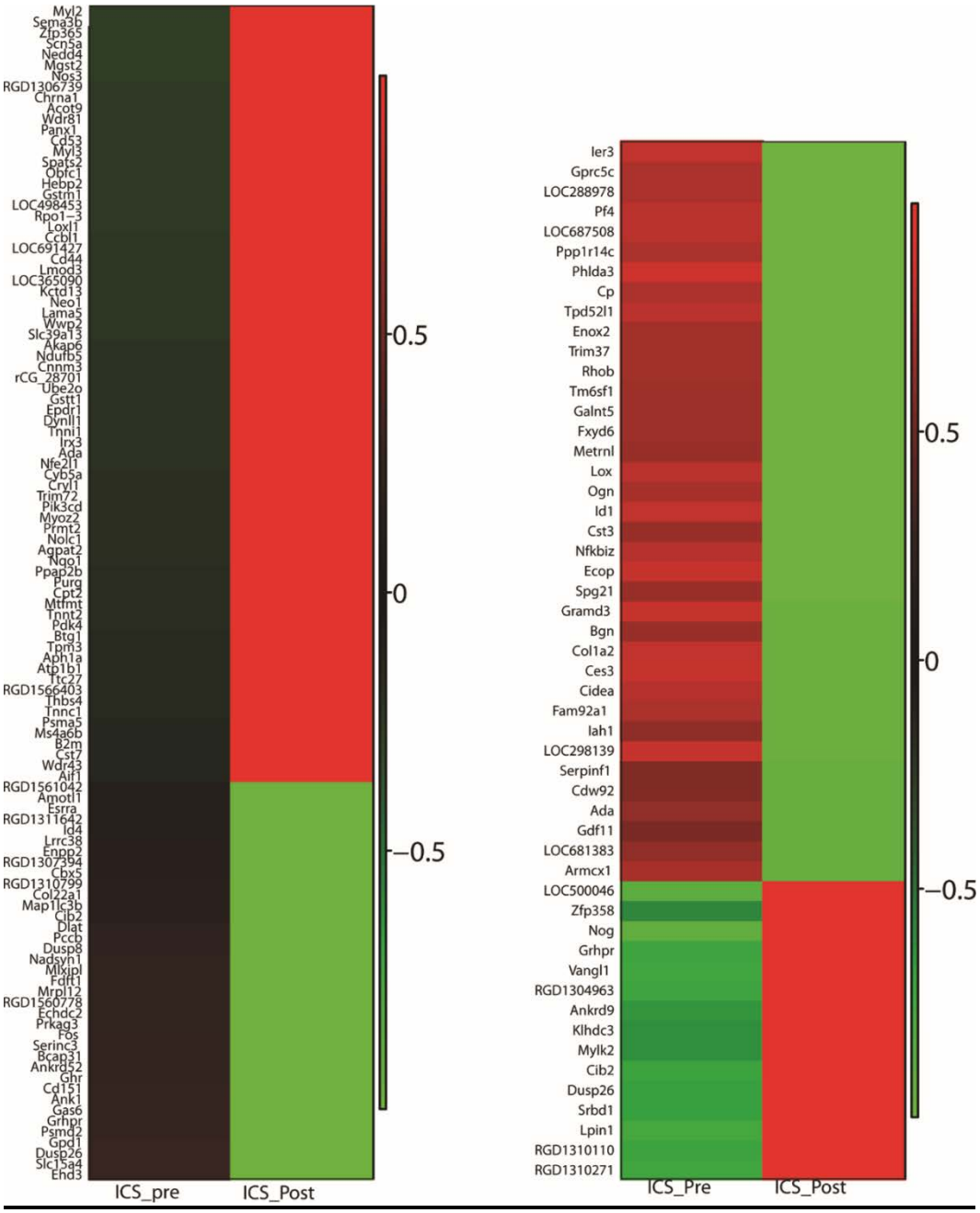

Figure 2.9: Heatmap showing the correlation between differentially expressed genes With isometric contractile strength at a) 1 week (left) and (b) 4 weeks after injection (right).Green represents negative correlation and red represents positive correlation.

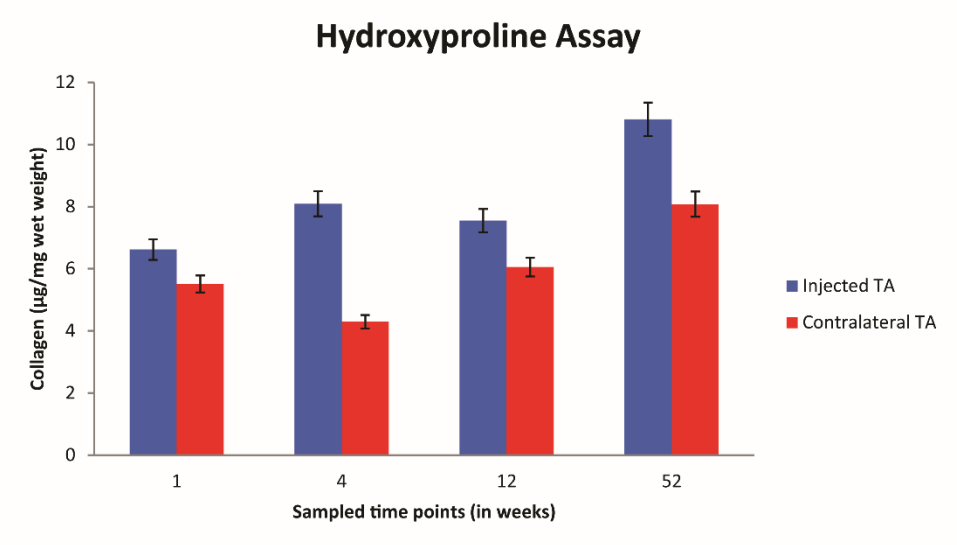

Figure S2.1: Hydroxyproline assay in BoNT-A injected muscle
The assay was performed to estimate collagen content with respect to the contralateral muscle.

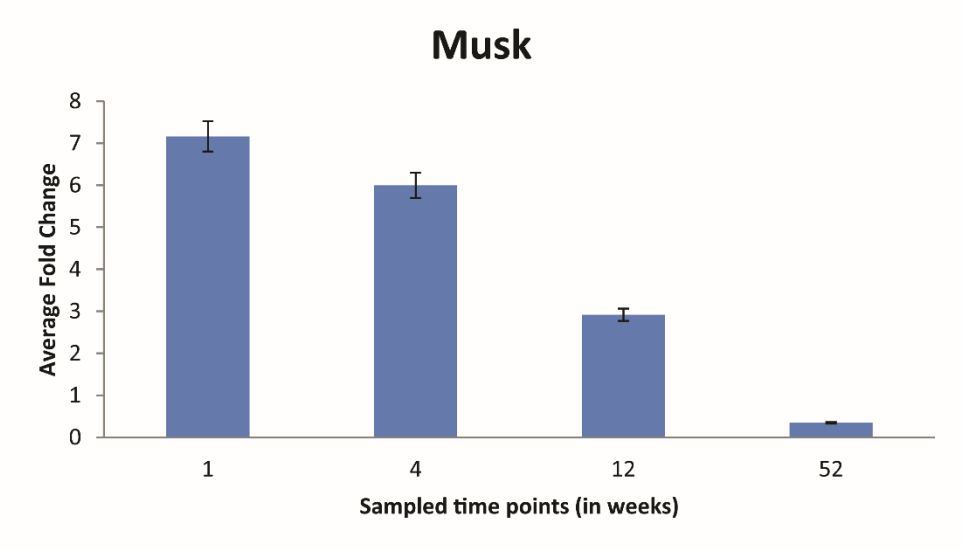

Figure S2.2: Musk expression as established through RT-PCR. Expression of MuSK was not identified using our gene chip.

2.10 References

1. McLaughlin JF, Bjornson KF, Astley SJ, Graubert C, Hays RM, Roberts TS, Price R, Temkin N. Selective dorsal rhizotomy: efficacy and safety in an investigator-masked randomized clinical trial. Dev Med Child Neurol. 1998;40(4):220–32.

2. Koman LA, Mooney III JF, Smith BP, Walker F, Leon JM. Botulinum toxin type A neuromuscular blockade in the treatment of lower extremity spasticity in cerebral palsy: a randomized, double-blind, placebo-controlled trial. J Pediatr Orthop. 2000;20(1):108.

3. Truong DD, Stenner A, Reichel G. Current clinical applications of botulinum toxin. Curr Pharm Des. 2009;15(31):3671-80.

4. Dolly JO, Lawrence GW, Meng J, Wang J, Ovsepian SV. Neuro-exocytosis: botulinum toxins as inhibitory probes and versatile therapeutics. Curr Opin Pharmacol. 2009;9(3):326–35.

5. Simpson LL. Identification of the major steps in botulinum toxin action. Annu Rev Pharmacol Toxicol. 2004;44:167–93.

6. De Paiva A, Meunier FA, Molgó J, Aoki KR, Dolly JO. Functional repair of motor endplates after botulinum neurotoxin type A poisoning: biphasic switch of synaptic activity between nerve sprouts and their parent terminals. Proc Natl Acad Sci. 1999;96(6):3200–5.

7. Meunier FA, Schiavo G, Molgó J. Botulinum neurotoxins: from paralysis to recovery of functional neuromuscular transmission. J Physiol-Paris. 2002;96(1):105–13.

8. Duchen LW. Changes in the electron microscopic structure of slow and fast skeletal muscle fibres of the mouse after the local injection of botulinum toxin. J Neurol Sci. 1971;14(1):61–74.

9. Lebeda FJ, Cer RZ, Stephens RM, Mudunuri U. Temporal characteristics of botulinum neurotoxin therapy. Expert Rev Neurother. 2010;10(1):93–103.

10. Ma J, Elsaidi GA, Smith TL, Walker FO, Tan KH, Martin E, Martin E, Koman LA, Smith BP. Time course of recovery of juvenile skeletal muscle after botulinum toxin A injection: an animal model study. Am J Phys Med Rehabil. 2004;83(10):774–80.

11. Shen J, Ma J, Lee C, Smith BP, Smith TL, Tan KH, Koman LA. How muscles recover from paresis and atrophy after intramuscular injection of botulinum toxin A: study in juvenile rats. J Orthop Res. 2006;24(5):1128–35.

12. Smith LR, Meyer G, Lieber RL. Systems analysis of biological networks in skeletal muscle function. Wiley Interdiscip Rev Syst Biol Med. 2013;5(1):55–71.

13. Wang Y, Winters J, Subramaniam S. Functional classification of skeletal muscle networks. I. Normal physiology. J Appl Physiol. 2012;113(12):1884–901.

14. Minamoto VB, Hulst JB, Lim M, Peace WJ, Bremner SN, Ward SR, Lieber RL. Increased efficacy and decreased systemic-effects of botulinum toxin A injection after active or passive muscle manipulation. Dev Med Child Neurol. 2007;49(12):907–14.

15. Edwards CA, O'brien WD. Modified assay for determination of hydroxyproline in a tissue hydrolyzate. Clin Chim Acta. 1980;104(2):161–7.

16. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative CT method. Nat Protoc. 2008;3(6):1101–8.

17. RDevelopment C. TEAM: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0;

18. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):R80.

19. Wu Z, Irizarry RA, Gentleman R, Murillo FM, Spencer F. A model based background adjustment for oligonucleotide expression arrays. J Am Stat Assoc. 2004; 99: 909–917.

20. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013;41(D1):D991–D995.

21. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559.

22. Kayala MA, Baldi P. Cyber-T web server: differential analysis of high-throughput data. Nucleic Acids Res. 2012;40(W1):W553–W559.

23. Goodman MN, Dluz SM, McElaney MA, Belur E, Ruderman NB. Glucose uptake and insulin sensitivity in rat muscle: changes during 3-96 weeks of age. Am J Physiol-Endocrinol Metab. 1983;244(1):E93–E100.

24. Piec I, Listrat A, Alliot J, Chambon C, Taylor RG, Bechet D. Differential proteome analysis of aging in rat skeletal muscle. FASEB J. 2005;19(9):1143–5.

25. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2008;4(1):44–57.

26. Wang X, Blagden C, Fan J, Nowak SJ, Taniuchi I, Littman DR, Burden SJ. Runx1 prevents wasting, myofibrillar disorganization, and autophagy of skeletal muscle. Sci Signal. 2005;19(14):1715.

27. Mann CJ, Perdiguero E, Kharraz Y, Aguilar S, Pessina P, Serrano AL, Muñoz-Cánoves P. Aberrant repair and fibrosis development in skeletal muscle. Skelet Muscle. 2011;1(1):21–21.

28. Raffaello A, Laveder P, Romualdi C, Bean C, Toniolo L, Germinario E, Megighian A, Danieli-Betto D, Reggiani C, Lanfranchi G. Denervation in murine fast-twitch muscle: short-term physiological changes and temporal expression profiling. Physiol Genomics. 2006;25(1):60–74.

29. Millino C, Fanin M, Vettori A, Laveder P, Mostacciuolo ML, Angelini C, Lanfranchi G. Different atrophy-hypertrophy transcription pathways in muscles affected by severe and mild spinal muscular atrophy. BMC Med. 2009;7(1):14.

30. Mo K, Razak Z, Rao P, Yu Z, Adachi H, Katsuno M, Sobue G, Lieberman AP, Westwood JT, Monks DA. Microarray analysis of gene expression by skeletal muscle of three mouse models of Kennedy disease/spinal bulbar muscular atrophy. PloS One. 2010;5(9):e12922.

31. Numberger M, Dürr I, Kues W, Koenen M, Witzemann V. Different mechanisms regulate muscle-specific AChR gamma-and epsilon-subunit gene expression. EMBO J. 1991;10(10):2957.

32. Witzemann V, Brenner HR, Sakmann B. Neural factors regulate AChR subunit mRNAs at rat neuromuscular synapses. J Cell Biol. 1991;114(1):125–41.

33. Kim N, Stiegler AL, Cameron TO, Hallock PT, Gomez AM, Huang JH, Hubbard SR, Dustin ML, Burden SJ. Lrp4 is a receptor for Agrin and forms a complex with MuSK. Cell. 2008;135(2):334–42.

34. Goodsell DS. Neuromuscular synapse. Biochem Mol Biol Educ. 2009;37(4):204–10.

35. Lain E, Carnejac S, Escher P, Wilson MC, Lømo T, Gajendran N, Brenner HR. A novel role for embigin to promote sprouting of motor nerve terminals at the neuromuscular junction. J Biol Chem. 2009;284(13):8930–9.

36. Yumoto N, Kim N, Burden SJ. Lrp4 is a retrograde signal for presynaptic differentiation at neuromuscular synapses. Nature. 2012;489(7416):438–42.

37. Sekiguchi K, Kanda F, Mitsui S, Kohara N, Chihara K. Fibrillation potentials of denervated rat skeletal muscle are associated with expression of cardiac-type voltage-gated sodium channel isoform Nav1. 5. Clin Neurophysiol. 2012;123(8):1650–5.

38. Kostrominova TY, Dow DE, Dennis RG, Miller RA, Faulkner JA. Comparison of gene expression of 2-mo denervated, 2-mo stimulated-denervated, and control rat skeletal muscles. Physiol Genomics. 2005;22(2):227–43.

39. Kimura T, Takahashi MP, Fujimura H, Sakoda S. Expression and distribution of a small-conductance calcium-activated potassium channel (SK3) protein in skeletal muscles from myotonic muscular dystrophy patients and congenital myotonic mice. Neurosci Lett. 2003;347(3):191–5.

40. Hundal HS, Marette A, Ramlal T, Liu Z, Klip A. Expression of β subunit isoforms of the Na+,K+-ATPase is muscle type-specific. FEBS Lett. 1993 Aug 16;328(3):253–8.

41. Engel A, Franzini-Armstrong C. Myology: basic and clinical. McGraw-Hill, Medical Pub. Division; 2004.

42. Midrio M. The denervated muscle: facts and hypotheses. A historical review. Eur J Appl Physiol. 2006;98(1):1–21.

43. Basco D, Nicchia GP, D'Alessandro A, Zolla L, Svelto M, Frigeri A. Absence of aquaporin-4 in skeletal muscle alters proteins involved in bioenergetic pathways and calcium handling. PloS One. 2011;6(4):e19225.

44. Ackermann MA, Ziman AP, Strong J, Zhang Y, Hartford AK, Ward CW, Randall WR, Kontrogianni-Konstantopoulos A, Bloch RJ. Integrity of the network sarcoplasmic reticulum in skeletal muscle requires small ankyrin 1. J Cell Sci. 2011;124(21):3619–30.

45. Lecker SH, Jagoe RT, Gilbert A, Gomes M, Baracos V, Bailey J, Price SR, Mitch WE, Goldberg AL. Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. FASEB J. 2004;18(1):39–51.

46. Dobrzyn A, Dobrzyn P. Stearoyl-CoA desaturase—a new player in skeletal muscle metabolism regulation. J Physiol Pharmacol. 2006;57(Suppl 10):31–42.

47. Dobrzyn A, Ntambi JM. The role of stearoyl-CoA desaturase in the control of metabolism. Prostaglandins Leukot Essent Fatty Acids. 2005;73(1):35–41.

48. Voss M, Beha A, Tennagels N, Tschank G, Herling A, Quint M, Gerl M, Metz-Weidmann C, Haun G, Korn M. Gene expression profiling in skeletal muscle of Zucker diabetic fatty rats: implications for a role of stearoyl-CoA desaturase 1 in insulin resistance. Diabetologia. 2005;48(12):2622–30.

49. Cantó C, Gerhart-Hines Z, Feige JN, Lagouge M, Noriega L, Milne JC, Elliott PJ, Puigserver P, Auwerx J. AMPK regulates energy expenditure by modulating NAD+ metabolism and SIRT1 activity. Nature. 2009;458(7241):1056–60.

50. Bouzakri K, Zachrisson A, Al-Khalili L, Zhang BB, Koistinen HA, Krook A, Zierath JR. siRNA-based gene silencing reveals specialized roles of IRS-1/Akt2 and IRS-2/Akt1 in glucose and lipid metabolism in human skeletal muscle. Cell Metab. 2006 Jul;4(1):89–96.

51. Olesen J, Kiilerich K, Pilegaard H. PGC-1α-mediated adaptations in skeletal muscle. Pflüg Arch-Eur J Physiol. 2010;460(1):153–62.

52. O'Leary MF, Hood DA. Denervation-induced oxidative stress and autophagy signaling in muscle. Autophagy. 2009;5(2):230–1.

53. Abruzzo PM, di Tullio S, Marchionni C, Belia S, Fanó G, Zampieri S, Carraro U, Kern H, Sgarbi G, Lenaz G, Marini M. Oxidative stress in the denervated muscle. Free Radic Res. 2010;44(5):563–76.

54. Kondo H, Miura M, Itokawa Y. Oxidative stress in skeletal muscle atrophied by immobiIization. Acta Physiol Scand. 1991;142(4):527–8.

55. Muller FL, Song W, Jang YC, Liu Y, Sabia M, Richardson A, Van Remmen H. Denervation-induced skeletal muscle atrophy is associated with increased mitochondrial ROS production. Am J Physiol-Regul Integr Comp Physiol. 2007;293(3):R1159–R1168

56. Reid M, Jahoor F. Glutathione in disease. Curr Opin Clin Nutr Metab Care. 2001;4(1):65–71

57. Kondo H, Miura M, Nakagaki I, Sasaki S, Itokawa Y. Trace element movement and oxidative stress in skeletal muscle atrophied by immobilization. Am J Physiol - Endocrinol Metab. 1992;262(5):E583–E590.

58. Maret W, Krężel A. Cellular zinc and redox buffering capacity of metallothionein/thionein in health and disease. Mol Med. 2007;13(7-8):371.

59. Maret W. Metallothionein redox biology in the cytoprotective and cytotoxic functions of zinc. Exp Gerontol. 2008;43(5):363–9.

60. Jackman RW, Kandarian SC. The molecular basis of skeletal muscle atrophy. Am J Physiol-Cell Physiol. 2004;287(4):C834–C843.

61. Plant PJ, Bain JR, Correa JE, Woo M, Batt J. Absence of caspase-3 protects against denervation-induced skeletal muscle atrophy. J Appl Physiol. 2009;107(1):224–34.

62. Bodine SC, Latres E, Baumhueter S, Lai VK-M, Nunez L, Clarke BA, Poueymirou WT, Panaro FJ, Na E, Dharmarajan K, Pan ZQ. Identification of ubiquitin ligases required for skeletal muscle atrophy. Sci Signal. 2001;294(5547):1704.

63. Kollias HD, McDermott JC. Transforming growth factor-β and myostatin signaling in skeletal muscle. J Appl Physiol. 2008;104(3):579–87.

64. Amthor H, Nicholas G, McKinnell I, Kemp CF, Sharma M, Kambadur R, Patel K. Follistatin complexes Myostatin and antagonises Myostatin-mediated inhibition of myogenesis. Dev Biol. 2004;270(1):19–30.

65. Sullivan KA, Kim B, Feldman EL. Insulin-like growth factors in the peripheral nervous system. Endocrinology. 2008;149(12):5963–71.

66. Sacheck JM, Ohtsuka A, McLary SC, Goldberg AL. IGF-I stimulates muscle growth by suppressing protein breakdown and expression of atrophy-related ubiquitin ligases, atrogin-1 and MuRF1. Am J Physiol-Endocrinol Metab. 2004;287(4):E591–E601.

67. Schneider MR, Wolf E, Hoeflich A, Lahm H. IGF-binding protein-5: flexible player in the IGF system and effector on its own. J Endocrinol. 2002;172(3):423–40.

68. Sabourin LA, Rudnicki MA. The molecular regulation of myogenesis. Clin Genet. 2000;57(1):16–25.

69. Chen C-M, Stott NS, Smith HK. Effects of botulinum toxin A injection and exercise on the growth of juvenile rat gastrocnemius muscle. J Appl Physiol. 2002;93(4):1437–47.

70. Walsh K, Perlman H. Cell cycle exit upon myogenic differentiation. Curr Opin Genet Dev. 1997;7(5):597–602.

71. Ebert SM, Dyle MC, Kunkel SD, Bullard SA, Bongers KS, Fox DK, Dierdorff JM, Foster ED, Adams CM. Stress-induced skeletal muscle Gadd45a expression reprograms myonuclei and causes muscle atrophy. J Biol Chem. 2012;287(33):27290–301.

72. Jen Y, Weintraub H, Benezra R. Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. Genes Dev. 1992;6(8):1466–79.

73. Stitt TN, Drujan D, Clarke BA, Panaro F, Timofeyva Y, Kline WO, Gonzalez M, Yancopoulos GD, Glass DJ. The IGF-1/PI3K/Akt pathway prevents expression of muscle atrophy-induced ubiquitin ligases by inhibiting FOXO transcription factors. Mol Cell. 2004;14(3):395–403.

74. Fox MA, Ho MS, Smyth N, Sanes JR. A synaptic nidogen: developmental regulation and role of nidogen-2 at the neuromuscular junction. Neural Develop. 2008;3(1):1–17.

75. Bennett RD, Mauer AS, Pittelkow MR, Strehler EE. Calmodulin-like protein upregulates myosin-10 in human keratinocytes and is regulated during epidermal wound healing in vivo. J Invest Dermatol. 2008;129(3):765–9.

76. Chen S-J, Ning H, Ishida W, Sodin-Semrl S, Takagawa S, Mori Y, Varga J. The early-immediate gene EGR-1 is induced by transforming growth factor-β and mediates stimulation of collagen gene expression. J Biol Chem. 2006;281(30):21183–97.

77. Sureshbabu A, Okajima H, Yamanaka D, Shastri S, Tonner E, Rae C, Szymanowska M, Shand JH, Takahashi SI, Beattie J, Allan GJ. IGFBP-5 induces epithelial and fibroblast responses consistent with the fibrotic response. Biochem Soc Trans. 2009;37(4):882.

78. Sureshbabu A, Tonner E, Allan GJ, Flint DJ. Relative Roles of TGF-β and IGFBP-5 in Idiopathic Pulmonary Fibrosis. Pulm Med. 2011;2011.

79. Yasuoka H, Hsu E, Ruiz XD, Steinman RA, Choi AM, Feghali-Bostwick CA. The fibrotic phenotype induced by IGFBP-5 is regulated by MAPK activation and egr-1-dependent and-independent mechanisms. Am J Pathol. 2009;175(2):605–15.

Chapter 3- Botulinum Neurotoxin-A Effects on Skeletal Muscle from Gene Co-Expression Networks: A Time Course Study.

3.1 Abstract

Background- Botulinum Neurotoxin A (BoNT-A) is a potent neurotoxin, which causes temporary paralysis of striated muscle and consequently is used for clinical applications such as pain management and movement disorders. While physiological effects of treatment with BoNT-A in muscle have been studied extensively, mechanistic understanding of its temporal transcriptional response is less understood. The primary objective of this study was to utilize a systems approach to decipher temporal transcriptional response of muscle to BoNT-A; and better assess its functional recovery. Methods- We systematically analyzed data containing transcriptional measurements from adult rat muscle treated with BoNT-A (t= 1, 4, 12 and 52 weeks after treatment). A ranked list of genes, varying across time was calculated using previously published Bayesian statistic. Top 2000 from the ranked list was used as seed set for analysis. Weighted gene co-expression network analysis (WGCNA) was used for generation and analysis of the co-expression network derived from the seed set. Results- 19 co-expressed modules were detected in our network and were re-clustered into groups (n=5). Quantifying average expression and co-expression patterns across these groups revealed temporal aspects of muscle's response to BoNT-A. Our results indicated an association of groups 1 and 5 with week 1, and groups 2 and 3 with week 4 samples. Functional analysis revealed enrichment of group 1 with metabolism; group 5 with contradictory functions of atrophy and cellular recovery; and groups 2 and 3 with extracellular matrix (ECM) and non-fast fiber isoforms. Topological positioning of two highest ranked genes- Dclk1 and Ostalpha within group 5, in conjunction with their

significant expression suggested possible mechanistic role in recovery from BoNT-A induced atrophy. Phenotypic correlations of groups with titin and myosin protein content further emphasized the effect on the sarcomeric contraction machinery in the early phase of chemodenervation. Conclusion- This study provides a network approach to assessing changes associated with cross-sectional temporal data. Grouping of modules revealed a hierarchical functional response to changes occurring during the course of BoNT-A-induced paralysis with an early metabolic response and later response affecting ECM. Additionally, our results provide an unbiased validation of the response documented in our previous work.

3.2 Introduction

Botulinum neurotoxin type A (BoNT-A) is a potent neurotoxin that *functions to temporarily* paralyze striated muscle by inhibiting exocytosis of acetylcholine (ACh), a neurotransmitter. This inhibition causes a series of downstream events leading to the absence of an action potential at the neuromuscular junction necessary for muscle excitation and contraction [1]. The ability of BoNT-A to cause a prolonged, albeit temporary paralysis has made it a useful therapeutic agent in addressing diseases with neurological and neuro-muscular consequences e.g. [2–5]. The physiological changes associated with muscle after treatment with BoNT-A have been studied extensively, in both human and murine models, at single and multiple time points e.g. [6–11]. Such studies have been crucial in determining the dose response and efficacy of BoNT-A, for varied therapeutic use. However, only a handful of studies have focused on deciphering

transcriptional control of BoNT-A in striated muscle [12], and much fewer in assessing the transcriptional changes over the course of time [13]. It is imperative to gain a comprehensive understanding of BoNT-A action for better use clinically. As a first step, we utilize genomic approaches to deciphering transcriptional regulation occurring after BoNT-A treatment in murine models over a period of 1 year. We recently analyzed the functional changes associated with mammalian skeletal muscle after BoNT-A treatment via differential analysis of genes, and interpreted our results using the framework of established physiological muscle networks [14]. The data was acquired over a period of 1 year with measurements made at 1, 4, 12, and 52 weeks after BoNT-A injection. In contrast to our previous study, here we utilize a data driven, co-expression networks approach to analyzing the cross-sectional, temporal data. The basic premise of co-expression studies is that genes that are functionally related tend to be highly correlated and hence often co-expressed in cells. Such studies focus less on differentially expressed genes, but more on groups of genes that tend to be strongly "co-expressed" [15, 16]. We hypothesized that such an approach would provide comprehensive insights into hitherto uncharacterized mechanistic changes and biomarkers underlying the muscle's response to BoNT-A treatment, additionally instantiating functional pathways previously implicated in changes associated with BoNT-A treatment [14].

We utilized an approach called Weighted Gene Co-expression Network Analysis (WGCNA) that elucidates gene relationships based on their co-expression profiles, integrated into a higher order network structure [17]. Modules identified through this approach have been shown to be functionally cohesive (even at low sample

sizes) in single [18–20] and multiple time point studies [21]. We elucidated module "groups" that were differentially correlated across samples from different time points, suggesting that by employing WGCNA on our time course data, we were able to extract modules that exhibited a temporal component to their regulation. We also identified genes not previously associated with chemodenervated muscle and speculated on their putative roles in muscle by analyzing their network topology and transcriptional factor binding sites. Finally, we evaluated the influence of gene modules on selected phenotypic traits such as isometric contraction force, myosin and titin content by correlating their measurements with the transcriptional changes (Figure S3.1).

3.3 Methods

3.3.1 Data acquisition

The data utilized in this study is publicly available through the Gene Expression Omnibus accession GSE52350. A detailed description of data generation and pre-processing is available in the original publication [14]. In brief, samples were obtained for 5 sets of rats (n=4/set) that included the Tibialis anterior (TA) muscle of BoNT-A injected rats at 1, 4, 12 and 52 weeks after injection and control tissue from the contralateral TA of saline injected rats sacrificed at 12 weeks. Minimal changes with muscle protein expression [22] and skeletal muscle glucose uptake [23] in rats until an age of >18 months prompted the use of a single control time point in the original study. Gene expression measurements were made using Affymetrix Rat Genome 230 2.0 arrays (GPL1355). 31099 probes from these 20 were GCRMA normalized (log2 based

expression) and outliers removed resulting in a reduced dataset comprising of 13751 genes across 19 arrays. This reduced dataset of the original publication was utilized for analysis and served as the input data for our current analysis.

3.3.2 Real time quantitative PCR

Real-time quantitative PCR (qPCR) was carried out utilizing cDNA prepared from RNA samples used for microarray analysis, to validate the expression of *Dclk1* and *Ostalpha*. RNA was reverse-transcribed into cDNA using the SuperScript First-Strand Synthesis System (Life Technologies, Grand Island, NY, USA). Samples were diluted 1:100 and qPCR was performed using KAPA SYBR FAST Master Mix (Kapa Biosystems, Woburn, MA, USA) and the Eppendorf Mastercycler system (Eppendorf, Hamburg, Germany). Primers for *GAPDH* (forward sequence: *AGACAGCCGCATCTTCTTGT* and reverse sequence: *TGATGGCAACAATGTCCACT)* was designed in Oligo 6.8 (Molecular Biology Insights, Cascade, CO, USA; Allele Biotechnology, San Diego, CA, USA) while *Dclk1* and *Ostalpha* were ordered premade (Integrated DNA Technologies, Coralville, IA, USA). A temperature gradient was used to determine the optimal reaction temperature for each primer based on the DNA melting temperature curve and single product production on an agarose gel. Samples were run in triplicate using the following protocol: samples were heated to 95°C for 2 minutes, then run through 40 cycles of heating at 95°C for 15 seconds, cooling to 55°C for 15 seconds, and heating for 20 seconds to the optimal primer temperature a determined by the temperature gradient described above. The triplicate results of each

gene from qPCR data were normalized with respect to the housekeeping gene *GAPDH*. Fold change was computed in accordance with reference [24].

3.3.3 Description of Phenotypic Measurements

Phenotypic measurements were made on the same 15 adult rats prior to sacrifice [25] after treatment with BoNT-A. The phenotypic traits measured were 1.Isometric contraction force measured in the injected leg before and after injection (ISO Pre and ISO Post). 2. Titin (isoforms 1 and 2) and 3.mysoin chains (I/IIA/IIB/IIX). Quantitation for titin and myosin was performed on both the injected and contralateral legs of BoNT-A treated rats. Missing measurements were treated as NA for any computation.

3.3.4 Multivariate empirical Bayes statistic

Tai and Speed [26] described a method to rank genes for cross-sectional replicated developmental microarray time course studies utilizing a multi-sample multivariate empirical Bayes model with structured means. Briefly, a linear model was fit to each condition separately and the temporal profile of each sample was compared using the regression coefficients. A ranking was generated using a structured mean design. The interested reader is directed to the original paper for detailed statistics [22]. A second set of control measurements were obtained by duplicating the first set and varying it very slightly, to allow for 2 degrees of freedom within the design matrix. . This was based on the premise that we would observe very minimal difference in gene expression of control samples over time [14]. Of the 13751 genes that were ranked, a small subset containing 2000 highly ranked genes, representing ~15% of the total gene set was utilized as the seed set to build our co-expression network.

3.3.5 Co-expression network generation and modularity detection

We utilized an open source, R/Bioconductor package called WGCNA [27] for analysis. Since we were particularly interested in exploring the association amongst genes under the treatment condition; we chose to work with the top 2000 genes across 15 BoNT-A treated samples rather than pooled data. The correlation was computed using biweight mid-correlation, a robust correlation metric that works beneficially at lower sample sizes. A soft power β was suitably chosen as outlined in [17]. Dissimilarity computed using the topological overlap matrix (TOM) served as input to hierarchical clustering. The hybrid algorithm with a dynamic cut height, part of the WGCNA library was used for modularity detection. Genes that didn't fall into a module were excluded from our analysis. The 1st principal component ("eigen gene") that captured maximum variance for each module was computed. Clustering of module eigen genes resulted in groups of highly correlated modules. Correlating eigen genes with phenotypic traits identified module groups that were strongly correlated with a particular phenotype. For each module, the module membership measure (kME) and intramodular connectivity (kWithin) were computed to identify gene hubs (with normalized kWithin and kME >0.95).

3.3.6 Protein interaction network

Protein interactions with combined score >0.2, for all genes within module M3 (group 5) were extracted from STRING [28] and imported into Cytoscape [29]. A low confidence threshold ensured that all possible interactions for relatively less annotated genes such as *Dclk1* and *Ostalpha* were included. We extracted the immediate

interaction partners and partners of partners (2-step neighborhood) for *Dclk1*. No interactions were found among genes in M3 for *Ostalpha.*

3.3.7 Visualization and Functional enrichment analysis

Network modules generated using WGCNA were exported and visualized in Cytoscape [29] an open source network visualization software. In order to constrain the number of edges for ease of visualization, we exported the top 25 percentile of edges from each module to Cytoscape. ClueGO [30], a plug-in available through Cytoscape was utilized to annotate gene functions and identify functional enrichment of modules using the latest updates of Gene Ontology (Biological process) and KEGG pathways. Heatmaps were generated in R, using functions from gplots and WGCNA libraries.

3.3.8 Identifying Transcription factor binding sites

MotifDB, a library in R, was utilized to extract promoter sequences 1000bp (1kb) upstream of genes using rat build rn5 as reference [31]. Sequence based motif searches were also performed using the same package.

3.4 Results and Discussion

In the current study, we have used microarray data previously generated by our group, from skeletal muscle of rats treated with BoNT-A extracted at 4 time points (t= 1, 4, 12, 52 weeks) after treatment [14], in combination with functional properties of the muscles measured over the same time period [25]. The expression data consisted of transcript levels for 31,099 microarray probes, representing 13,751 genes across 15 samples (see Methods).

3.4.1 Ranking time varying genes using the empirical Bayes statistic

Given the complexity of interpreting and deriving a robust co-expression network from the entire set of 13,571 genes, we chose to work with a smaller subset of genes that could comprehensively capture the changes in BoNT-A muscle, across time. We utilized the empirical Bayes statistic [26] to rank our list of 13,751 genes (see Methods) and chose the top ~15% (2,000) highly ranked genes as our "seed set" for network reconstruction.

To ascertain that the genes contained in the seed set indeed played a role in treated muscle, we compared our seed set with the significantly differential genes (at any time point) from our previous study [14]. We identified an overlap of 61% which included several markers of BoNT-A treatment such as the cholinergic receptors- *Chrna1, Chrnd, Chrng, Chrne*; transcription factors and atrophy markers such as *Cdkn1a, Runx1, Gadd45a*; Calcium handling proteins- *Sln*; Immature myosin isoforms- *Myh3, Myh8*; Oxidative stress markers- *Mgst2, Mt1a, Mt2a*, among others. However, a lower fraction of genes associated with mitochondrial metabolism were identified in our seed set, indicative of a fairly quick stabilization of mitochondrial metabolism gene expression across the time course. Interestingly, the two highest-ranking genes in our seed set were *Ostalpha* and *Dclk1* with no currently evidenced role in skeletal muscle function (expression validated using qPCR, Figure 3.1).

3.4.2 Reconstructing the BoNT-A transcriptional network

Using WGCNA [27], we reconstructed the BoNT-A transcriptional network. The initial adjacency matrix was first computed by raising the gene-by-gene correlation

matrix to a power ß in order to eliminate spurious/weak correlations. ß was chosen at 8 (R^2 = 0.92). Clustering of the dissimilarity matrix resulted in 19 co-expressed modules (Figure 3.2A) with each module ranging in size between 20 - 360 genes (Figure 3.2B). 76 genes that did not belong to any module were excluded from further analysis for the purposes of this study (Figure 3.2A, shown in light grey). The modules were re-clustered into five groups (Figure 3.2C) based on their module eigen gene correlation (see Methods).

3.4.3 Systems elucidation of BoNT-A treatment in muscle

Prior to performing a systematic analysis of the groups identified we attempted to understand the overarching role of the groups across time. To this extent, we computed the module eigen-gene correlations (see Methods), which provided insight into the collective behavior of the modules within groups. We observed strong intra-group correlation (along the diagonal, Figure 3.3A) while the inter-group correlation was fairly non-significant except for a strong anti-correlation between groups 1 and 5.

Additionally we computed the *average expression,* which captures the collective behavior of all co-expressed genes within a module (Figure 3.3B). Average expression patterns in conjunction with eigen-gene correlations allowed us to gain a broad view of the temporal distribution of groups. We observed that the grouping of modules corresponded broadly with the time points of the study. For instance, groups 1 and 5 showed opposing average expression patterns in the samples from week 1 while reversing expression trends at all other time points. This suggested an activation of genes within group 1 and an inhibition of genes regulated within group 5, at week 1.

Average expression patterns along with the eigen-gene correlations suggested genes with contrasting functions were captured in groups 1 and 5 and were strongly associated with samples from week 1 after injection. In the following sections we provide a detailed analysis of the functions captured by each of the groups that further justify the observed correlation patterns.

1. Group 1

An enrichment analysis of repressed modules from group 1(M5, M7 and M19) revealed functions associated specifically with metabolism and cellular homeostasis, which is consistent with the expected reduction in energy requirement due to, reduced muscle activity. For instance, module M19 was enriched for several genes associated with metabolism ($p<0.05$) particularly glucose, carbohydrate and phosphate metabolism such as Acls5, Acsm5, Acss1, St8sia5, Pde4b, Pde4d, Gpt2, Irs1, Fbp2, Prkcz, Phkg1, Ppp1r3c, Ppp1r3d, Eif4e. Aqp4, Gpd1, Fbp2, Pde4a, Slc16a3 and Ppp1r3c. The strong co-expression of Aqp4 (hub, see Methods) [32] with other metabolic genes within this module suggested a role for Aqp4 in influencing metabolic activity of the chemodenervated muscle. Likewise, M5 was particularly enriched for several genes associated with ion homeostasis necessary for maintenance of action potential (such as Cav3, Ank3, Ptpn3, and Kcna5) and metabolism. Regulation of ion homeostasis and metabolism are tightly coupled in skeletal muscle under most conditions [33]. The presence of functionally associated modules (such as M19 and M5) within a single group further reinforces the ability of our approach to identify functionally cohesive groups from time course studies.

2. Group 5

Modules of group 5 (M2, M3, M6, M8 and M18) accounted for 50% of the seed set and were broadly associated with contradictory functions of atrophy and cellular recovery after chemodenervation. Relevant gene ontology categories ($p<0.05$) such as regulation of cell death; wound healing, skeletal muscle development, neuron differentiation, response to organic substance and oxidative stress was identified within this group. Table S3.1 lists the enrichment identified for relevant modules in groups 1 and 5.

For instance, module M18, showed an enrichment for markers associated with oxidative stress and metal ion imbalance such as metallothioneins *(Mt1a, Mt2a)* [34–36] while module M3 showed significant enrichment for contrasting functions of atrophy and recovery. M3 contained genes associated with the NMJ, for example, *Chrna1* and *Chrnd1 (*ACh receptor subunits) were found to be strongly co-expressed with genes implicated in the recovery of the neuromuscular junction - *Lrp4*, *Emb* and *Casp3* [37–39]. Absence of *Casp3* has been shown to protect muscle from denervation induced atrophy [39] while *Emb* and *Lrp4* may serve as retrograde signals for the recovery of NMJ. Other pro-growth markers such as the Myogenic regulatory factors (MRFs)- *Myog* and *Myod1* [40], *Runx1* (sustains muscle under atrophic conditions [41]) and inhibitors associated with TGF β pathway, *Ltbp1* and *Postn.* Though the exact role of MRFs in post mitotic muscle is yet to be clearly understood, their co-expression with factors necessary for maintaining the trophicity of muscle suggests their role in activating the necessary programs after

chemodenervation. *Sln* , a marker for BoNT-A treatment [13], involved in Ca^{2+} handling across the sarcoplasmic reticulum was identified to be strongly co-expressed in this module. In addition to these markers of muscle recovery we identified several markers of atrophy within M3 including genes of the TGF- β pathway such as *Tgfb2, Fst, Rhoa, Rhoc* and *Cdkn1a, Gadd45a* [35, 42].

A further advantage of co-expression network analysis over pairwise analysis of differential expression is that it allows one to decipher the functional role of genes based on their topology [16]. As mentioned earlier, two highly ranked genes *Dclk1* and *Ostalpha*, with no evidenced role in treated or normal skeletal muscle, exhibited significant expression in our samples (Figure 3.1). These genes were identified as being strongly co-expressed within M3. *Dclk1* is known to catalyze protein-protein interactions associated with neurogenesis and maintenance of the nervous system both peri and postnatally [43, 44]. There is little evidence for the function of *Dclk1* in mature skeletal muscle, however, one prior publication has pointed to a spike in expression of *Dclk1* from activated satellite cell populations of adult murine skeletal muscle [45]. However, no prior evidence exists for the role of *Ostalpha* (a known bile transporter [46]) in skeletal muscle function. Immediate neighbors of *Dclk1* and *Ostalpha* (hub) in module M3 included several biomarkers of skeletal muscle function, cell cycle markers- *Gadd45a* [47], *Cdkn1a* [13] and genes associated with the NMJ- *Emb, Chrnd*, and *Chrna* (Figure 3.4). *Casp3* for which *Dclk1* is a substrate in apoptotic neural cells [48], was also co-expressed in M3.

With little to no evidence for the role of these genes in skeletal muscle, we assessed if either Dclk1 or *Ostalpha* contained putative transcription factor (TF) binding sites for TFs co-expressed in M3. We identified a total of 16 TFs in the module M3 including *Egr1, Egr2, Id1/2/3/4, Myc, Myog, Myod1, Nfkb2, Runx1, Zfp810, Stat3 and Sox4* [49]. Several of these TFs are known to be active in skeletal muscle with an important role in atrophying skeletal muscle [35]. Interestingly, we observed that *Dclk1* contained binding sites for muscle relevant TFs *Myod1, Runx1 and Stat3* [50, 51] on the promoter sequence 1kb upstream while *Ostalpha* contained binding sites for *Myc* [52] and *Myod1*. We also mapped the genes of module M3 onto protein interactions networks to identify known protein interactions between proteins encoded [28]. *Dclk1* displayed several genes associated with skeletal muscle in its 2-step neighborhood including TFs identified above. (Figure 3.5). Interaction partners for *Dclk1* revealed enrichment for processes associated mainly with recovery of skeletal muscle/muscle adaptation (Table S3.2).

Based on these results, we believe that further exploration for the roles of *Ostalpha* and specifically *Dclk1,* in skeletal muscle recovery, particularly within the early period (1-4 weeks) after chemodenervation is warranted.

Grouping of modules such as M3 and M18 together further asserts the "regulatory" role of this group in the maintenance and recovery of muscle trophicity after BoNT-A treatment.

3. Groups 2 and 3

Among the remaining groups, group 2 (M4, M10, M13, M16) and to an extent group 3 (M1, M9, M11, M14, M15) - exhibit higher expression in samples from 4 weeks (Figure 3.3B). Correlation between M10 (group 2) and M1, M11 and M15 (group 3) suggested a possible functional link between them. Interestingly enough, each of these modules was enriched for genes associated with the extracellular matrix (ECM). The ECM is a carbohydrate rich connective tissue surrounding the skeletal muscle providing a structural support to the muscle (fibrillar ECM) protecting each muscle fiber (basal lamina). M10 contained collagens of the basal lamina *Col4a1*, *Col4a2*, that were strongly co-expressed with *Fbn1*, *Loxl2*, and fibrillar collagens such as *Col5a1*, *Col14a1* while M11 contained genes such as *S100a4*, *Loxl1*, *Col1a1*, and *Col1a2* that affect ECM dynamics.

Reduced contractile activity triggers the onset of muscular atrophy, an adaptive response by the muscle with an associated shift in isoform composition of fast muscle towards a more mixed state, including cardiac and slow fiber types in most cases. Analysis of the remaining group 3 modules, showed an overrepresentation of genes associated with sarcomeric contraction. Interestingly, M14 was associated with several non-fast fiber isoforms involved in contraction including *Tpm3*, *Myoz2*, *Myl2*, *Myh7*, *Myh10*, *Tnnc1*, *Tnnt1* and *Tnni1*. *Atp2a2* [53] and *Casq2* [54], two genes specifically involved in Ca^{2+} ion regulation for cardiac contraction were also identified in this module along with stretch responsive *Ankrd2* associated with muscle remodeling [55]. Onset of contraction requires the

mechanical coupling of the dihydropyridine receptors with the ryanodine receptors (RYRs). This module contains an embryonic isoform of the RYR –*Ryr3* that is strongly co-expressed with the sarcomeric genes *Myoz2*, *Myh10*, *Tnnt1*, *Tnnc1* and *Tnni1*. *Ryr3* has been suggested to function as a potent source of voltage independent excitation contraction coupling and known to be expressed primarily in developing/perinatal skeletal muscle [39]. It can be inferred that this increased expression of isoforms normally less expressed in adult skeletal tissue indicates a progression of the muscle towards a "hybrid" state, which is more pronounced by week 4.

No significant groups were identified as being particularly associated with either 12 or 52 weeks. However, we would like to point out that group 1 modules appear to reverse trends and show an increased average expression at later time points (specially 52 weeks) suggesting the recovery of metabolism in the muscle after 1week.

3.4.4 Assessment of phenotype to module correlation

Correlation analysis was performed with an aim to understand the effects of transcriptional changes at the phenotypic level. The following sections provide a discussion of the correlation identified between each phenotype measure and the modules identified in our network.

1. Isometric Contraction

Modules of group 5 correlated positively with isometric contraction before injection (ISO Pre, Figure 3.6A) while being significantly negatively correlated

($p<0.05$) with isometric contraction measured after injection (ISO Post, Figure 3.6A) with opposing patterns exhibited by group 1 genes. This suggests active and opposing roles for genes associated with groups 1 and 5 in the impaired contractility, particularly at 1 week after injection. Of note, module M11 that was enriched for genes associated with several non-fast fiber isoforms of contractile genes was also strongly anti-correlated ($p<0.0001$) with contraction strength after treatment suggesting that expression of these isoforms impedes muscle contraction after chemodenervation (Figure S3.2).

2. Titin

Titin is a fairly large protein (~3MDa) that links myosin to the Z-disk via the M-line and is involved in a variety of functions such as defining the length and organization of the myosin and actin filaments, maintaining the stability of the sarcomere and subsequently controlling the mechanical activity of the muscle [57]. Correlation of two titin isoforms revealed that most modules from group 5 were anti-correlated ($p<0.05$) while group 1 modules were positively correlated with titin (Figure 3. 6A). Though no changes in correlations between the injected and contralateral legs were found, the patterns of correlation suggested a differential influence of genes within these groups on titin turnover [58]. Additionally again, M11 containing the slow muscle isoforms was anti-correlated ($p <0.05$) with isoforms titin1 and titin2 suggesting a debilitating effect of expression of slow isoforms on titin protein dynamics.

3. Myosin heavy chain

Several modules from group 3 were anti-correlated (albeit $p>0.05$) with Myosin IIB while being significantly correlated with Myosin IIA, in injected muscle (Figure 3.6B). The correlation patterns for both Myosin IIA and IIB appear to be reversed, in measurements made on the contralateral leg suggesting a positive change in the dynamics of Myosin IIA turnover over Myosin IIB, particularly at 4 weeks after BoNT-A treatment in the injected muscle. This is consistent with the observed change in MHC quantification (specifically Myosin IIA/IIB) being the largest at 4 weeks after treatment between contralateral and injected muscle in rats treated with BoNT-A (Figure S3.3).

3.5 Conclusion

The physiological response of a system to a stimulus like BoNT-A stems from the underlying transcriptional and epigenetic changes associated with the system. In this study, we analyzed the transcriptional response of skeletal muscle secondary to the insult of BoNT-A at the neuromuscular junction. Utilizing a network theoretic approach, we assessed the response of the skeletal muscle to an injection of BoNT-A over a period of 1 year. Clustering of the co-expression network on a seed set of 2000 genes resulted in modules that were subsequently clustered into 5 groups. Expression patterns revealed dramatic regulation of - metabolism (group 1) and processes associated with muscle trophicity (group 5) in samples from week 1. A reduced energy requirement combined

with the onset of atrophy due to reduced muscle contractility after chemodenervation was reflected in the strong anti-correlation of expression from groups 1 and 5. Strong co-expression of transcription factors *Myod1*, and *Runx1 and* highly expressed genes *Dclk1* and *Ostalpha* with putative binding sites for these transcription factors further emphasized a potential role for these genes in skeletal muscle recovery, warranting further investigations. The increased expression of modules associated with ECM and slow/non skeletal isoforms of sarcomeric proteins at 4 weeks suggested a change to the physical composition of muscle starting at 4 weeks after treatment [14]. The muscle recovers transcriptionally to the pre-treatment state in samples beyond 12 weeks. Correlation of phenotypic data (titin and myosin protein content) with the groups provided insight into the dynamics of the contractile proteins over time. Our analysis of transcriptional response to BoNT-A treatment of skeletal muscle, not only identified mechanisms of response consistent with our previous work, but also identified putative markers, setting the stage for further experiments with implications for clinical use of BoNT-A.

3.6 Acknowledgements

The content of chapter 3 is a modified presentation of the manuscript submitted to BMC medical genomics titled “Botulinum Neurotoxin-A Effects on Skeletal Muscle from Gene Co-Expression Networks: A Time Course Study” by Mukund K, Ward SR, Lieber RL, Subramaniam S.. The dissertation author was the primary author for this material. The authors would also like to thank Dr. Margie Mathewson for performing

the PCR experiments and Dr. Vivianne Minamoto for providing data on the phenotypic assays. Funding for this study was provided by the Department of Veterans Affairs Grant RX000670 (RL), NIH grants R24HD050837 (RL), AR057013 (SRW), Allergan, Inc. (RL) and National Heart, Lung, and Blood Institute Grants HL087375-02 (SS), HL106579 (SS) and HL108735 (SS) and NSF Collaborative Grant STC-0939370 (SS)

3.7 Figures

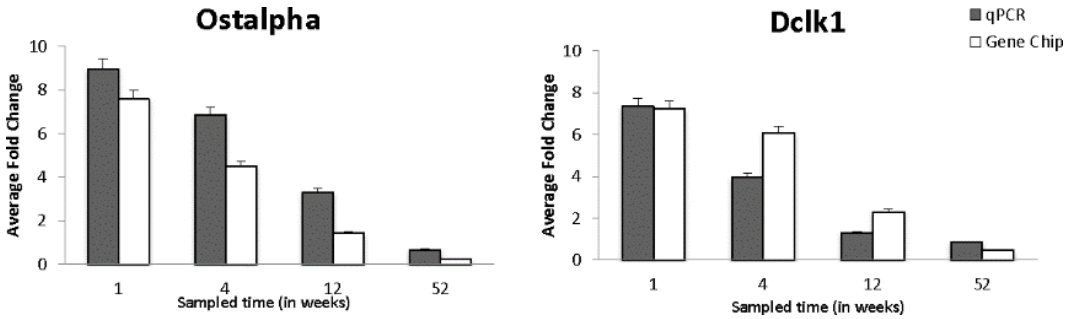

Figure 3.1: Comparison of quantitative real-time PCR data with microarray expression data.
Expression fold changes from GCRMA normalized data is shown in comparison with the real time PCR fold changes measured. The fold change was measured with respect saline injected rat control samples.

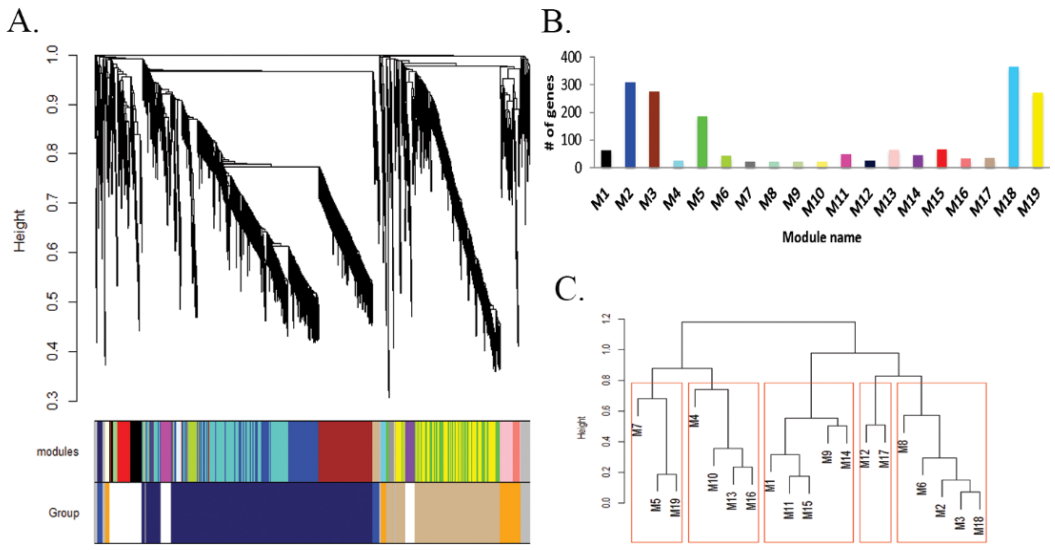

Figure 3.2: Identifying differential co-expressed modules from BoNT-A co-expression network.

A- Hierarchical clustering of 2000 genes from BoNT-A treated samples: The upper panel shows the clustering dendrogram with the middle and lower panels indicating the modules and their corresponding groups (designated by blue, white, dark blue, orange and tan colors) respectively. B- Size distribution of the modules identified where each color in the bar plot corresponds to the "modules" panel of Figure 3.2A and is identified by a letter. C: Hierarchical clustering of co-expressed modules: groups were identified by clustering module eigen genes. This corresponds with the lower panel of Figure 3.2A.

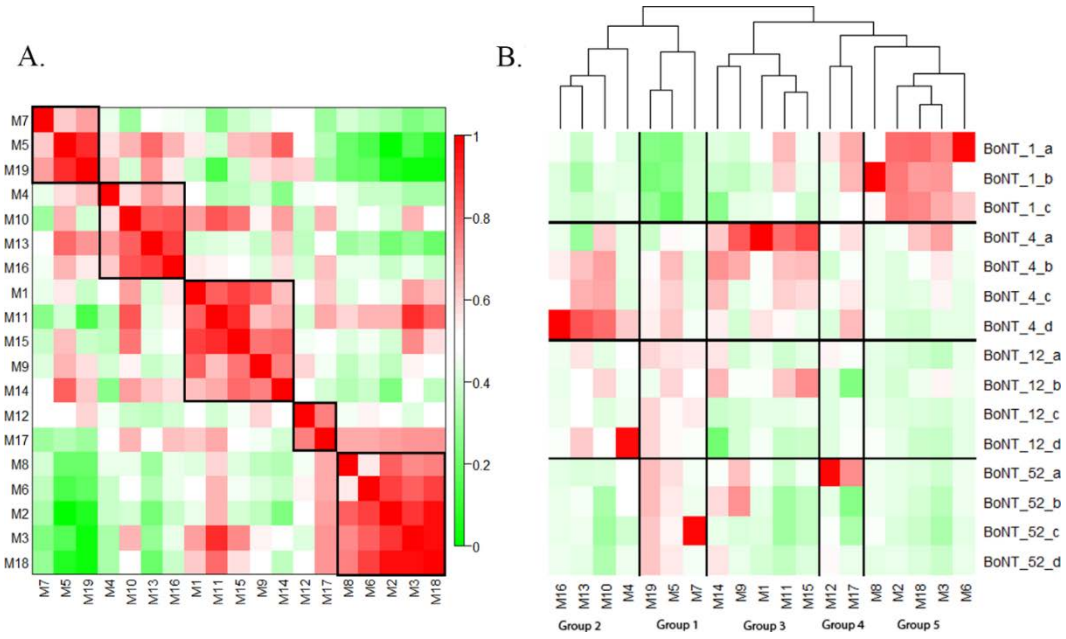

Figure 3.3: Eigen gene correlation and average expression of modules in BoNT-A co-expression network

A- Correlation heatmap of the module eigen genes where each tile in the heatmap represents the scaled correlation between the module eigen genes (colored as per legend with red –correlated and green- anti-correlated). Groups are represented in black squares along the diagonal. High intra-group correlation is observed. B- Average expression heatmap for all modules identified in our network (colored as per legend with red – correlated and green- anti-correlated)

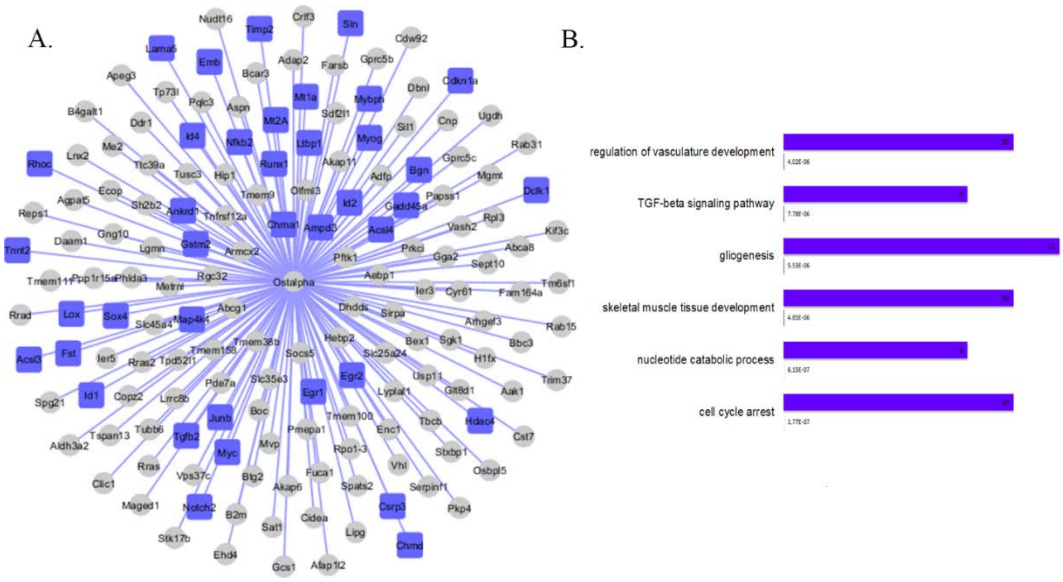

Figure 3.4: The *Ostalpha* co-expression sub-network
A- Immediate neighbors of *Ostalpha* identified as being co-expressed in module M3. Nodes that have been previously referenced in muscle literature have been shown in blue. B- The bar chart represents the relevant gene ontologies categories for this sub network identified through the ClueGO plugin with p-values and number of genes/category indicated.

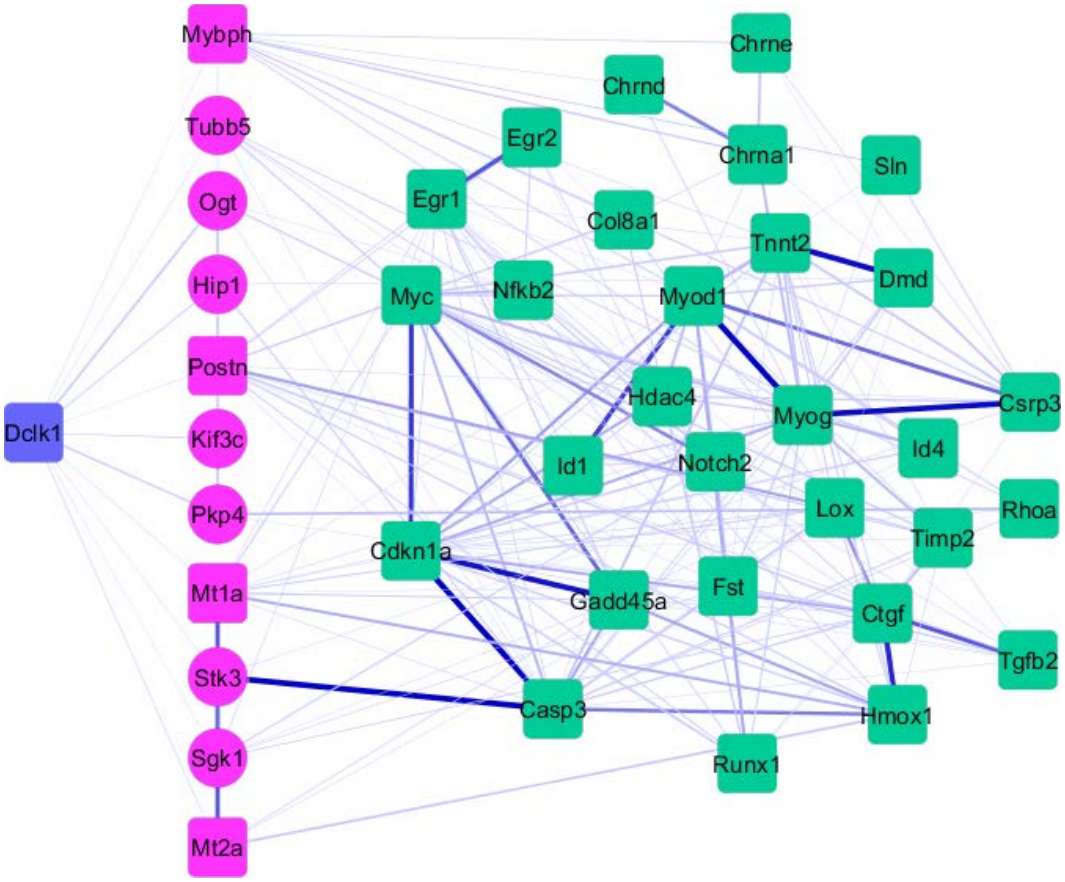

Figure 3.5: Protein interaction map for *Dclk1*
Select protein interaction partners of Dclk1within M3 are shown here. Thickness and color of edges indicate the combined score for the interaction from STRING database. Nodes in pink are the 1-step neighbors for *Dclk1* and nodes in green are its 2-step neighbors. Square nodes indicate genes previously referenced in muscle literature.

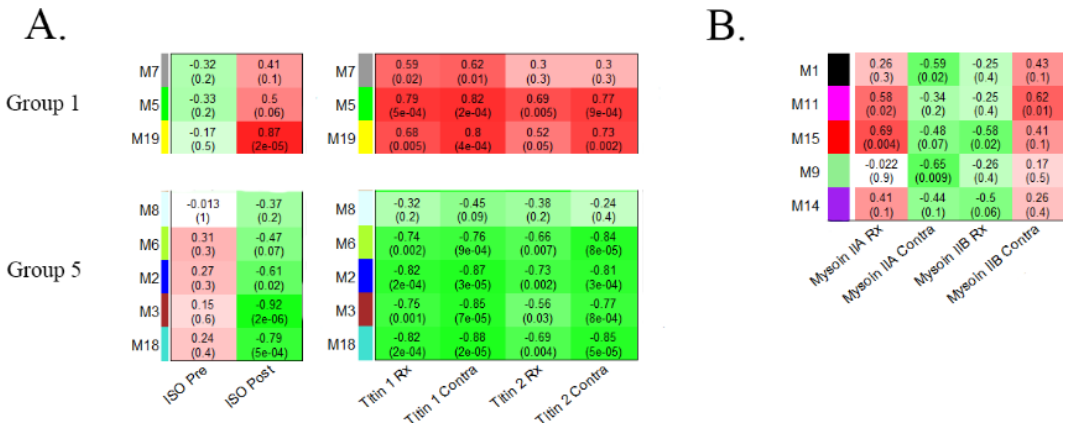

Figure 3.6: Correlation of modules with phenotypic measurements

Each row in the table corresponds to a module, and each column to a phenotypic measurement. The phenotypic measurements include 6A (left) - Isometric contraction strength measured before and after injection on the BoNT-A injected leg (ISO Pre, ISO Post) in two groups- group 1 and group 5. 6A (right) - Titin isoforms 1 and 2, measured in the injected muscle of treated rats (Titin 1Rx, Titin 2 Rx) and their contralateral leg (Titin 1 Contra, Titin 2 Contra) in two groups- group 1 an group 5. 6B- Myosin chains (IIA/IIB) measured in the injected and contralateral muscle of treated rats. Numbers in the table report the correlations of the corresponding module eigen genes with the measure phenotypes with the corresponding p-values printed below in brackets. The table is color coded by correlation red- correlated, green-anti correlated. Intensity of the color represents strength of correlation.

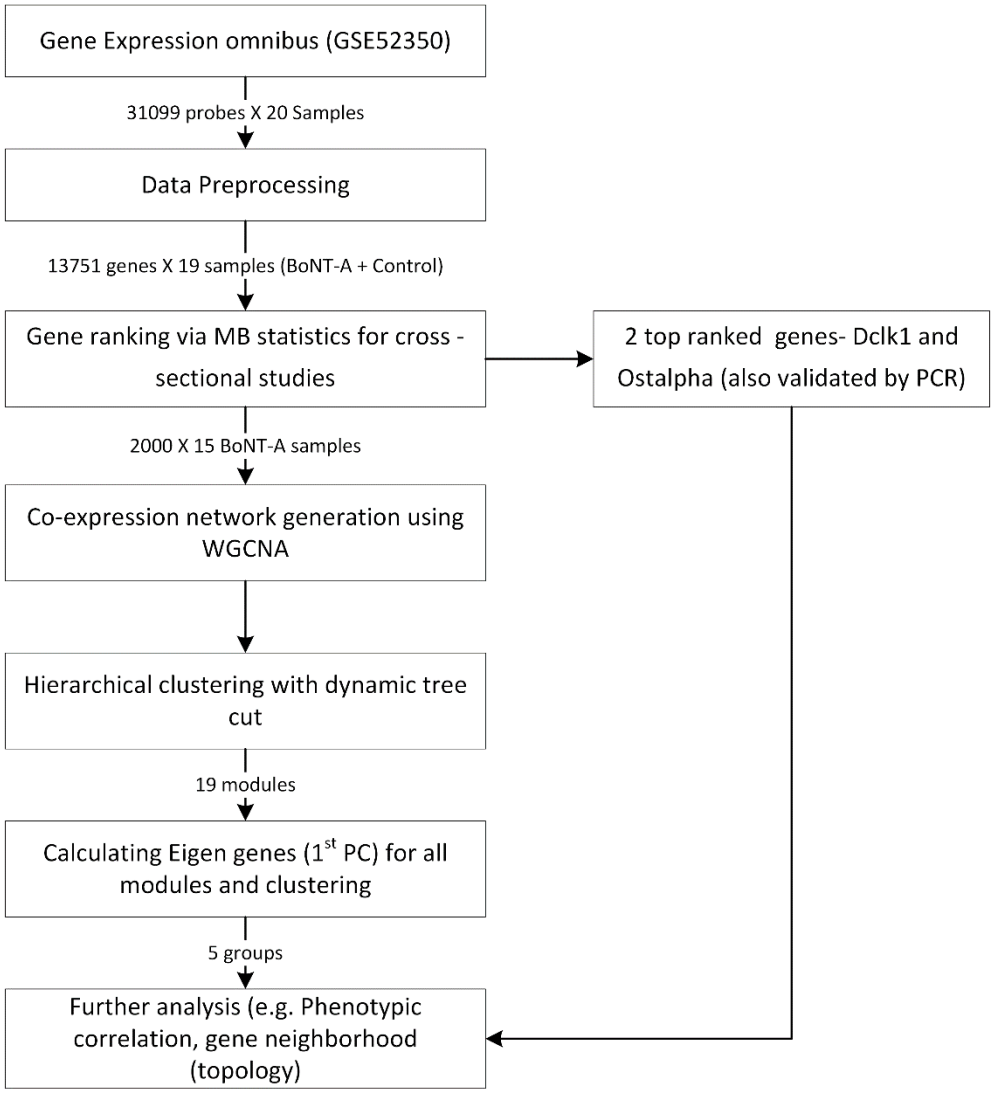

Figure S3.1: A simplified workflow outlining the steps taken in our analysis. Co-expression network generation using WGCNA was performed using their standard workflows.

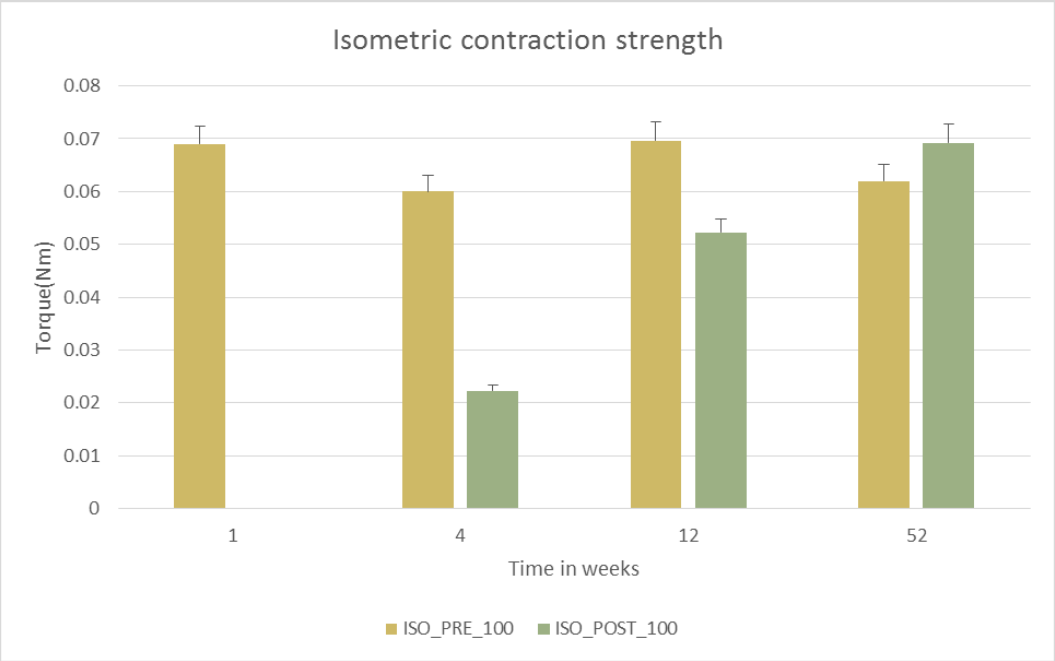

Figure S3.2: Isometric contraction strength before and after BoNT-A
The isometric contraction strength was measured in injected TA of all samples before (ISO_PRE_100) and after (ISO_POST_100) BoNT-A injection across time (1 year period).

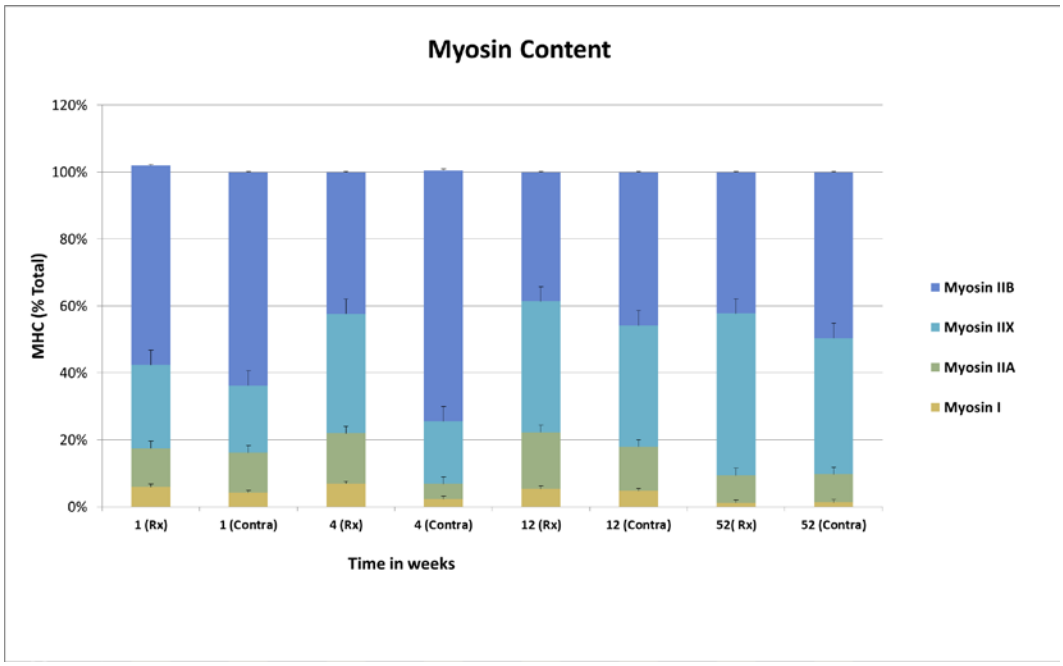

Figure S3.3: Myosin content before and after BoNT-A.
Content of myosin isoforms were measured for samples in both the injected (Rx) and contralateral (contra) TA across time (1 year period).

3.8 References

1. Simpson LL: Identification of the major steps in botulinum toxin action. *Annu Rev Pharmacol Toxicol* 2004, 44:167–193.

2. Dolly JO, Lawrence GW, Meng J, Wang J, Ovsepian SV: Neuro-exocytosis: botulinum toxins as inhibitory probes and versatile therapeutics. *Curr Opin Pharmacol* 2009, 9:326–335.

3. Rosales RL, Chua-Yap AS: Evidence-based systematic review on the efficacy and safety of botulinum toxin-A therapy in post-stroke spasticity. *J Neural Transm* 2008, 115:617–623.

4. Lang AM: Focused review: Botulinum toxin type A therapy in chronic pain disorders. *Arch Phys Med Rehabil* 2003, 84:S69–S73.

5. Foran PG, Mohammed N, Lisk GO, Nagwaney S, Lawrence GW, Johnson E, Smith L, Aoki KR, Dolly JO: Evaluation of the Therapeutic Usefulness of Botulinum Neurotoxin B, C1, E, and F Compared with the Long Lasting Type A BASIS FOR DISTINCT DURATIONS OF INHIBITION OF EXOCYTOSIS IN CENTRAL NEURONS. *J Biol Chem* 2003, 278:1363–1371.

6. Hulst JB, Minamoto VB, Lim MB, Bremner SN, Ward SR, Lieber RL: Systematic test of neurotoxin dose and volume on muscle function in a rat model. *Muscle Nerve* 2014, 49:709–715.

7. Gracies J-M: Physiological effects of botulinum toxin in spasticity. *Mov Disord* 2004, 19:S120–S128.

8. Meunier FA, Schiavo G, Molgó J: Botulinum neurotoxins: from paralysis to recovery of functional neuromuscular transmission. *J Physiol-Paris* 2002, 96:105–113.

9. Lee H-M, Chen J-JJ, Wu Y-N, Wang Y-L, Huang S-C, Piotrkiewicz M: Time course analysis of the effects of botulinum toxin type a on elbow spasticity based on biomechanic and electromyographic parameters. *Arch Phys Med Rehabil* 2008, 89:692–699.

10. Garner CG, Straube A, Witt TN, Gasser T, Oertel WH: Time course of distant effects of local injections of botulinum toxin. *Mov Disord* 1993, 8:33–37.

11. Ma J, Elsaidi GA, Smith TL, Walker FO, Tan KH, Martin E, Koman LA, Smith BP: Time course of recovery of juvenile skeletal muscle after botulinum toxin A injection: an animal model study. *Am J Phys Med Rehabil* 2004, 83:774–780.

12. Ma J, Shen J, Lee CA, Elsaidi GA, Smith TL, Walker FO, Rushing JT, Tan KH, Koman LA, Smith BP: Gene expression of nAChR, SNAP-25 and GAP-43 in skeletal muscles following botulinum toxin A injection: a study in rats. *J Orthop Res* 2005, 23:302–309.

13. Shen J, Ma J, Lee C, Smith BP, Smith TL, Tan KH, Koman LA: How muscles recover from paresis and atrophy after intramuscular injection of botulinum toxin A: study in juvenile rats. *J Orthop Res* 2006, 24:1128–1135.

14. Mukund K, Mathewson M, Minamoto V, Ward SR, Subramaniam S, Lieber RL: Systems Analysis of Transcriptional Data Provides Insights Into Muscle's Biological Response to Botulinum Toxin. *Muscle Nerve* 2014, 50:744–758.

15. de la Fuente A: From "differential expression"to "differential networking"–identification of dysfunctional regulatory networks in diseases. *Trends Genet* 2010, 26:326–333.

16. Nayak RR, Kearns M, Spielman RS, Cheung VG: Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. *Genome Res* 2009, 19:1953–1962.

17. Zhang B, Horvath S: A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005, 4:1128.

18. Maschietto M, Tahira AC, Puga R, Lima L, Mariani D, da Silveira Paulsen B, Belmonte-de-Abreu P, Vieira H, Krepischi AC, Carraro DM: Co-expression network of neural-differentiation genes shows specific pattern in schizophrenia. *BMC Med Genomics* 2015, 8:23.

19. Haas BE, Horvath S, Pietiläinen KH, Cantor RM, Nikkola E, Weissglas-Volkov D, Rissanen A, Civelek M, Cruz-Bautista I, Riba L: Adipose Co-expression networks across Finns and Mexicans identify novel triglyceride-associated genes. *BMC Med Genomics* 2012, 5:61.

20. Miller JA, Oldham MC, Geschwind DH: A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci* 2008, 28:1410.

21. Chang X, Liu S, Yu Y-T, Li Y-X, Li Y-Y: Identifying modules of coexpressed transcript units and their organization of Saccharopolyspora erythraea from time series gene expression profiles. *PLoS One* 2010, 5:e12126.

22. Piec I, Listrat A, Alliot J, Chambon C, Taylor RG, Bechet D: Differential proteome analysis of aging in rat skeletal muscle. *FASEB J* 2005, 19:1143–1145.

23. Goodman MN, Dluz SM, McElaney MA, Belur E, Ruderman NB: Glucose uptake and insulin sensitivity in rat muscle: changes during 3-96 weeks of age. *Am J Physiol-Endocrinol Metab* 1983, 244:E93–E100.

24. Schmittgen TD, Livak KJ: Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* 2008, 3:1101–1108.

25. Ward SR, Minamoto VB, Suzuki KP, Hulst JB, Bremner SN, Lieber RL: Persistent Degradation of Muscle Structure and Function one-year after a Single Botulinum Toxin Injection. *Muscle Nerve* .

26. Chuan Tai Y, Speed TP: On gene ranking using replicated microarray time course data. *Biometrics* 2009, 65:40–51.

27. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008, 9:559.

28. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 2013, 41:D808–D815.

29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003, 13:2498.

30. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J: ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009, 25:1091–1093.

31. Shannon P: MotifDb: An Annotated Collection of Protein-DNA Binding Sequence Motifs. *R Package Version 160* 2014.

32. Basco D, Nicchia GP, D'Alessandro A, Zolla L, Svelto M, Frigeri A: Absence of aquaporin-4 in skeletal muscle alters proteins involved in bioenergetic pathways and calcium handling. *PloS One* 2011, 6:e19225.

33. Hood DA, Irrcher I, Ljubicic V, Joseph A-M: Coordination of metabolic plasticity in skeletal muscle. *J Exp Biol* 2006, 209:2265–2275.

34. Abruzzo PM, di Tullio S, Marchionni C, Belia S, Fanó G, Zampieri S, Carraro U, Kern H, Sgarbi G, Lenaz G: Oxidative stress in the denervated muscle. *Free Radic Res* 2010, 44:563–576.

35. Lecker SH, Jagoe RT, Gilbert A, Gomes M, Baracos V, Bailey J, Price SR, Mitch WE, Goldberg AL: Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression. *FASEB J* 2004, 18:39–51.

36. Muller FL, Song W, Jang YC, Liu Y, Sabia M, Richardson A, Van Remmen H: Denervation-induced skeletal muscle atrophy is associated with increased mitochondrial ROS production. *Am J Physiol-Regul Integr Comp Physiol* 2007, 293:R1159–R1168.

37. Kim N, Stiegler AL, Cameron TO, Hallock PT, Gomez AM, Huang JH, Hubbard SR, Dustin ML, Burden SJ: Lrp4 is a receptor for Agrin and forms a complex with MuSK. *Cell* 2008, 135:334–342.

38. Lain E, Carnejac S, Escher P, Wilson MC, Lømo T, Gajendran N, Brenner HR: A novel role for embigin to promote sprouting of motor nerve terminals at the neuromuscular junction. *J Biol Chem* 2009, 284:8930–8939.

39. Plant PJ, Bain JR, Correa JE, Woo M, Batt J: Absence of caspase-3 protects against denervation-induced skeletal muscle atrophy. *J Appl Physiol* 2009, 107:224–234.

40. Chargé SB, Rudnicki MA: Cellular and molecular regulation of muscle regeneration. *Physiol Rev* 2004, 84:209–238.

41. Wang X, Blagden C, Fan J, Nowak SJ, Taniuchi I, Littman DR, Burden SJ: Runx1 prevents wasting, myofibrillar disorganization, and autophagy of skeletal muscle. *Sci Signal* 2005, 19:1715.

42. Jackman RW, Kandarian SC: The molecular basis of skeletal muscle atrophy. *Am J Physiol-Cell Physiol* 2004, 287:C834–C843.

43. Deuel TA, Liu JS, Corbo JC, Yoo S-Y, Rorke-Adams LB, Walsh CA: Genetic interactions between doublecortin and doublecortin-like kinase in neuronal migration and axon outgrowth. *Neuron* 2006, 49:41–53.

44. Shu T, Tseng H-C, Sapir T, Stern P, Zhou Y, Sanada K, Fischer A, Coquelle FM, Reiner O, Tsai L-H: Doublecortin-like kinase controls neurogenesis by regulating mitotic spindles and M phase progression. *Neuron* 2006, 49:25–39.

45. Pallafacchina G, François S, Regnault B, Czarny B, Dive V, Cumano A, Montarras D, Buckingham M: An adult tissue-specific stem cell in its niche: A gene profiling analysis of< i> in vivo</i> quiescent and activated muscle satellite cells. *Stem Cell Res* 2010, 4:77–91.

46. Ballatori N, Li N, Fang F, Boyer JL, Christian WV, Hammond CL: OST alpha-OST beta: a key membrane transporter of bile acids and conjugated steroids. *Front Biosci J Virtual Libr* 2009, 14:2829.

47. Ebert SM, Dyle MC, Kunkel SD, Bullard SA, Bongers KS, Fox DK, Dierdorff JM, Foster ED, Adams CM: Stress-induced skeletal muscle Gadd45a expression reprograms myonuclei and causes muscle atrophy. J Biol Chem 2012, 287:27290–27301.

48. Schenk GJ, Engels B, Zhang Y-P, Fitzsimons CP, Schouten T, Kruidering M, Ron de Kloet E, Vreugdenhil E: A potential role for calcium/calmodulin-dependent protein kinase-related peptide in neuronal apoptosis: in vivo and in vitro evidence. Eur J Neurosci 2007, 26:3411–3420.

49. Zhang H-M, Chen H, Liu W, Liu H, Gong J, Wang H, Guo A-Y: AnimalTFDB: a comprehensive animal transcription factor database. Nucleic Acids Res 2012, 40:D144–D149.

50. Yang Y, Xu Y, Li W, Wang G, Song Y, Yang G, Han X, Du Z, Sun L, Ma K: STAT3 induces muscle stem cell differentiation by interaction with myoD. Cytokine 2009, 46:137–141.

51. Kami K, Senba E: In vivo activation of STAT3 signaling in satellite cells and myofibers in regenerating rat skeletal muscles. J Histochem Cytochem 2002, 50:1579–1589.

52. Veal EA, Jackson MJ: C-myc is expressed in mouse skeletal muscle nuclei during post-natal maturation. *Int J Biochem Cell Biol* 1998, 30:811–821.

53. Ji Y, Lalli MJ, Babu GJ, Xu Y, Kirkpatrick DL, Liu LH, Chiamvimonvat N, Walsh RA, Shull GE, Periasamy M: Disruption of a single copy of the SERCA2 gene results in altered Ca2+ homeostasis and cardiomyocyte function. *J Biol Chem* 2000, 275:38073–38080.

54. Viatchenko-Karpinski S, Terentyev D, Györke I, Terentyeva R, Volpe P, Priori SG, Napolitano C, Nori A, Williams SC, Györke S: Abnormal calcium signaling and sudden cardiac death associated with mutation of calsequestrin. *Circ Res* 2004, 94:471–477.

55. Kojic S, Radojkovic D, Faulkner G: Muscle ankyrin repeat proteins: their role in striated muscle function in health and disease. Crit Rev Clin Lab Sci 2011, 48:269–294.

56. Ward CW, Rodney GG: Does a lack of RyR3 make mammalian skeletal muscle EC coupling a “spark-less” affair? J Physiol 2008, 586:313–314.

57. Engel A, Franzini-Armstrong C: *Myology: Basic and Clinical*. McGraw-Hill, Medical Pub. Division; 2004.

58. Minamoto VB, Hulst JB, Lim M, Peace WJ, Bremner SN, Ward SR, Lieber RL: Increased efficacy and decreased systemic-effects of botulinum toxin A injection after active or passive muscle manipulation. *Dev Med Child Neurol* 2007, 49:907–914.

Chapter 4- Dysregulated Mechanisms Underlying Duchenne Muscular Dystrophy from Co-expression Network Preservation Analysis

4.1 Abstract

Background: Duchenne Muscular Dystrophy (DMD) is an X-linked recessive disorder with its primary insult on the skeletal muscle. Severe muscle wasting, chronic inflammation and fibrosis characterize dystrophic muscle. Here we identify dysregulated pathways in DMD utilizing a co-expression network approach as described in Weighted Gene Co-expression Network Analysis (WGCNA). Specifically, we utilize WGCNA's "preservation" statistics to identify gene modules that exhibit a weak conservation of network topology within healthy and dystrophic networks. Preservation statistics rank modules based on their topological metrics such as node density, connectivity and separability between networks. **Methods**: Raw data for DMD was downloaded from Gene Expression Omnibus (GSE6011) and suitably preprocessed. Co-expression networks for each condition (healthy and dystrophic) were generated using the WGCNA library in R. Preservation of healthy network edges was evaluated with respect to dystrophic muscle and vice versa using WGCNA. Highly exclusive gene pairs for each of the low preserved modules within both networks were also determined using a specificity measure. **Results**: A total of 11 and 10 co-expressed modules were identified in the networks generated from 13 healthy and 23 dystrophic samples respectively. 5 out of the 11, and 4 out of the 10 modules were identified as exhibiting none-to-weak preservation. Functional enrichment analysis identified that these weakly preserved modules were highly relevant to the condition under study. For instance, weakly preserved dystrophic module D2 exhibited the highest fraction of genes exclusive to DMD. The highly specific gene pairs identified within these modules

were enriched for genes activated in response to wounding and affect the extracellular matrix including several markers such as SPP1, MMP9 and ITGB2. **Conclusion**: The proposed approach allowed us to identify clusters of genes that are non-randomly associated with the disease. Furthermore, highly specific gene pairs pointed to interactions between known markers of disease and identification of putative markers likely associated with disease. The analysis also helped identify putative novel interactions associated with the progression of DMD.

4.2 Introduction

Duchenne muscular dystrophy (DMD), is a lethal form of dystrophinopathy characterized by marked deficiency or absence of subsarcolemmal cytoskeletal protein-dystrophin. Absence of this protein is caused due to frame shift mutations of the dystrophin gene [1]. Dystrophin, part of the dystroglycan complex plays a crucial role in maintaining the integrity of the muscle fiber. Absence of dystrophin causes uneven mechanical force transmissions leading to sarcolemmal ruptures and subsequent atrophy. Clinical manifestations of DMD occur by second year of birth and progressively degrade with time. The first decade of life is marked by developmental delays, and steady decreases in the strength of the limbs and torso with subsequent loss of ambulation. Respiratory and cardiac complications arise by the second decade of life leading to death [2]. Here we utilize a co-expression networks approach to gain insights into molecular interactions dysregulated in dystrophic skeletal muscle with respect to healthy muscle.

Co-expression networks are being increasingly used for deciphering disease mechanisms and providing systems level views of dysregulated pathways [3, 4]. The basic premise of co-expression analysis is that strongly correlated genes are likely to be functionally associated. Weighted Gene Co-expression Network Analysis (WGCNA) is an open source tool that performs co-expression analysis using a network theoretic approach. WGCNA integrates expression differences across samples into a higher order network structure, elucidating relationships among genes based on their co-expression profiles [5, 6].

Here, we propose to utilize a set of statistics implemented in WGCNA, called preservation statistics, to elucidate global differences in mechanisms underlying the early phase of DMD [7]. Traditionally, these statistics have been utilized to identify modules of genes that are topologically preserved between two networks. In contrast to this approach, we propose to utilize these statistics to identify modules that do not exhibit a preservation of topology between networks. This is based on the premise that such modules would represent a cohort of gene interactions that are vastly different between conditions and point to dysfunctional pathways and interactions.

In our current study we utilize a previously published dataset on DMD containing a cohort of healthy and affected individuals (mostly children) - representing the early phase of DMD development [8]. Briefly, we evaluated differential mechanisms between dystrophic and healthy skeletal muscle using the following approach; first, co-expression networks were generated independently for healthy and dystrophic samples; second, clustering each of the co-expression networks resulted in several groups of

biologically relevant genes (modules) for each condition; and finally, preservation of modular topology from one condition was detected with respect to the second condition, allowing us to identify differences in gene connectivity patterns between conditions.

The results of our differential analysis reveal convergent molecular mechanisms consistent with published studies in addition to providing us novel hypothesis on gene interactions associated with the early phase of DMD.

4.3 Methods

4.3.1 Data Acquisition

The raw (.CEL) files for GSE6011 was downloaded from GEO [30]. This data consists of 37 Affymetrix HG-U133A microarrays with 24 juvenile DMD samples, between ages 1.5-61 months, and 13 age matched controls [8].

4.3.2 Data processing

The data set was preprocessed using Bioconductor/R packages *affy* and *WGCNA*. Data was MAS 5.0 normalized using functions in *affy* and any array with an average inter sample correlation <2 SDs (σ) below the mean was removed [9]. This resulted in the removal of a single array - GSM139506.CEL (2.54 σ below mean) from the study. All probes with missing Entrez gene identifiers were excluded from this study, resulting in a data set comprising of the expression values for 11101 probes. Multiple probes were collapsed into a gene based on variance resulting in a final reduced expression data set comprising of 7996 genes and 36 samples. A subset of 4000 most varying genes was used to construct the co-expression networks, in an effort to minimize

computational complexity and eliminate low varying genes that may contribute minimally to the co-expression matrix. The number 4000 was chosen as it represents roughly half the total number of genes (7996 genes) identified after pre-processing. This method of gene list selection is agnostic to their pathophysiological role in the muscle.

4.3.3 Co-expression network generation and modularity detection

The topological overlap measure (TOM) in WGCNA between genes i and j is defined as

$$TOM_{ij} = \frac{\sum_{k=1}^{N} A_{i,k}.A_{k,j} + A_{i,j}}{min(k_i, k_j) + 1 - A_{i,j}}$$

Where A is the weighted adjacency matrix given by $A_{ij} = |cor(x_i, x_j)|^{\beta}$ and $\beta \geq 1$ is the soft thresholding power. TOM takes continuous values between 0 and 1, where 0 for a gene pair indicates no similarity between the genes while 1 indicates a direct link. The soft thresholding power β for each dataset in our study was ascertained as prescribed in the original publication [5].

Co-expression networks from the adjacency matrices of healthy and dystrophic samples were generated using the "TOMsimilarity" function available via WGCNA. Hierarchical clustering on the topological dissimilarity (1-TOM) was performed using the function "flashClust". The tree cut height was dynamically determined using the function "cutreeDynamic" in WGCNA, for identifying modules in each of our networks. Additional files 4 and 5 provide a list of all genes identified in each of the networks (healthy and dystrophic respectively) and their corresponding module assignments.

4.3.4 Preservation of modules

"Module preservation" or preservation statistics implemented in WGCNA allows us to detect the conservation of gene pairs between two networks (test and reference) [7]. Briefly, three types of network based module preservation statistics have been identified by this method, namely

Density based preservation statistics: determine if nodes remain highly connected in the test network. Four independent measures account for this statistic.

Connectivity based preservation statistics determine the extent to which connectivity patterns between nodes in the reference network are similar to the test network. Three independent measures of the network account for this statistic.

Separability based preservation statistics determine if network modules remain distinct from one another in the test network.

Network based statistics employed by WGCNA do not require identification of modules within the test network to ascertain the conservation of reference network modules within the test network. This is in contrast to several existing methods that ascertain module preservation as discussed in the original publication. The authors of the original publication have shown that using this method it is possible to identify sets of preserved co-expression across species.

As these statistics measure distinct aspects of module preservation, two composite measures have been defined

Median rank: A composite measure that is based on observed preservation values and is less dependent on module sizes. It is defined as the mean of median ranks

computed for connectivity and density measures of each module (0.5 ($medianRank_{connectivity}$+$medianRank_{density}$).

$Z_{summary}$: A permutation based composite Z statistic that is used to assess the significance of observed statistics and is defined as the mean of Z scores computed for density and connectivity measures ($0.5(Z_{denstiy} + Z_{connectivity})$). An associated empirical p-value is also calculated by the algorithm.

We utilize median rank to identify module preservation and $Z_{summary}$ to assess significance of module preservation via permutation testing. Based on the number of modules within each of our networks, a median rank of 8 was chosen as a cutoff to detect weak preservation. Permutation was performed 200 times given the computational complexity involved for our network sizes. Based on the thresholds prescribed in [7], modules with a $Z_{summary}$ score >10 indicate preservation, 2 to 10 indicate weak to moderate preservation and <2 indicate no preservation in the permutations.

4.3.5 Network specific gene pairs

Condition specific interactions for a given pair of genes *i* and *j* was defined as [17]:

$$Specificity_{cond1} = \frac{TOM_{ij(cond1)}}{TOM_{ij(cond1)} + TOM_{ij(cond2)}}$$

Where, $TOM_{ij\ (cond)}$ is the normalized TO for the gene pairs *i-j* in the given condition (healthy or disease). We considered gene pairs to be condition specific, if the specificity was >0.95 and were in the top 1% of the gene pairs ranked on TOM similarity in any given module. Number of edges in an undirected network is computed as n (n-

1)/2, where n is the number of nodes. Considering the top 1% allowed us to focus only on the strongest co-expression patterns within the module, rather than noise.

4.3.6 Enrichment Analysis and visualization

The results presented correspond to the top term identified in the highest-ranking cluster (as of this analysis) using the annotation clustering feature available in DAVID [31], with Gene Ontology's Biological process functional annotations. Additional files 6 and 7 provide the top 3 functional annotation clusters identified for each of the modules within the healthy and dystrophic networks respectively. Cytoscape [32] and Bioconductor [33] were utilized for generating the figures in this paper.

4.4 Results and Discussion

4.4.1 Network construction and modularity detection

WGCNA was utilized to construct unsigned weighted co-expression networks from 13 healthy and 23 DMD muscle samples across 4000 most varying genes (see Methods). Briefly, unsigned network adjacency matrices were obtained by raising the Pearson correlation matrices to a power $\beta = 5$ for each condition [5]. The adjacencies were transformed to similarity matrices for subsequent clustering. When represented as networks, each entry of the similarity matrix *ij* corresponds to weight on the edge between genes *ij*. The strength of similarity between two genes depends not only on the correlation but also on their shared network neighborhood [5]. Clustering based on such a similarity allowed for identification of gene groups that were biologically relevant.

Hierarchical clustering of the two weighted networks resulted in eleven modules for the network from healthy samples (N1-N11, see Methods) and ten modules from the network derived from dystrophic samples (D1-D10, see Methods). Figure S4.1 shows the clustering dendrograms and corresponding modules identified in both networks. Genes that did not cluster were excluded from further analysis for the purposes of this study.

4.4.2. Functional characterization of modules identified in healthy and dystrophic networks

Modules identified using WGCNA have been repeatedly shown to be biological relevant to the condition under study [4, 9]. We utilized functional enrichment analysis as a method to assess the functional coherence of modules identified within each of the networks.

1. Characterizing modules of the healthy network

Enrichment of modules from the healthy network revealed several functions routinely associated with healthy skeletal muscle such as striated muscle contraction, energy generation and extracellular matrix organization (Table 4.1).

Skeletal muscle contraction occurs via the coordinated movement of several proteins particularly the actin-myosin complex within the sarcomere, incident upon a changing Ca^{2+} flux. Several genes encoding the sarcomeric proteins such as MYH2, MYH6, MYH7, MYBPC2 TPM1, TPM3, TNNC1/2, TNNI1, TNNT1, MYOM2, MYOZ1, MYOZ2, and MYOZ3 were identified in modules N1 and N8 [10]. The extracellular matrix (ECM) surrounding the skeletal muscle plays an

important role in force transmission and affects the mechanical properties of the skeletal muscle. Several genes associated with the ECM and focal adhesion such as COL4A1, COL4A2, COL5A2, COL6A1, COL6A2, ITGA6, ITGB1, CAV1, CTNNB1, ACTB and LAMN4 were identified in modules N1 and N8 [11]. Muscle contraction and relaxation depend primarily upon energy derived from hydrolysis of adenosine triphosphate (ATP) within the mitochondria. Glycogen/glucose and lipid metabolism serve as major sources of ATP within muscle. Several genes associated with such metabolism were identified within modules N6 and N9 with genes such as NDUFB3, NDUFB5, FABP4, AACS, ADIPOQ, SDHA, SDHB and SCD.

2. Characterizing modules of the dystrophic network

Though the same 4000 genes were used to const ruct the co-expression network in each case, modules cluster differently based on their co-expression. Subsequent enrichment of modules from the dystrophic network revealed functions particularly associated with dystrophic muscle such as response to wounding (Table 4.2).

For instance, module D2 contained several genes associated with wounding and inflammatory response, including several cathepsins and MHC class II antigen processing and presentation genes such as HLA-DPA1, HLA-DMA, HLA-DPB1, HLA-DQA1, HLA-DRA, and HLA-DRB1 [12]. Chronic inflammatory processes are known to initiate fibrosis within dystrophic muscle [13]. Concurrently, ECM adapts dramatically altering both the manifestation and function within dystrophic

muscle. We observe the co-expression of ECM markers affecting fibrosis such as fibronectin (FN1) (a fibroblast marker) and lumican (LUM) - both known to influence collagen expression within this module [14].

Modules D4 and D5 were associated with apoptosis and proteolytic processes within the muscle - more specifically ubiquitin-proteosome system [15] with genes such as genes of the 23s proteasome (PSMA2/3, PSMD9/12, PSME2/4), ubiquitin conjugating enzymes (UBE2B, UBE2D1), ubiquitin ligases (UBE3A, UBE3C), ubiquitin peptidases (USP11, USP6) co-expressed with cullins (CUL4A and CUL5) that serve as scaffolds for ubiquitin ligases.

4.4.3. Identifying functional differences between healthy and dystrophic muscle- a systems approach

The functional annotation clustering results above suggested a mutual exclusivity of certain functions between dystrophic and healthy muscle implying a difference in the topology of connections for genes within these networks.

In order to systematically assess and quantify differential gene co-expression, we performed a "preservation" analysis. This allowed us to identify modules that were fairly unique in terms of their gene co-expression within a given network compared to another. We utilized a method implemented in WGCNA called "*modulePreservation*" [7]. In contrast to the idea of the original paper which aimed at identifying modules preserved between conditions, we aimed to identify modules "weakly preserved" across conditions (see Methods). We hypothesized that modules that were either weakly

preserved or non-preserved in either condition might point to dysregulated pathways in disease that were either acquired or lost with respect to a healthy skeletal muscle.

1. Assessing differential co-expression in healthy muscle with respect to dystrophy

In order to evaluate how the topology of the healthy network differed from the dystrophic network, we computed the preservation (density, connectivity and separability statistics) of modules from the healthy network (reference network) as compared to the dystrophic network (test network). Lower preservation statistics suggested a loss of co-expression structure between these gene pairs in the dystrophic network.

Based on the median preservation score, we identified a total of 5 interesting modules. Two modules N1 and N8 from the healthy network were non-preserved in the dystrophic network while three other modules N2, N3 and particularly N7 exhibited weak preservation (Figure 4.1A). $Z_{summary}$, a permutation statistic (see Methods) defined for assessing significance of the observed preservation also revealed a low preservation of these modules (Table 4.3). Broadly, loss of healthy muscle function and weakened contractility in dystrophic muscle triggers the activation of atrophic pathways leading to severe muscle wasting, changes to the extracellular matrix, fibrosis and necrosis over time [16]. Accordingly, unpreserved modules N1 and N8 were associated with genes necessary for striated muscle contraction, while the weakly preserved modules (N2, N3 and N7) were associated with ECM and cytoskeletal framework of the skeletal muscle.

We utilized a co-expression specificity measure [17] (see Methods) to elucidate co-expressed genes pairs (edges) from these 5 modules. We observed that modules exhibiting none-to-low preservation in the healthy network consistently had a higher fraction of gene pairs exclusive to the healthy network than their preserved counterparts (Table 4.4). For instance, ~35% of the gene pairs considered (599/1738) within N1 were specific to the healthy network (Figure 4.1B). Several of the genes involved are known markers influencing skeletal muscle contraction such as ANKRD2, TNNC2, SMAD3, HSPB1, CRYAB, SDC4, and MYOD1.

It was interesting to observe however that a majority of the genes identified as being part of these interactions were ion- binding as per GO's molecular function ontology (zinc and copper, $p<10^{-4}$). A visual inspection of subset of the exclusive genes pairs identified reveals strong co-expression between several zinc binding genes such as metallothioneins (MT1E/F/H/X), ZNF593, and genes affecting muscle contraction (ANKRD1, MYOD1, SMAD3, HSPB1). Metallothioneins have been postulated to be associated with a host of functions ranging from chaperones for synthesis of metalloproteins, to reservoirs of essential metals (Zn and Cu) in healthy tissue [18]. Specifically, metallothioneins (MTs) exhibit specific redox properties and have been speculated to selectively control release and uptake of Zinc [18]. However, it is interesting to note that MTs were co-expressed with genes affecting muscle contraction only within the healthy network, suggesting a link between zinc homeostasis, and muscle contraction in healthy muscle. The

exclusivity of connections to the healthy network further emphasizes the possibility of an aberration in zinc homeostasis and its effect on contraction in DMD.

2. Assessing differential co-expression in dystrophic muscle with respect to healthy tissue

A similar analysis with dystrophic network as the reference network, allowed us to identify gene pairs that were not conserved in healthy tissue. As proposed earlier, we speculated that identifying non/weakly preserved modules in the dystrophic network could point to gene associations that are *gained* in dystrophy. We identified two modules- D1 and D8 that exhibited no preservation in the healthy network while two other modules D3 and to a greater extent D2 were weakly preserved (Figure 4.2A). The $Z_{summary}$ scores (see Methods) likewise revealed a low preservation of these modules via permutation testing (Table 4.5).

Juvenile dystrophic muscle, in general, exhibits atrophy and is pre-necrotic, with pathways associated with wounding and inflammation being subsequently activated. The functional enrichment identified within these four modules (Table 4.2) corroborated our approach, highlighting functions that are more pronounced in dystrophic muscle compared to healthy tissue.

These modules also exhibited higher specificity of connections to the dystrophic network than their preserved counterparts (Table 4.6). For instance, the highest specificity was observed for module D2 with nearly 45% of its gene pairs as being specific to dystrophy (specificity> 0.95). Interestingly, the dystrophic-specific gene interactions identified in module D2 corresponded with interactions

categorized as a part of the inflammatory and tissue repair/remodeling repertoire of genes, as witnessed in models of skeletal muscle injury, particularly dystrophy (Figure 4.2B).

For instance, expression of SPP1, a multifunctional cytokine (also called early T-cell activation-1 (Eta-1), osteopontin), is linked with macrophage infiltration, resulting in a chronic inflammatory response observed in dystrophic muscle [13,19]. VSIG4, a regulator of T-cell activation expressed mostly in macrophages is strongly co-expressed within D2[19]. Though the exact mechanisms by which skeletal muscle attracts and allows entry of neutrophils and macrophages in dystrophic muscle are not well understood, there is evidence suggesting that ITGB2 is required to control the functional activities of neutrophils and macrophages within muscle [20]. Fibrosis observed in DMD, is largely activated in response to chronic inflammatory processes initiated within dystrophic muscle [13] and broadly refers to the accumulation of excess connective tissue (ECM)[11]. SPP1 which is also expressed in fibrotic lesions is considered a marker for disease severity in DMD [21]. SPP1 is required for differentiation of myofibroblasts [22], an important class of fibroblastic cells required for wound healing, present abundantly within dystrophic muscle. Fibronectin (FN1) serves as a marker for fibroblast activation in muscle [23]

SPP1, in addition to modulating fibrotic responses, promotes cell-cell and cell-matrix adhesions through its interaction with integrins, and CD44 [24]. Mature focal adhesion complexes containing genes such as ACTN1, fail to form in the

absence of SPP1 ($SPP1^{-/-}$) within myofibroblast cultures [22] suggesting similar pathways for adhesion in dystrophy. Interestingly, within this module we also identify MMP9- a matrix metalloproteinase whose increased expression, particularly in the pathology of DMD, is associated with breakdown of cytoskeleton-ECM components leading to sarcolemmal damage and fiber necrosis [25, 26] Additionally, MMP9 is also suggested to act as an inflammatory stimulus for mediating neutrophil and macrophage infiltration within the dystrophic skeletal muscle [27, 28].

SPP1 is subject to extensive posttranslational modification via glycosylation, phosphorylation and sulphation. Specific posttranslational modifications have been associated with altered properties and function of SPP1 [29]. Interestingly, ACP5, a phosphatase required for mineralization of cartilage and bone matrix resorption [28] was recently demonstrated to be responsible for phosphorylation of SPP1 in endometrial tissue. Though no evidence for role of ACP5 or post-translation modification of SPP1 in dystrophic muscle exists, the co-expression of ACP5 with SPP1 suggests a possible role in dystrophy warranting further investigation.

Additionally, SPP1 shows high specificity interactions with certain ECM markers including CTSK, LUM, VCAN and VIM (Figure 4.2B). Though there is no direct evidence for the interaction of these markers with SPP1, the extant understanding of the ECM markers combined with the high specificity of co-

expression observed in our network module suggest possible associations with SPP1 in dystrophic muscle.

Overall, our results indicate that the modules exhibiting low preservation statistics contain several gene pairs that are likely to be associated with the disease progression. Though it is conceivable that not all the genes identified within these less-preserved modules play a role in disease, several high specificity gene pairs identified were noted and hypothesized to play a significant role in pathogenesis of DMD.

4.5 Conclusion

An analysis of modules exhibiting a low preservation between dystrophic and healthy conditions showed that these modules showed a higher specificity among gene pairs pertinent to the condition under study. We illustrated the application of using preservation statistics to detecting modules functionally associated with dysregulated pathways in disease, as exemplified by the inflammatory module D2. This approach enabled identifying putative biomarkers, such as ACP5 identified within module D2, likely to be associated with the disease.

In summary, our method provided a simple approach to identifying differences between conditions, which can be utilized for exploratory analysis of dysregulated pathways in disease using a published set of statistics.

4.6 Acknowledgements

This chapter is a modified presentation of the manuscript published in BMC research notes titled “Dysregulated mechanisms underlying Duchenne muscular dystrophy from co-expression network preservation analysis” by Mukund K, Subramaniam S. The dissertation author was the primary author for this material.

This study was supported by a National Science Foundation Collaborative Grant STC-0939370, a National Science Foundation Grant DBI 0835541, and National Heart, Lung, and Blood Institute Grants HL106579 and HL108735.

4.7 Tables

Table 4.1: Enrichment of modules identified in the healthy network
This table represents the top functional enrichment term from the highest ranking annotation cluster identified for each module of the healthy network. The annotation clusters were ranked and identified using DAVID's annotation clustering tool [31].

Module	#Nodes	Top Term	p value
N1	590	striated muscle contraction	7.59E-07
N2	323	extracellular structure organization	2.90E-08
N3	109	actin cytoskeleton organization	1.50E-02
N4	598	modification-dependent macromolecule catabolic process	2.36E-04
N5	349	intracellular protein transport	7.69E-06
N6	93	generation of precursor metabolites and energy	8.27E-39
N7	215	chromatin assembly or disassembly	8.06E-04
N8	125	muscle organ development	3.74E-03
N9	171	fatty acid metabolic process	3.85E-05
N10	1102	intracellular protein transport	5.12E-05
N11	319	ribosomal small subunit biogenesis	4.28E-05

Table 4.2: Enrichment of modules identified in the dystrophic network
This table represents the top functional enrichment from the highest ranking annotation cluster identified for each module of the diseased network. The annotation clusters were ranked and identified using DAVID's annotation clustering tool [31]

Module	#Nodes	Top Term	p-value
D1	247	cytoskeleton organization	1.89E-04
D2	156	response to wounding	5.09E-07
D3	121	blood vessel development	5.07E-03
D4	377	modification-dependent macromolecule catabolic process	1.56E-03
D5	540	ubiquitin-dependent protein catabolic process	5.24E-04
D6	874	generation of precursor metabolites and energy	7.03E-30
D7	180	muscle system process	2.67E-04
D8	301	RNA splicing	3.52E-10
D9	75	extracellular matrix organization	1.22E-05
D10	1089	positive regulation of ligase activity	9.73E-05

Table 4.3: Permutation based $Z_{summary}$
This table reports the composite measure $Z_{summary}$ and its associated p-value obtained by permuting modules labels in the dystrophic (test) network to assess preservation of modules in the healthy network. Median rank based on the observed statistics are also reported here.

Module	**Size**	**Median rank**	**Z summary**	**log p values (Z summary)**
N1	**590**	**12**	**6.03**	**-13.80**
N2	**323**	**8**	**9.66**	**-29.99**
N3	**109**	**8**	**5.25**	**-8.66**
N4	598	4	23.84	-139.55
N5	349	6	15.97	-71.44
N6	93	2	10.69	-27.93
N7	**215**	**9**	**5.53**	**-9.57**
N8	**125**	**12**	**2.33**	**-2.69**
N9	171	5	10.26	-28.77
N10	1102	1	51.15	-581.53
N11	319	3	18.85	-91.87

Table 4.4: Healthy network specificity
This table represents the fraction of edges identified as being exclusive to the healthy modules as compared to the dystrophic network. #Gene pairs represent the top 1% of the edges calculated as 0.01*(n (n-1)/2)

Module Name	**#Genes (n)**	**#Gene pairs considered**	**Gene pair specificity (%)**
N1	590	1738	34.46
N2	323	520	15.00
N3	109	59	16.95
N4	598	1785	3.08
N5	349	607	2.31
N6	93	43	6.98
N7	215	230	16.09
N8	125	78	16.67
N9	171	145	4.83
N10	1102	6067	0.12
N11	319	507	2.37

Table 4.5: Permutation based $Z_{summary}$

This table reports the composite measure $Z_{summary}$ and its associated p-value obtained by permuting modules labels in the healthy (test) network to assess preservation of modules in the dystrophic network. Median rank based on the observed statistics are also reported here.

Module	**Size**	**Median rank**	**Z summary**	**log p values (Z summary)**
D1	**247**	**11**	**2.50**	**-3.03**
D2	**156**	**9**	**2.35**	**-2.78**
D3	**121**	**8**	**7.36**	**-17.83**
D4	377	2	16.89	-73.04
D5	540	4	20.28	-113.58
D6	874	7	18.91	-120.35
D7	180	5	9.60	-28.01
D8	**301**	**11**	**5.36**	**-12.79**
D9	75	5	4.80	-6.16
D10	1089	1	50.95	-568.57

Table 4.6: Edges exclusive to the dystrophic modules.
This table lists the fraction of edges identified as being exclusive to the dystrophic modules with respect to the healthy network. #Gene pairs represent the top 1% of the edges calculated as 0.01*(n (n-1)/2)

Module Name	**#Genes**	**#Gene pairs**	**Specificity (%)**
D1	247	304	20.39
D2	156	121	45.45
D3	121	73	4.11
D4	377	709	2.12
D5	540	1455	4.60
D6	874	3815	3.96
D7	180	161	3.11
D8	301	452	3.32
D9	75	28	0.00
D10	1089	5924	0.20

4.8 Figures

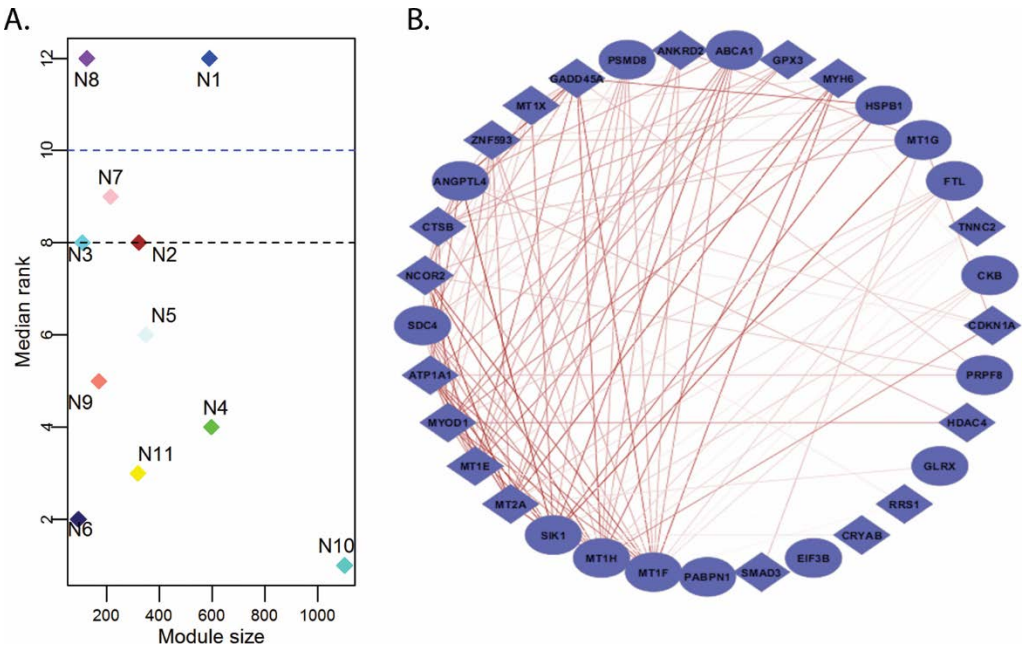

Figure 4.1: Module preservation between test (dystrophic) and reference (healthy) network.
A- Scatter plot identifying the median rank of module preservation between test (dystrophic) and reference (healthy) networks. B- Co-expression between genes identified in module N1- The co-expression patterns for a subset of genes identified within module N1 are shown here. All interactions have a specificity of >0.95. Darker the line, stronger is the strength of co-expression between the gene pairs.

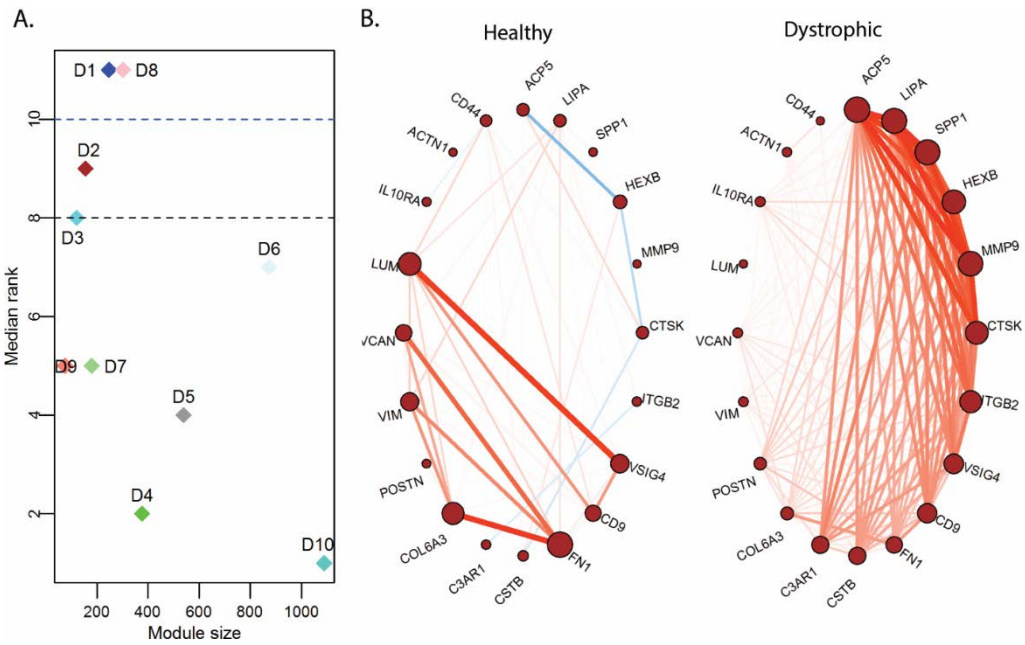

Figure 4.2: Module preservation between test (healthy) and reference (dystrophic) network.

A- Scatter plot identifying the median rank of module preservation between test (healthy) and reference (dystrophic) networks. B- Pearson correlation between a subset of genes identified in module D2 from the dystrophic network- For the same set of genes from module D2, we also identify correlation patterns in the healthy network (left). The size of the node is proportional to the sum of all correlation strengths at the node in the network shown. Red lines indicate positive correlation and blue indicate negative correlation.

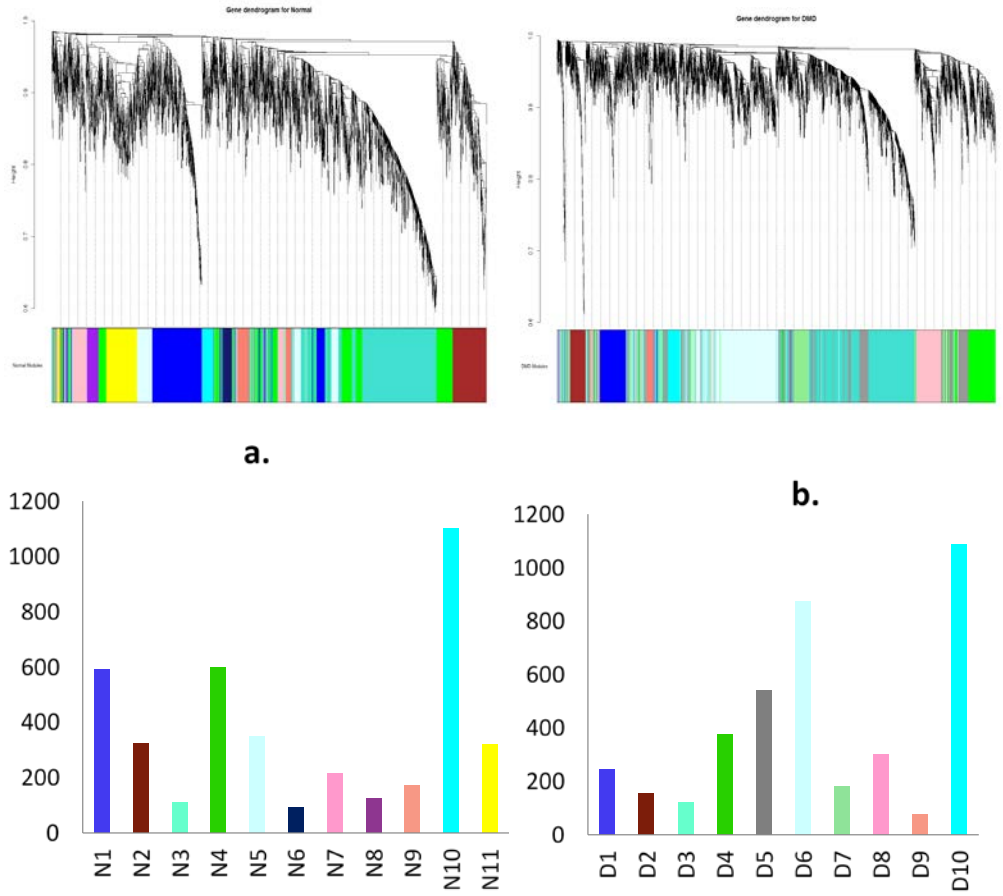

Figure S4.1: Hierarchical clustering results for a) Healthy and b) Dystrophic networks The upper section represents the cluster dendrogram of the differentially expressed genes identified for a) and b). The lower section (bar charts) indicates the modules identified and their respective sizes after hierarchical clustering. Each module is represented by the same colors in the dendrogram for ease of visualization.

4.9 References

1. Nowak KJ, Davies KE: Duchenne muscular dystrophy and dystrophin: pathogenesis and opportunities for treatment. *EMBO Rep* 2004, 5:872–876.

2. Engel A, Franzini-Armstrong C: *Myology: Basic and Clinical*. McGraw-Hill, Medical Pub. Division; 2004.

3. Nayak RR, Kearns M, Spielman RS, Cheung VG: Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. *Genome Res* 2009, 19:1953–1962.

4. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH: Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011, 474:380–384.

5. Zhang B, Horvath S: A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 2005, 4:1128.

6. Langfelder P, Horvath S: WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008, 9:559.

7. Langfelder P, Luo R, Oldham MC, Horvath S: Is my network module preserved and reproducible?. *PLoS Comput Biol* 2011, 7:e1001057.

8. Pescatori M, Broccolini A, Minetti C, Bertini E, Bruno C, D'amico A, Bernardini C, Mirabella M, Silvestri G, Giglio V: Gene expression profiling in the early phases of DMD: a constant molecular signature characterizes DMD muscle from early postnatal life throughout disease progression. *FASEB J* 2007, 21:1210–1226.

9. Miller JA, Oldham MC, Geschwind DH: A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci* 2008, 28:1410.

10. Lieber RL: *Skeletal Muscle Structure, Function, and Plasticity*. Lippincott Williams & Wilkins Baltimore, MD:; 2009.

11. Gillies AR, Lieber RL: Structure and function of the skeletal muscle extracellular matrix. *Muscle Nerve* 2011, 44:318–331.

12. Englund P, Lindroos E, Nennesmo I, Klareskog L, Lundberg IE: Skeletal muscle fibers express major histocompatibility complex class II antigens independently of inflammatory infiltrates in inflammatory myopathies. *Am J Pathol* 2001, 159:1263–1273.

13. Porter JD, Khanna S, Kaminski HJ, Rao JS, Merriam AP, Richmonds CR, Leahy P, Li J, Guo W, Andrade FH: A chronic inflammatory response dominates the skeletal muscle molecular signature in dystrophin-deficient mdx mice. *Hum Mol Genet* 2002, 11:263–272.

14. Hantaï D, Labat-robert J, Grimaud JA, Fardeau M: Fibronectin, laminin, type I, III and IV collagens in Duchenne's muscular dystrophy, congenital muscular dystrophies and congenital myopathies: an immunocytochemical study. *Connect Tissue Res* 1985, 13:273–281.

15. Bodine SC, Latres E, Baumhueter S, Lai VK-M, Nunez L, Clarke BA, Poueymirou WT, Panaro FJ, Na E, Dharmarajan K: Identification of ubiquitin ligases required for skeletal muscle atrophy. *Sci Signal* 2001, 294:1704.

16. Deconinck N, Dan B: Pathophysiology of duchenne muscular dystrophy: current hypotheses. *Pediatr Neurol* 2007, 36:1.

17. Oldham MC, Horvath S, Geschwind DH: Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc Natl Acad Sci* 2006, 103:17973–17978.

18. Palmiter RD: The elusive function of metallothioneins. *Proc Natl Acad Sci* 1998, 95:8428–8430.

19. Forbes SJ, Rosenthal N: Preparing the ground for tissue regeneration: from mechanism to therapy. *Nat Med* 2014, 20:857–869.

20. Marino JS, Tausch BJ, Dearth CL, Manacci MV, McLoughlin TJ, Rakyta SJ, Linsenmayer MP, Pizza FX: β2-Integrins contribute to skeletal muscle hypertrophy in mice. *Am J Physiol-Cell Physiol* 2008, 295:C1026–C1036.

21. Pegoraro E, Hoffman EP, Piva L, Gavassini BF, Cagnin S, Ermani M, Bello L, Soraru G, Pacchioni B, Bonifati MD, others: SPP1 genotype is a determinant of disease severity in Duchenne muscular dystrophy. *Neurology* 2011, 76:219.

22. Lenga Y, Koh A, Perera AS, McCulloch CA, Sodek J, Zohar R: Osteopontin expression is required for myofibroblast differentiation. *Circ Res* 2008, 102:319–327.

23. Gabbiani G: The myofibroblast in wound healing and fibrocontractive diseases. *J Pathol* 2003, 200:500–503.

24. Denhardt DT, Noda M, O'Regan AW, Pavlin D, Berman JS: Osteopontin as a means to cope with environmental insults: regulation of inflammation, tissue remodeling, and cell survival. *J Clin Invest* 2001, 107:1055–1061.

25. Dahiya S, Givvimani S, Bhatnagar S, Qipshidze N, Tyagi SC, Kumar A: Osteopontin-stimulated expression of matrix metalloproteinase-9 causes cardiomyopathy in the mdx model of Duchenne muscular dystrophy. *J Immunol* 2011, 187:2723–2731.

26. Shin J, Tajrishi MM, Ogura Y, Kumar A: Wasting mechanisms in muscular dystrophy. *Int J Biochem Cell Biol* 2013, 45:2266–2279.

27. Luttun A, Lutgens E, Manderveld A, Maris K, Collen D, Carmeliet P, Moons L: Loss of matrix metalloproteinase-9 or matrix metalloproteinase-12 protects apolipoprotein E–deficient mice against atherosclerotic media destruction but differentially affects plaque growth. *Circulation* 2004, 109:1408–1414.

28. Bradley LM, Douglass MF, Chatterjee D, Akira S, Baaten BJ: Matrix metalloprotease 9 mediates neutrophil migration into the airways in response to influenza virus-induced toll-like receptor signaling. *PLoS Pathog* 2012, 8:e1002641.

29. Pagel CN, Wijesinghe DKW, Esfandouni NT, Mackie EJ: Osteopontin, inflammation and myogenesis: influencing regeneration, fibrosis and size of skeletal muscle. *J Cell Commun Signal* 2013:1–9.

30. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M: NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 2013, 41:D991–D995.

31. Dennis Jr G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003, 4:P3.

32. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003, 13:2498.

33. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J: Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004, 5:R80.

Chapter 5- Functional relationships amongst diseases affecting skeletal muscle: a network theoretic approach

5.1 Abstract

The use of systems biology in skeletal muscle research is fairly recent and so far, has largely focused on identifying disease marker genes and functional relationships from a handful of related diseases. Consequently limiting the discovery of new and unknown disease relationships. Here we utilize a quantitative framework to extract functional similarities between 20 diseases affecting the muscle. Using this framework we first identified 7 disease clusters via hierarchical clustering of the disease co-expression network, with permutation testing identifying 20 network associations to be significant ($p<0.05$). We next explored the whole spectrum of muscle diseases by mapping the network onto the human protein-protein interaction network. A common protein signature underlying more than half the muscle diseases was identified using this approach. Functional enrichment analysis of 23 modules identified as part of this signature indicated a statistically non-random dysregulation of muscle bioenergetic pathways and calcium homeostasis. Next, the mechanistic similarities between the 20 significant disease associations was assessed using previously defined "functional modules" of the muscle. In particular, a detailed analysis of the functional similarity between two significantly associated diseases ALS and CP revealed dramatic adaptation of the Ca^{2+} machinery, albeit with a few differences, in both ALS and CP muscle. Lastly, a significant over-representation of drugs targets within the disease clusters highlighted prime therapeutic opportunities.

5.2 Introduction

Human skeletal muscle is a versatile tissue, with science devoting many several decades to its understanding. Muscle, along with its metabolic and regulatory machinery has revealed complexities in composition, structure and function [1]. Precisely coordinated activity of each of its components is essential for normal functioning with factors intrinsic (such as genetic, epigenetic, and developmental) and environmental (such as hormonal, immune), shaping the destiny of muscular health and associated motor activity. A disruption of any component within this complex system of interactions lead to disorders of the muscle, typically characterized by muscle fiber loss, reduced motor output and possibly death. Epidemiological, clinical, and physiological studies have contributed immensely to understanding the pathogenesis and manifestation of individual muscle diseases, revealing information about similarities amongst them [2,3].

Recent advancements in genomic technologies have enabled newer opportunities for extracting disease similarities. Bioinformatics is being increasingly used in muscle research to gain a more comprehensive understanding of muscle dynamics, particularly in biomarker discovery [4–7]. Only a few studies however have capitalized on computational techniques for extracting similarities underlying highly similar muscle diseases, on a much larger scale. . For instance, Blandin et al [8] utilized the yeast-two hybrid (Y2H) system high throughput technology combined with co-expression networks to generate a muscular LGMD-centered interaction network (LGMD- Limb girdle muscular dystrophy) identifying a total of 1018 proteins

connected by 1492 direct binary interactions particularly enriched for the cytoskeleton and extracellular matrix.. In our current study we utilize a quantitative framework to assess relationships between 20 diverse diseases affecting the muscle, categorized broadly into 5 categories, based on their pathological/clinical presentation (Table 5.1, see Methods) [9].

Briefly, the diseases considered here included the emery dreifuss muscular dystrophy (EDMD), limb girdle muscular dystrophy (LGMDs) – caused due to mutation in muscle structural genes; dystrophinopathies caused due to frame-shift (duchenne muscular dystrophy, DMD) or in-frame mutation (becker muscular dystrophy, BMD) of the DMD gene- all characterized by progressive weakening and wasting of skeletal muscle. Mitochondrial myopathies caused due to mutation of mitochondrial DNA (MT-TL1 gene) such as progressive external opthalmoplegia (PEO) and mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS). Other idiopathic metabolic diseases such as acute quadriplegic myopathy (AQM) and chronic fatigue syndrome (CFS), characterized by preferential loss of thick filaments and in-excitability of muscle were also considered for this study. Additionally, host of heterogeneous diseases called inflammatory myopathies (Polymyositis-PM, Dermatomyositis-DM, Inclusion body myopathies IBM and hereditary inclusion body myositis – HIBM) caused due to infiltration of immune cells into muscle were included here. Other diseases that affect muscle composition and function secondary to neurodegeneration such as amyotrophic lateral sclerosis (ALS), spastic paraplegia (SP) and cerebral palsy (CP) were also part of the current study.

In spite of the observable physiological similarities between diseases identified under the same category, diseases exhibit considerable heterogeneity in intensity, etiology, phenotypic manifestation and gene expression. For instance, though muscle in both inclusion body myopathies (HIBM and IBM) exhibits chronic inflammation with visible vacuoles, IBM is mostly idiopathic, while mutations in GNE and MYH2 cause HIBM[10].

Given the heterogeneity associated with these diseases, we utilized techniques from systems biology, to assess the functional relationship amongst them, by integrating various sources of publicly available data - transcriptomic, protein, and drug. Particularly, we first generated an association network using the co-expression between genes based on available transcriptional data and identified 7 disease clusters, with 20 significant inter-disease associations. We quantitatively assessed gene modules regulated across diseases. The gene modules were defined using two approaches- from the human protein-protein interaction network (PPIN) [11] and based on muscle function [12]. Using the protein interaction network we discovered a protein “signature” underlying more than half the diseases considered here mainly associated with deficient bioenergetics, and calcium dysregulation. We also utilized previously published muscle functional modules to compare the mechanistic changes associated with less explored disease associations such as ALS and CP. Overrepresentation of druggable targets with known drugs, was assessed in the disease signature modules. Additionally, drug information was used to extract a smaller subset of drugs uniquely associated with 3 disease clusters further highlighting their importance as therapeutic opportunities. Thus,

this analysis allowed for a synergistic identification of mechanisms shared among muscle diseases, which may or may not share clinical similarities.

5.3 Methods

5.3.1 Data acquisition and processing

The list of diseases available under the Medical Subject Headings (MeSH) terms “neuromuscular”, “musculoskeletal” and “muscular” diseases as a guideline for identifying muscular diseases of interest [13].

All available (RNAseq + microarray platforms) information was downloaded from GEO [14], and was manually surveyed for maximum coverage of muscle diseases in the MeSH headings identified above. A single platform GPL96 (Affymetrix HG-U133A offered the highest coverage of muscle diseases surveyed. Choosing studies from one platform alone (GPL96) allowed us to limit possible noise arising due to platform differences. Additionally, we selected studies that had at least two disease and two control samples. We also eliminated studies on muscle diseases where blood plasma, derived cells or connective tissue was sample tissue. Filtering the data for accuracy, and experimental context using our constraints, we had microarrays for 19 human diseases characterized under the MeSH headings described above. In addition to these 19 diseases we included a study on “Cerebral Palsy”, a movement disorder characterized by contractures of the muscle, though it was not available under MeSH terms, giving a total of 20 disease included in our analysis.

We RMA normalized the gene expression data (disease and control) in each microarray study for which raw .CEL files were available. Studies with series matrix files were downloaded as is. ComBat cross-array normalization was performed on diseases, which had more than one associated GSE (3 diseases -LGMD2A, DMD and JDM, Table 5.1), to remove study artifacts upon combining data [15]. Multiple probes were accounted for using the "collapseRows" function of WGCNA library in R [16]. All probes with missing ENTREZ gene identifiers were excluded from his study. Reduced data sets containing log2-based normalized expression values of 12789 genes for each disease state was subsequently obtained. The normalized expression values were then z transformed thus allowing for comparison of gene expression values across various microarray studies and disease types.

5.3.2 Identifying disease similarity

The T-statistic provides insight into the difference in mean expression of a gene across conditions, and was used as a measure of gene's differential activity (differential gene activity, DGA) in each disease state. The T-test was performed on the z-transformed expression values using Cyber-T's regularized t-test [17]. In contrast to student t-test, Cyber-T implements a Bayesian framework to compute a regularized variance of the measurements associated with each gene under each condition. The partial pearson correlation of the DGA scores was used to quantify disease similarity based on expression profiles. The use of the partial spearman correlation coefficient has been previously shown to be effective in factoring out dependencies such as variation in tissue types and experimental conditions [18].

Disease clusters were identified using hierarchical clustering of the partial pearson correlation between DGA scores. The complete method of clustering, with 1-partial correlation was used as the distance metric for clustering.

5.3.3 Disease-gene based disease overlap

A comprehensive list of genetic factors affecting the 20 muscle diseases was downloaded from various resources such as OMIM[19], PheGenI [20] and ClinVar [21] in addition, to a publicly available database- DisGeNET (that integrates information on gene-disease associations from several public data sources and the literature [22]. The statistical overlap between diseases based on the disease-gene list was identified using the basic hypergeometric model with a null that disease- associated genes are randomly drawn from the space of all genes considered.

5.3.4 Muscle "functional modules" and functional module activity score

A recently published [12] list of functional modules within muscle was appended and utilized in this study (Table 5.3). The genes identified in each "functional module" represent a group of biomarkers significantly associated with a functional subfamily that falls under the broad functional pathway associated with muscle (Table S5.1) as per previously published research [12]. It is acknowledged that several of the genes represented under these categories are multi-functional; yet, we place them in functional modules that are most relevant with respect to muscle. Functional module activity (FMA) score associated with each functional module i in disease k was calculated as the mean of DGA scores, of its component genes, with the scores reflecting the state of the module in a particular disease. Specifically, a negative FMA values

reflects on a general downregulation of the module in disease with respect to controls and vice versa for positive FMA values. Significant functional modules associated with each disease were identified via permutation testing.

5.3.5 Human protein interaction network and protein module activity score

The human PPIN was downloaded from the STRING database (v9.1), containing both direct (physical) and indirect (functionally derived) protein interactions [11]. Limiting the interactions to cutoff of >0.85 ensured only high quality interactions among proteins to be retained in our network. A total of 148030 unique interactions among 10341 proteins were identified at this cutoff. Clustering of the PPIN was performed using MCL clustering (plugin available in Cytoscape), which has been shown to be highly robust in clustering PPIN [23,24]. 278/10341 proteins were unclustered, with the remaining being clustered into 1766 protein modules, with module sizes ranging from 2 to 256 genes/proteins. 1025/1766 modules (total of 8581 proteins) with at least 3 genes/module were considered for further analysis.

Analogous to Suthram et al [18], the protein module activity score was calculated for each protein module *i* in a disease *k* as the mean of DGA scores, of its component genes. In the end, we obtained a vector of protein module activity (PMA_{ik}) scores for each disease representing the activity level of each protein module in each disease state (see Methods). Significance testing identified protein modules with a $p<0.05$ associated with each disease. Threshold for module differential expression was identified as the 90^{th} percentile of |mean PMA| scores from significant modules across

diseases or disease clusters and was considered as the cutoff for identifying significantly differential expression for the given group of diseases.

5.3.6 Significance testing

To assign significance to the observed disease correlations, we created a background distribution of disease correlations expected at random. First we shuffled the disease and the control sample labels and calculated DGA values associated with each gene, for each disease state using the randomized data. We then computed the corresponding disease correlations. Finally, we repeated the whole process 100 times to create a background distribution of disease correlations (test statistic) which was utilized to determine a permutation based p-value by ascertaining the number of the times the permuted statistic exceeded the observed statistic. A similar approach was adopted for generating background distributions for PMA and FMA scores.

5.3.7 Network visualization and functional enrichment

All network visualization was performed using Cytoscape software [25]. All data processing steps and pipelines implemented in this analysis were written in R (v 3.2.2)/ Bioconductor. Enrichment was identified using ClueGO a functional enrichment analysis plugin available via cytoscape [26].

5.3.8 Drug data

The drug gene interaction database (DGIdb) [27] was utilized to identify the list of all expert-curated list of proteins/genes that serve as druggable targets in our study. A list of currently approved drugs from the FDA was obtained from Drugbank [28],

while drugs treating the disease clusters considered here was obtained from Medscape (http://www.medscape.com/).

5.4 Results and Discussion

5.4.1 Clustering muscle diseases based on differential gene activity identifies both well and less characterized disease associations.

Expression data for the 20 muscle diseases was downloaded from GEO, and suitably processed (see Methods). The role of a gene, in each disease state was quantified as the associated T-statistic of z-normalized expression values between case and controls (see methods, Figure 5.1A) and is referred to as the *differential gene activity* (DGA) score. Subsequently, each of the 12789 genes had an associated DGA score for each disease state. The partial correlation of DGA scores was computed to quantify disease similarity. Hierarchical clustering of diseases based on the partial correlations provided 7 clusters, with several of them consistent with known disease similarities (Figure 5.1B, Table 5.1). For instance, disease clusters identified included muscular dystrophies (BMD, DMD, LGMD2I and LGMD2A) mitochondrial disorders (MELAS and PEO) while others are less characterized in muscle literature such cluster containing CP, ALS and AQM. 20 (~10%) of the 190 possible disease interactions were identified as being highly significant ($p<0.05$) through significance testing (see Methods) with several of them not captured via clustering such as between DMD and JDM (Figure 5.1C). A few of the associations marginally missed the significance

threshold and were not captured for instance the association between LGMD2B and DMD, BMD and DMD.

We next ascertained if the observed significant disease correlations (based on DGA scores) shared known genetic associations. List of genes associated with the 20 diseases was compiled from publicly available databases (see Methods). A pair of diseases was considered to significantly share disease-genes if the hypergeometric p-value of overlap was less than 0.05. 26 of 190 possible disease interactions were identified to share a significant genetic basis. 6/26 interactions overlapped with our 20 significant DGA based associations (Table 5.2A), A one-sided fisher's exact test p-value of 0.036 (Table 5.2B), implied that the genetic disease similarity was significantly captured by DGA based correlations in this study. It is to be noted that relatively low albeit significant p-value of fisher's exact test reflects the possibility that certain disease associations did not pass our significance threshold or were not captured using disease-gene lists, as certain diseases are better studied than others. For example, genes associated with ALS (895 genes) where far greater than those associated with AQM (26 genes).

In the following sections we utilize two approaches to exploring muscle diseases using the computed DGA scores as a basis for disease similarity- First, we explore the whole spectrum of muscle diseases by mapping it onto the human PPIN. Next, we mechanistically compare significant pairwise disease associations, by mapping them onto previously published functional modules of the muscle.

5.4.2 Deficient bioenergetics and calcium homeostasis form the common protein signature underlying diseases affecting the muscle

It is often argued that proteins encoded by expressed genes give a more deterministic view of changes occurring in pathophysiology. The protein modules represent a group of strongly interacting proteins identified within a human-PPIN, as detected through modularity detection algorithms [23,24]. With this is mind, we sought to answer if certain protein modules were commonly regulated in the diseases considered and refer to them as the protein signature modules underlying diseases affecting the muscle.

To this end, a catalog of 1025 protein modules were first generated by querying a large-scale human PPIN available through STRING [11] (see Methods) with module sizes varying from 4 – 256 proteins. A protein module activity (PMA) score was calculated for each protein module, defined as the mean of the DGA scores for proteins/genes in each module, for each disease state. Signature protein modules were identified utilizing a two-fold approach- first, all modules, which had their absolute PMA values significantly higher than random ($p<0.05$) in more than half the diseases ($n>10$) were extracted. 45 modules passed this selection criterion. Next, using a threshold for module differential expression (see Methods), 23/45 modules were identify to exceed the threshold and were defined as the underlying protein signature. A complete list of signature modules identified and their top enrichment term is provided in Table S5.2).

The functional enrichment for the 23 signature modules indicated several modules being associated with mitochondrial function (e.g. modules 41, 168, 271, 355, 656), mitochondrial structure (e.g. 537), metabolism (e.g. 426, 632), calcium homeostasis in muscle (e.g. 334), and the extracellular matrix (e.g. 153, 416) (Figure 5.2).

Early research on mitochondrial morphology and its abnormalities in muscle disorders suggested that the widespread occurrence of mitochondrial anomalies observed did not necessarily imply a primary deficiency in efficacy of mitochondrial function, with muscle meeting its energy requirements [29]. However, more recent research has repeatedly suggested deficient bioenergetics underlie the pathology of several muscular and neuromuscular diseases in mammalian models [30–33]. Pathology of neuromuscular diseases such as ALS exhibits mitochondrial dysfunction as a major event in its progression [34,35]. Reduced efficiency in the action of the tricarboxylic acid (TCA) cycle has been also assessed in diseased muscle associated with inflammatory myopathies [36], dystrophy [37] and its well documented effects on mitochondrial diseases such as MELAS, PEO [38]. Mounting evidence suggests that the pathological muscle wasting observed in dystrophies (e.g. DMD) might be due to reduced ATP availability required for maintenance of Ca^{2+} homeostasis and fiber regeneration [39]. Bioenergetic pathway enzymes have recently shown to be relevant biomarkers of muscular and neuromuscular disease progression [40].

Ca^{2+} homeostasis in muscle largely determines its contraction and relaxation properties. This is tightly regulated by the Ca^{2+} signaling apparatus within muscle

comprising of the ryanodine receptors, sarcoplasmic endoplasmic reticulum calcium pumps (SERCA), troponin complex, calsequestrin; in addition to Ca^{2+} binding proteins such as parvalbumin, sarcolipin, phospholamban and calpains. Ca^{2+} signaling and handling molecules have been shown to be altered in various diseases such the strong dysregulation of proteins ATP2A1 (SERCA pump), sarcolipin (SLN, which inhibits SERCA) and calsequestrin (CASQ, restrains Ca^{2+} to the sarcoplasmic reticulum), are strongly dysregulated in ALS and DMD [41]. Likewise, regulation of ASPH (regulator of ryanodine receptors) and SLN have also been observed in muscle from diseases such as CP [42]. Figure 5.3 provides a representative sample of 4/23 signature modules and their enrichment (Ai–iv, and respective enrichment in Bi-iv).

Although existing research on several muscle diseases (such as ALS, DMD, BMD, and CP) has shown varying extents of mitochondrial dysfunction and calcium dysregulation in their pathomechanism, our approach points to widespread, statistically non-random dysregulation of mitochondrial function and calcium homeostasis associated with most muscle diseases including the relatively less characterized disease such as AQM and CFS. Further, the absence of modules associated with structural sarcomeric proteins (myosins, actins, z – disc, dystroglycan) at our significance threshold emphasizes the vital role of muscle bioenergetics and calcium signaling and homeostasis pathways in the pathogenesis of diseases affecting muscle.

5.4.3 Muscle specific mechanistic changes underlying pairwise disease associations

The above approach using protein modules from the human PPIN allowed us to infer a common program of functional changes underlying majority of the muscle

diseases. Subsequently, we sought to elucidate mechanistic similarities among significant pairwise disease associations identified, in a more muscle specific context. Utilizing a modified set of previously published muscle functional networks [12], we defined a set of 23 “functional modules” to capture key biomarkers associated with major pathways/specific functions within muscle, as per extant understanding (Table 5.3, Table S5.1). A functional module activity (FMA) score was calculated for each disease state as the mean of the DGA scores for genes associated with each functional module. The FMA score reflect the collective behavior of genes in each functional module in a given disease state. An associated p-value for each functional module was also associated via permutation testing. Functional mechanisms shared amongst the 20 significant disease-pairs were assessed by identifying functional modules (with $p<0.05$) shared between them. Table 5.4 presents a subset of significant disease-pairs, which shared four, or more significant functional modules shared between them.

For instance, JDM and DMD had 15/23 functional modules (FM) overrepresented in both DMD and JDM ($p<0.05$). The shared functional modules included atrophy, inflammation, ECM and cytoskeleton, all members of the excitation contraction coupling family (FM IDs- 3,4,5), members of contraction (7,8), mitochondrial energy metabolism (11,12,13,14,15), inflammation, and fiber type maintenance (19 and 21). Both DMD and JDM represent myopathies, where the primary insult of the disease is on the skeletal muscle however, JDM is a systemic autoimmune vasculopathy characterized by weakness of proximal muscles and skin rashes with its histopathology showing evidence for necrosis, fiber size variation, and a muscle

degeneration/regeneration phenotype [43]. It has been previously suggested to share many pathologic similarities muscle of children affected with DMD. A comparison of the expression profiles of children with DMD and JDM have revealed similarities in gene cascades involving muscular atrophy, deficits in mitochondrial metabolism and contraction, along with upregulation of extracellular matrix and cytoskeletal cascades [44] consistent with the functional overlap observed using our method.

Finding relevant functional modules consistent with the current understanding of similarities between JDM and DMD further justifies the efficacy of the adopted approach in identifying functional modules affected in more than one diseases state, in a context specific manner. To further elucidate disease associations much less explored, we focused on two diseases ALS and CP and their overlapping functional modules and associated FMA scores.

5.4.4 Calcium dysregulation in patients with ALS and CP

ALS and CP both represent neurological diseases with their primary insult on upper and/or lower motor neurons. While ALS is a neurologically progressive disease, CP is not, with both disorders exhibiting *progressive musculoskeletal* weakness and increased spasticity. While ALS muscle is additionally characterized by denervation atrophy and spasticity, there is distinctive shortening and subsequent weakness of CP muscle [45,46]. We identified 6 functional modules as being significantly dysregulated in both ALS and CP (Table 5.4). The associated FMA scores reflect the state of the functional module in the disease, specifically a negative FMA scores reflects a general downregulation of genes associated with the functional module in the particular disease.

We observed that 5/6 functional modules identified above are similarly regulated in both ALS and CP (Table 5.5), particularly functional modules associated with mitochondrial metabolism (FM IDs 12-15). There is abundant evidence in literature for mitochondrial dysfunction particularly electron transport chain dysregulation and its role in ALS [47,48], albeit in neurons. Our results indicate similar programs of mitochondrial dysregulation to be associated with *muscle* in patients with ALS and CP. Though no detailed studies in muscle exist to corroborate mitochondrial dysfunction in CP, Smith et al [42,49] also show a general downregulation of mitochondrial transcripts. Comparison of the expression values for genes associated with these functional modules showed an R^2 of 0.9.

Interestingly, the FMA scores associated with calcium dynamics/homeostasis reflected differential regulation between ALS and CP (Table 5.5). While, cellular Ca^{2+} dysregulation in ALS has been reported within affected neurons, characterized by endoplasmic reticulum Ca^{2+} depletion and mitochondrial Ca^{2+} overload [50]; dramatic Ca^{2+} dysregulation within muscle from CP patients has also been suggested to occur [42,49]. We observed a few notable differences between ALS and CP functional modules. ATP2A1 and ATP2A2, two-muscle specific, fast fiber sarco (endo) plasmic reticulum Ca^{2+} ATPase (SERCA pumps) were very strongly downregulated in ALS. The energy demanding fast isoform SERCA pumps serve to rapidly replenish the sarcoplasmic reticulum (SR) Ca^{2+} store by pumping the cytosolic calcium back to the SR, bringing about muscle relaxation [51]. The reduced need for regulation of the SERCA pumps is reflected in the downregulation of its two strong regulators- SLN and

PLN. Upregulation of voltage independent neuronal/smooth muscle isoform of the ryanodine receptor (RYR3) further emphasizes leakage of SR Ca^{2+} into the cytosol [52]. Activation of non-skeletal muscle isoforms further emphasizes a shift in the fibers towards slower/more mixed phenotype in ALS. In CP, increase in Ca^{2+} can be inferred by the massive upregulation of PVALB, which selectively binds to free Ca^{2+} to reduce the quantity of free intracellular Ca^{2+} (subsequently, bringing about muscle relaxation). Though no significant changes are observed with respect the SERCA pumps or ryanodine receptors, FKBP1A and PDE4D that prevent channel leaking were significantly downregulated and PLN that controls the Ca^{2+} intake by the SERCA pumps was significantly upregulated in CP. On the other hand, upregulation of ASPH, TRDN and CASQ1 indicate that muscle actively trying to sequester intracellular Ca^{2+} to the stores. Figure 5.4 represents associated fold changes for select genes from the calcium homeostasis functional module.

Taken together this indicates increased cytosolic Ca^{2+} in both diseases, however, in ALS- the Ca^{2+} homeostasis machinery associated with SR appears to be severely challenged by the disease with increased leakage of Ca^{2+} from the intracellular stores and a constrained uptake of Ca^{2+} back into SR. In contrast to this, CP displays a use-dependent decrease in capacity of the SR albeit muscle's efforts to actively recover its Ca^{2+} stores. This dramatic adaptation in both ALS and CP muscle might additionally lead to altered muscle contractile properties and mitochondrial functions.

5.4.5 Drug targets are overrepresented in disease associated protein modules

In spite of the current advancements in understanding of muscular/neuromuscular disease pathophysiology, several of the diseases discussed here are as yet untreatable with high rates of morbidity and mortality with limited therapeutic options. Given the lack of drug/therapeutic availability for muscle diseases, we aimed to identify if therapeutic opportunities could be inferred from our quantitative framework in two contexts.

1. We ascertained if there was an overrepresentation of druggable targets in the common protein signature to support the hypothesis that common drug targets and subsequently drugs can treat a variety of muscle diseases. 47/134 proteins in the disease signature modules belonged to at least one druggable category (as categorized in the drug gene interaction database [27], hypergeometric p-value of overlap <0.05) targeted by at least one drug. 30 of these proteins were further identified as targets for 34 FDA approved drugs, treating a variety of other diseases providing avenues for exploration of therapeutic options (Figure 5.3A, Table S5.3). The absence of appropriate drugs for treating diseases of the muscle, with several of current treatment offering only symptomatic relief; and the presence of drug target over-representation within our network prompted us to explore if there are unique drugs shared by disease clusters.

2. We next hypothesized that protein modules uniquely regulated in each of the disease clusters may be enriched for drug targets uniquely regulating the disease cluster. For instance, the protein modules associated with 3 disease clusters explored are shown in Figure 5.5A (DM/JDM cluster, IBM/PM cluster and DMD/BMD/LGMD cluster),

with the numbers identifying shared and uniquely regulated protein modules identified in the human PPIN, within the 3 disease clusters. FDA approved drugs targeting the proteins in uniquely regulated protein modules were identified. We identified a total of 50 FDA approved drugs to be targeting proteins in the uniquely regulated modules within the IBM/PM cluster, 30 drugs targeting proteins within the DMD/BMD/LGMD cluster and 182 for the DM_JDM cluster (Figure 5.5B, Table S5.4). Several drugs such as methotrexate and cyclosporine – immunosuppresents currently used in symptomatic relief in DM, JDM were identified associated with the relevant cluster. Interestingly, Sirolimus a FDA approved drug for prophylaxis against organ rejection is currently suggested for symptomatic relief in patients with DM. Our analysis showed that Sirolimus targets genes in the IBM/PM clusters suggesting therapeutic opportunities for Sirolimus in IBM/PM (Figure 5.5B). These provide evidence for further exploration of opportunities for drug repurposing for diseases affecting the muscle, as primary or secondary treatment.

5.5 Conclusion

In summary, this study demonstrates the value of an integrated approach in revealing disease relationships and highlights opportunities for therapeutic advancements for treating muscle diseases. Integrating protein information into the diseases similarity network allowed for identification of common dysregulation pathways across a variety of muscle diseases. Likewise, the knowledge of how diseases functionally relate to each other using muscle functional modules provide invaluable

insight into mechanistic differences in muscle, as witnessed through the differences identified in ALS and CP. Incorporating drug data into our quantitative framework allowed us to infer opportunities for exploring drug repurposing as option for treating diseases of the muscle. We aim to further incorporate more gene expression data from GEO and other similar repositories in the future, expanding the set of diseases affecting muscle for the disease-similarity network.

5.6 Acknowledgements

The content of chapter 5 is a modified presentation of material being prepared for submission currently titled “Functional relationships amongst diseases affecting skeletal muscle: a network theoretic approach” by Mukund K, Subramaniam S. The dissertation author is the primary author for this material.

5.7 Tables

Table 5.1**:** Diseases affecting muscle**.**
This table represents the 20 muscle diseases considered in the current study along with the major disease category, current evidence for genetic association and the Gene Expression Omnibus (GEO) accession of the studies corresponding to muscle diseases used in our study. Sarcopenia was not included in any major disease category as it is a age related loss of muscle tissue.

Major disease category	**Disease**	**Extant evidence for genetic association**	**GEO series**
Muscular Dystrophy	Becker muscular dystrophy (BMD)	DMD	GSE3307
	Duchenne muscular dystrophy (DMD)	DMD	GSE3307, GSE6011
	Emery Dreifuss muscular dystrophy (EDMD)	STA (EDMD1), LMNA (EDMD2)	GSE3307
	Facioscapulohumeral muscular dystrophy (FSHD)	FSHMD1A, region 4q35	GSE9397, GSE10760
	Limb-Girdle muscular dystrophies (LGMD) Type 2A	CAPN3	GSE3307, GSE11681
	LGMD Type 2B	DYSF	GSE3307
	LGMD Type 2I	FKRP	GSE3307
Inflammatory Myopathies	Juvenile dermatomyositis (JDM)	Mostly idiopathic with evidence for association with HLA alleles	GSE3112
	Dermatomyositis (DM)		GSE5370
	Polymyositis (PM)		GSE3307, GSE11971
Inclusion body Myopathies	Inclusion body myositis (IBM)		GSE3112
	Hereditary inclusion body myopathy (HIBM)	GNE, MYH2	GSE12648
Age related loss	Sarcopenia	Idiopathic	GSE1428

Table 5.1**:** Diseases affecting muscle**,** continued

Metabolic disorders affecting muscle	Mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes (MELAS)	MT-TL1	GSE1462
	Acute quadriplegic myopathy (AQM)	Idiopathic	GSE1017
	Chronic fatigues syndrome (CFS)	Idiopathic	CSE14577
	Progressive external ophthalmoplegia (PEO)	MT-TL1 and/or POLG, SLC25A4, and C10orf2	GSE1017
Neural diseases affecting muscle	Amyotrophic lateral sclerosis (ALS)	C9orf72,SOD1, TARDBP, FUS, ANG, ALS2, SETX, VAPB (familial); 80-90% are idiopathic (sporadic)	GSE3307
	Hereditary spastic paraplegia (SP)	ATL1, SPG4, SPG20, SPG7	GSE3307
	Cerebral Palsy (CP)	Mostly idiopathic	GSE11686

Table 5.2: Disease association overlap.
A. This table provides associations that overlap between associations calculated based on the DGA scores and associations that share a genetic basis (disease-gene list based). Hypergeometric p-values of the overlap are also presented. B. Contingency table to evaluate the hypothesis that significant disease associations also significantly shared disease genes.

A.

Disease 1	**Disease 2**	**Hypergeometric p-value**
DM	JDM	3.54E-12
DMD	JDM	2.08E-02
DMD	LGMD2A	9.09E-03
BMD	LGMD2I	5.38E-03
MELAS	PEO	6.95E-04
IBM	PM	7.26E-10

B.

	Correlations based on DGA scores			
		Significant	***Not Significant***	***Totals***
Correlations based on known disease genes	***Significant***	6	20	26
	Not Significant	14	150	164
	Totals	20	170	190

Table 5.3: Functional modules in muscle
This table represents the modified set of previously published functional modules associated with muscle. 23 subfamilies were identified as belonging to 14 main functional pathways associated with muscle.

Family	Subfamily	ID	Family	Subfamily	ID
Neuromuscular Junction (NMJ)	Components of NMJ	1	***Mitochondrial energy metabolism***	Mitochondrial electron transporters	13
	Synaptic basal lamina	2		Small molecule transporters	14
Excitation Contraction coupling (ECC)	Ion Channels of post synaptic muscle	3		Members of outer & inner mitochondrial membrane	15
	Ion transporters pumps/exchangers)	4		Associated signaling	16
	Calcium dynamics/homeostasis required for ECC	5	***Hypertrophy***	Hypertrophy	17
Contraction	Sarcomeric thin filament associated	6	***Atrophy***	Atrophy	18
	Sarcomeric thick filament associated	7	***Inflammation***	Inflammation	19
	Sarcomeric z-disc associated	8	***Regulators***	Myogenic and cell cycle regulators	20
Cytoskeleton	Cytoskeleton	9	***Fiber type maintenance***	Fiber type maintenance	21
Extracellular Matrix (ECM)	Components of ECM	10	***Vasculogenesis***	Angiogenic processes	22
Mitochondrial energy metabolism	Glycolytic metabolism	11	***Oxidative stress***	Oxidative stress	23
	Oxidative metabolism	12			

Table 5.4: A representative set of functional modules shared between significant disease pairs

This table represents the functional modules identified as overlapping ($p<0.05$) between the significant diseases associations identified on the left. The overlapping functional modules are identified using the IDs presented in Table 5.3.

Significant disease association	**Overlapping functional modules**
ALS-CP	2,5,12,13,14,15
DMD-BMD	5,7,10,12
DM-JDM	5,6,7,8,9,11,12,13,14,15,19,21
EDMD-FSHD	5,6,11,13
IBM-PM	5,9,10,11,12,15,16,21,23
JDM-DMD	3,4,5,7,8,9,10,11,12,13,14,15,18,19,21

Table 5.5: Overlapping functional modules between ALS and CP
This table provides a list of functional modules that were identified as being significantly shared between two diseases ALS and CP along with their computed functional activity scores.

ID	Functional module	ALS FMA score	CP FMA score
2	Synaptic basal lamina	1.42	1.53
5	Calcium dynamics/homeostasis required for ECC	-2.02	1.72
12	Oxidative metabolism	-2.62	-2.09
13	Mitochondrial electron transporters	-3.17	-2.14
14	Small molecule transporters	-2.99	-1.94
15	Members of outer & inner mitochondrial membrane	-1.18	-1.34

5.8 Figures

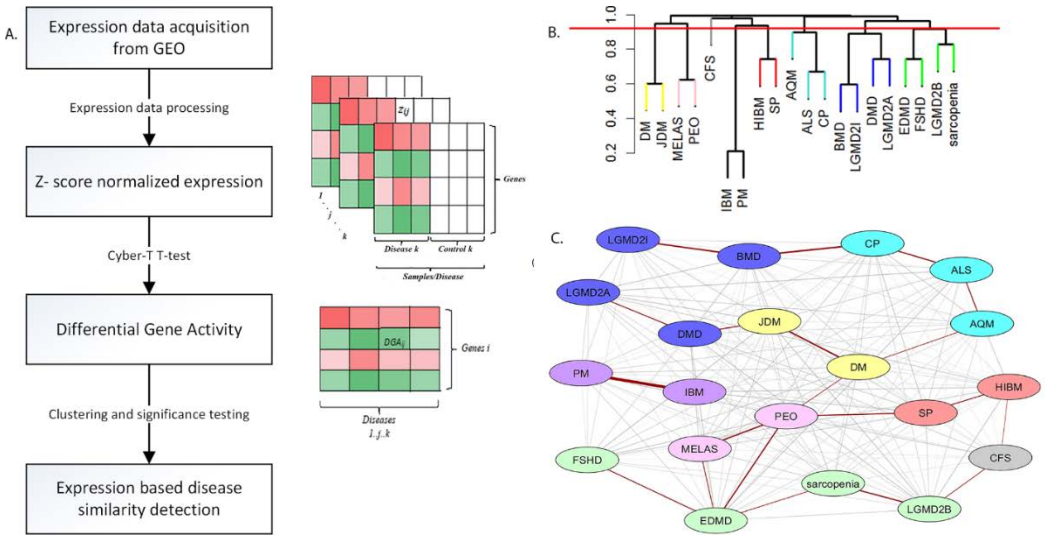

Figure 5.1: Extracting significant disease similarities from 20 diseases affecting muscle A- Shows the workflow involved in calculating the differential gene activity (DGA) score and hierarchical clustering of the scores to extract disease clusters based on DGA B- Shows the hierarchical clustering dendrogram (method- complete) of disease correlations. Tree cut height (red line) corresponds to a p-value of 0.05 and disease clusters identified below this line were identified to be significantly correlated. C- This network represent the 190 possible associations between the 20 diseases. Edges highlighted in red indicate the associations identified as being highly significant through permutation testing. The various node colors indicate the clusters the diseases belong to as identified through hierarchical clustering. The nodes colored in grey were not clustered.

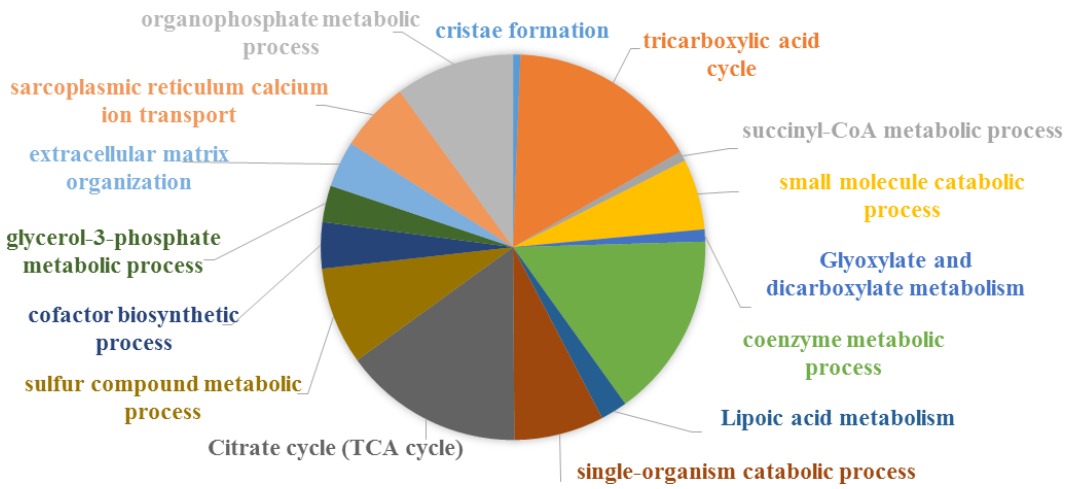

Figure 5.2: Combined functional enrichment of protein signature underlying diseases affecting the muscle

This provides a graphical representation of the top enrichment terms identified in the 23 protein modules (combined) that were commonly identified as underlying a majority of the diseases considered in this study.

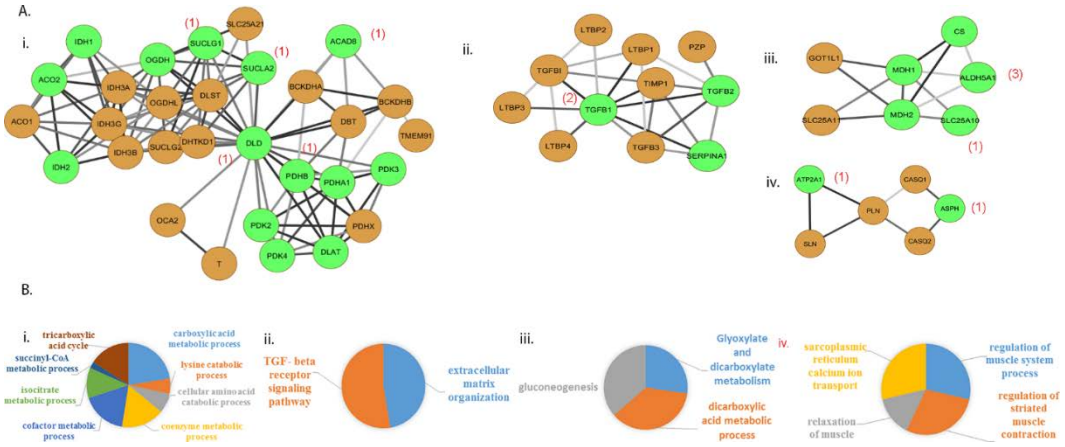

Figure 5.3: Representative set of the protein signature modules underlying diseases affecting the muscle

A (i-iv) - Represent 4/23 protein signature modules are indicated here with brown nodes representing proteins that belong to at least one druggable category as defined in the Drug-gene interaction database (DGIdb). Number shown in braces represent the number of FDA approved drugs targeting the proteins. B (i-iv) represents the corresponding top terms in the functional enrichment as identified using Gene Ontology's -Biological Process category and KEGG pathways in ClueGO- a cystoscope plugin.

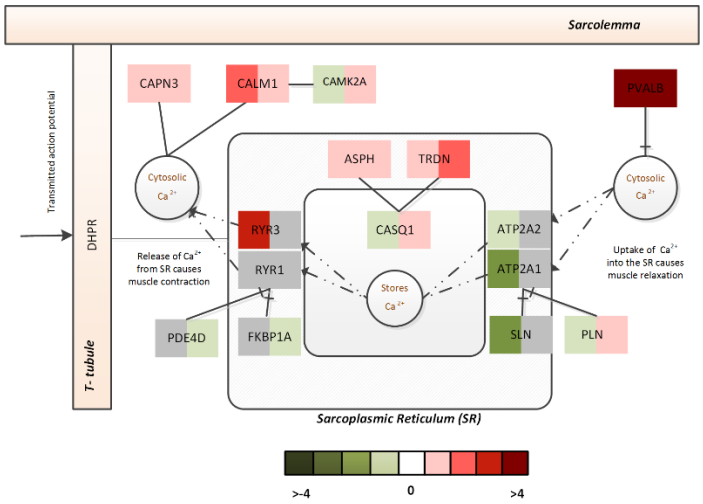

Figure 5.4: The Ca^{2+} homeostasis associated functional module in ALS and CP
This figure captures the fold changes associated with select genes of the Ca^{2+} homeostasis functional module with the left half indicating the fold change associated with ALS and the right half indicating the fold change associated with CP.

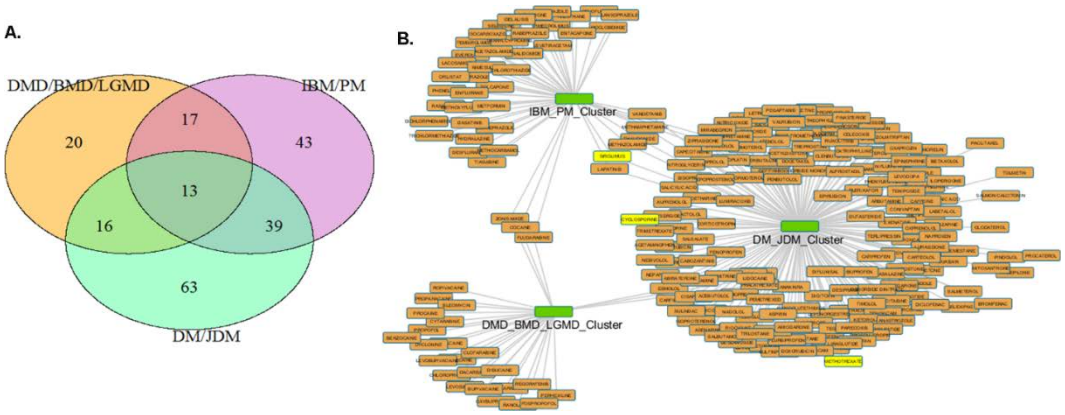

Figure 5.5: Drug disease network for 3 disease clusters
A- Shows the number of protein modules associated with each disease cluster considered e.g 13 protein modules were shared among all clusters, 20 modules were uniquely regulated in the DMD/BMD/LGMD cluster, 43 in the IBM/PM cluster and 63 in the DM/JDM cluster. B. Represents the FDA approved drugs (Table S5.3) associated with the protein modules uniquely regulated in each disease cluster. Nodes in yellow are the drugs currently utilized for treatment in the disease associated with the cluster.

5.9 References

1. Kierszenbaum AL, Tres L. Histology and cell biology: an introduction to pathology. Elsevier Health Sciences; 2015.

2. Askanas V, Engel WK. Inclusion-body myositis: muscle-fiber molecular pathology and possible pathogenic significance of its similarity to Alzheimer's and Parkinson's disease brains. Acta Neuropathol. (Berl.). 2008;116:583–95.

3. Jones DA, Round JM. Skeletal muscle in health and disease: a textbook of muscle physiology. Manchester University Press; 1990.

4. Laaksonen R, Katajamaa M, Paiva H, Sysi-Aho M, Saarinen L, Junni P, Lütjohann D, Smet J, Van Coster R, Seppänen-Laakso T, Lehtimäki T. A systems biology strategy reveals biological pathways and plasma biomarker candidates for potentially toxic statin-induced changes in muscle. PloS One. 2006;1:e97.

5. Dewey FE, Perez MV, Wheeler MT, Watt C, Spin J, Langfelder P, Horvath S, Hannenhalli S, Cappola TP, Ashley EA. Gene coexpression network topology of cardiac development, hypertrophy, and failure. Circ. Cardiovasc. Genet. 2011;4:26–35.

6. Azuaje FJ, Dewey FE, Brutsaert DL, Devaux Y, Ashley EA, Wagner DR. Systems-based approaches to cardiovascular biomarker discovery. Circ. Cardiovasc. Genet. 2012;5:360–7.

7. Gupta S, Kim S-M, Wang Y, Dinasarapu AR, Subramaniam S. Statistical insights into major human muscular diseases. Hum. Mol. Genet. 2014;23:3772–8.

8. Blandin G, Marchand S, Charton K, Danièle N, Gicquel E, Boucheteil J-B, Bentaib A, Barrault L, Stockholm D, Bartoli M, Richard I. A human skeletal muscle interactome centered on proteins involved in muscular dystrophies: LGMD interactome. Skelet Muscle. 2013;3.

9. Engel A, Franzini-Armstrong C. Myology: basic and clinical [Internet]. McGraw-Hill, Medical Pub. Division; 2004. Available from: http://books.google.com/books?id=zP5vQgAACAAJ

10. Tomé FM, Fardeau M. Hereditary inclusion body myopathies. Curr. Opin. Neurol. 1998;11:453–9.

11. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013;41:D808–15.

12. Smith LR, Meyer G, Lieber RL. Systems analysis of biological networks in skeletal muscle function. Wiley Interdiscip. Rev. Syst. Biol. Med. 2013;5:55–71.

13. Lipscomb CE. Medical subject headings (MeSH). Bull. Med. Libr. Assoc. 2000;88:265.

14. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Edgar R. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 2013;41:D991–5.

15. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118–27.

16. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

17. Kayala MA, Baldi P. Cyber-T web server: differential analysis of high-throughput data. Nucleic Acids Res. 2012;40:W553–9.

18. Suthram S, Dudley JT, Chiang AP, Chen R, Hastie TJ, Butte AJ. Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. PLoS Comput. Biol. 2010;6:e1000662.

19. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's online Mendelian inheritance in man (OMIM®). Nucleic Acids Res. 2009;37:D793–6.

20. Ramos EM, Hoffman D, Junkins HA, Maglott D, Phan L, Sherry ST, Feolo M, Hindorff LA. Phenotype–Genotype Integrator (PheGenI): synthesizing genome-wide association study (GWAS) data with existing genomic resources. Eur. J. Hum. Genet. 2014;22:144–7.

21. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, Maglott DR. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42:D980–5.

22. Piñero J, Queralt-Rosinach N, Bravo À, Deu-Pons J, Bauer-Mehren A, Baron M, Sanz F, Furlong Li. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. Database. 2015;2015:bav028.

23. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 2002;30:1575–84.

24. Brohee S, Van Helden J. Evaluation of clustering algorithms for protein-protein interaction networks. BMC Bioinformatics. 2006;7:1.

25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, min N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498.

26. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25:1091–3.

27. Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, Koval J, Das I, Callaway MB, Eldred JM, Miller CA. DGIdb: mining the druggable genome. Nat. Methods. 2013;10:1209–10.

28. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014;42:D1091–7.

29. Stadhouders AM, Sengers RCA. Morphological observations in skeletal muscle from patients with a mitochondrial myopathy. J. Inherit. Metab. Dis. 1987;10:62–80.

30. Wallace DC. Mitochondrial defects in cardiomyopathy and neuromuscular disease. Am. Heart J. 2000;139:s70–85.

31. Pieczenik SR, Neustadt J. Mitochondrial dysfunction and molecular pathways of disease. Exp. Mol. Pathol. 2007;83:84–92.

32. Wallace DC. A mitochondrial bioenergetic etiology of disease. J. Clin. Invest. 2013;123:1405–12.

33. Ramadasan-Nair R, Gayathri N, Mishra S, Sunitha B, Mythri RB, Nalini A, Subbannayya Y, Harsha HC, Kolthur-Seetharam U, Bharath MM. Mitochondrial Alterations and Oxidative Stress in an Acute Transient Mouse Model of Muscle Degeneration- Iimplications for muscular dystrophy and related muscle pathologies. J. Biol. Chem. 2014;289:485–509.

34. Cozzolino M, Carrì MT. Mitochondrial dysfunction in ALS. Prog. Neurobiol. 2012;97:54–66.

35. Dupuis L, Loeffler J-P. Neuromuscular junction destruction during amyotrophic lateral sclerosis: insights from transgenic models. Curr. Opin. Pharmacol. 2009;9:341–6.

36. Coley W, Rayavarapu S, Pandey GS, Sabina RL, Van der Meulen JH, Ampong B, Wortmann RL, Rawat R, Nagaraju K. The molecular basis of skeletal muscle weakness in a mouse model of inflammatory myopathy. Arthritis Rheum. 2012;64:3750–9.

37. Even PC, Decrouy A, Chinet A. Defective regulation of energy metabolism in mdx-mouse skeletal muscles. Biochem. J. 1994;304 (Pt 2):649–54.

38. Wallace DC. Mitochondrial Diseases in Man and Mouse. Science. 1999;283:1482–8.

39. Timpani CA, Hayes A, Rybalka E. Revisiting the dystrophin-ATP connection: How half a century of research still implicates mitochondrial dysfunction in duchenne muscular dystrophy aetiology. Med. Hypotheses. 2015;85:1021–33.

40. Santacatterina F, Chamorro M, Arenas CN de, Navarro C, Martín MA, Cuezva JM, Sánchez-Aragó M. Quantitative analysis of proteins of metabolism by reverse phase protein microarrays identifies potential biomarkers of rare neuromuscular diseases. J. Transl. Med. 2015;13:65.

41. Wang Y, Winters J, Subramaniam S. Functional classification of skeletal muscle networks. II. Applications to pathophysiology. J. Appl. Physiol. 2012;113:1902–20.

42. Smith LR, Lee KS, Ward SR, Chambers HG, Lieber RL. Hamstring contractures in children with spastic cerebral palsy result from a stiffer extracellular matrix and increased in vivo sarcomere length. J. Physiol. 2011;589:2625–39.

43. Peloro TM, Miller OF, Hahn TF, Newman ED. Juvenile dermatomyositis: a retrospective review of a 30-year experience. J. Am. Acad. Dermatol. 2001;45:28–34.

44. Tezak Z, Hoffman EP, Lutz JL, Fedczyna TO, Stephan D, Bremer EG, Krasnoselska-Riz I, Kumar A, Pachman LM. Gene expression profiling in DQA1* 0501+ children with untreated dermatomyositis: a novel model of pathogenesis. J. Immunol. 2002;168:4154–63.

45. Graham HK, Selber P. Musculoskeletal aspects of cerebral palsy. J. BONE Jt. Surg.-Br. Vol.-. 2003;85:157–66.

46. Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, Burrell JR, Zoing MC. Amyotrophic lateral sclerosis. The Lancet. 2011;377:942–55.

47. Borthwick GM, Johnson MA, Ince PG, Shaw PJ, Turnbull DM. Mitochondrial enzyme activity in amyotrophic lateral sclerosis: implications for the role of mitochondria in neuronal cell death. Ann. Neurol. 1999;46:787–90.

48. Crugnola V, Lamperti C, Lucchini V, Ronchi D, Peverelli L, Prelle A, Sciacco M, Bordoni A, Fassone E, Fortunato F, Corti S. Mitochondrial respiratory chain dysfunction in muscle from patients with amyotrophic lateral sclerosis. Arch. Neurol. 2010;67:849–54.

49. Smith LR, Pontén E, Hedström Y, Ward SR, Chambers HG, Subramaniam S, Lieber RL. Novel transcriptional profile in wrist muscles from cerebral palsy patients. BMC Med. Genomics. 2009;2:44.

50. Grosskreutz J, Van Den Bosch L, Keller BU. Calcium dysregulation in amyotrophic lateral sclerosis. Cell Calcium. 2010;47:165–74.

51. Periasamy M, Kalyanasundaram A. SERCA pump isoforms: their role in calcium transport and disease. Muscle Nerve. 2007;35:430–42.

52. Perez CF, López JR, Allen PD. Expression levels of RyR1 and RyR3 control resting free Ca2+ in skeletal muscle. Am. J. Physiol.-Cell Physiol. 2005;288:C640–9.

Chapter 6- Mechanisms underlying ischemic and idiopathic dilated cardiomyopathy utilizing signed differential co-expression network

6.1 Abstract

Differential co-expression networks allow us to identify groups of genes that are differentially co-regulated between conditions. The goal of this study was two-fold- First, we describe an approach to identifying differential co-expression from signed, weighted co-expression network from two condition data. Next, we utilize the defined approach to discern mechanisms underlying ischemic (ICM) and idiopathic dilated cardiomyopathy (IDCM) - two leading causes of human heart failure.

Briefly, the expression correlation matrix for each condition was converted to z-scores using Fisher's r to z transform. A signed scalar matrix of difference (δ) was obtained from the two transformed correlation matrices. A distance metric that takes into account the sign of difference was utilized during hierarchical clustering of the adjacency matrix derived from δ. Dynamic tree cut algorithms were employed to identify differentially co-expressed modules from this network. Signed network parameters including signed node connectivity and signed module centrality were defined in the process.

Functional analysis of differentially co-expressed modules generated using the above approach corresponded well with current understanding of aberrant processes underlying dilated cardiomyopathy (DCM). Topological assessment of the hubs identified within these modules point to strong associations with known markers of disease. These hubs were also enriched for variants in cis-eQTL within the heart left ventricle. Our analysis identified differential co-regulation of targets of SP/KLF family of transcriptional factors in ICM and IDCM. Further research is however needed to

delineate the precise differences in the mechanisms involving SP/KLF family of transcription factors.

6.2 Introduction

Dilated cardiomyopathy (DCM) is a progressive, largely irreversible complex disease of the heart characterized by left ventricular chamber enlargement and systolic dysfunction. Historically, the prevalence of DCM is documented to be 1: 2500 making it one of the leading causes of heart failure, often requiring heart transplantation [1]. However, recent genomic studies have estimated a much higher prevalence of 1:250 making it more widespread in the population than previously believed [2].

In addition to heritable causes of DCM, ischemia, toxic and metabolic insults, immune dysregulation, virus-mediated and valvular defects are known to be causative of DCM [1, 3]. Recent estimates suggest that nearly half the DCM population are non-ischemic in origin, with more than 70% of the non-ischemic cases being categorized as idiopathic (idiopathic DCM, IDCM), that is, no underlying cause can be identified [4]. Though both ischemic cardiomyopathy (ICM) and IDCM are clinically characterized by ventricular dilatation and impaired contractility of left ventricle, ICM exhibits ischemia of the heart (reduced blood flow due to arterial blockage causing a fibrotic phenotype with contractile dysfunction) leading to congestive heart failure. Utilizing clinical, pathological and echocardiographic measurements, studies have shown differential characteristics in treatment and prognosis for ICM and IDCM [5, 6].

Studies in cardiovascular genetics have adopted several expression and systems based approaches to understanding cardiac muscle activity particularly in the context of DCM and its various etiologies [7]. These studies have focused extensively on identifying biomarkers and genetic variants associated with DCM; identifying the effect of their expressivity and penetrance within the DCM population, across various etiologies. Studies have also identified several SNPs contributing to DCM [8,9]; and have implicated 42 genes – that affect structure of cardiac muscle - the sarcomere, the nuclear envelope and the sarcolemma and processes associated with muscle such as adhesion and inflammation [10]. Additionally, a recent systematic meta-analysis has added a putative list of 68 genes with possible roles in DCM [11].

Despite the steadily increasing deluge of high throughput data from studies on DCM, a comprehensive and systematic understanding of the genomic differences between ICM and IDCM is in its early stages. For example, studies that aimed at identifying differences between ICM and IDCM at an “omic” level in humans have largely focused on differential expression [12, 6]. A systems level understanding of the difference in the nucleoplasmic network between ICM and IDCM was recently published [13]. However, an overall understanding of the complex dynamics of interactions in ICM and IDCM is lacking. To this extent, we utilize a module based, differential co-expression approach to further the understanding of ICM and IDCM.

In contrast to differential expression studies that normally provide a static view of transcriptional regulation, co-expression networks allow us to explore the dynamics of gene-gene interactions across conditions [14]. Differential co-expression networks

allow us to identify groups of genes that are differentially co-regulated between conditions [15] where the genes necessarily need not be all differentially expressed. Differential network biology has adopted several approaches to identifying differential co-regulation such as DiffCoEx [16] and DICER [17]. With the caveat that correlation does not imply causation, these approaches have allowed us to infer differential transcriptional regulation by identifying gene modules that change concomitantly across conditions.

In this study, we construct a signed, weighted, differential co-expression network and utilize it in the analysis of transcriptomic data from end stage heart failure due to IDCM and ICM. While most studies have rarely focused on influence of "signed" edges on clustering, studies in social network analysis have shown observable difference in properties of signed and unsigned networks and their subsequent clustering [18].

The goal of our study is two-fold - First, we construct an approach to differential co-expression network analysis using signed co-expression networks. In the process, defining network parameters for signed and weighted networks such as node connectivity and signed module centrality. Next, we utilize the defined approach to identify a cohort of differentially co-expressed modules regulated in ICM and IDCM. The goal of our approach is to not delineate the differences between ICM and IDCM, but to find a set of gene modules that are differentially co-regulated across diseases, which can provide a starting point for further scientific endeavors in identifying the differences in etiology and pathogenesis of these two most common types of DCM.

6.3 Methods

6.3.1 Method for identifying signed differential co-expression from two class studies.

1. Generating the signed differential co-expression network

The correlation matrix using a suitable metric (e.g. spearman rank correlation) is first constructed for all 'n' genes across 'p' samples from each condition (say, condition 1 and 2). The correlation matrix 'ρ' for each condition is transformed using the Fisher's r-z transform [19].

$$\zeta = 0.5[\log(1+\rho)/\log(1-\rho)] \qquad \text{Eq.1}$$

This transformation ensured the correlations to be normally distributed (the variance of ζ was approximately constant for all values ρ) allowing us to compute a scalar matrix of difference in correlation between conditions 1 and 2 as

$$\delta = |\zeta_1| - |\zeta_2| \qquad \text{Eq. 2}$$

Thus, each entry of the difference matrix δ, is a scalar, which captures the difference between condition 1 and 2. When represented as a differential co-expression network, each entry of the matrix δ_{ij}, is the edge weight between the two genes i and j. A positive δ_{ij} represents higher magnitude of correlation in condition 1 while negative δ_{ij} represents a higher magnitude of correlation in condition 2, with the magnitude of δ_{ij} capturing the difference.

2. Clustering the signed differential co-expression network

Drawing on knowledge from analysis of weighted co-expression networks [14], we utilized a published dissimilarity measure called the topological overlap

measure to cluster our difference matrix. We define our adjacency matrix as a scaled δ matrix such that {-1<A<1}.

$$A = \frac{\delta}{\max(\text{abs}(\delta))} \quad \text{Eq.3}$$

The dissimilarity for clustering was used as 0.5(1+A) to account for both positive and negative edge weights.

3. Defining the network parameters

Signed node connectivity- Degree, is a node level parameter and has been previously defined for unweighted, signed networks as the difference of the total number positive edges to the total negative edges incident upon a node. We extend this definition to weighted and signed networks, with edge weights [-1, 1] and define- signed connectivity ($kTotal_s$) for each node u as

$$kTotal_s(u) = conn^+(u) - conn^-(u) \quad \text{Eq. 4}$$

Where, $conn^+(u)$ is the sum all positive edge weights incident at node u from all nodes within network N and $conn^-(u)$ is the sum of all its negative edge weights. Within module connectivity was similarly defined such that

$$kWithin_s(u) = kTotal_s(u) \ \forall u \in module\ M \quad \text{Eq.5}$$

$$kOut_s = kTotal_s(u) - kWithin_s(u) \ \forall u \in module\ M \quad \text{Eq. 6}$$

$kOut_s$ was defined as the connectivity of the node u to all other nodes not within module M. Module hubs were defined as genes with $kWithin_s$ in the upper and lower 5% quantile for each module.

Signed module centrality- We next define a module level parameter, which measures the centrality of a "module". Analogous to the definition of group degree centrality defined for unweighted, unsigned networks [52], we utilize connectivities to define a normalized signed module centrality (SMC) as follows

$$SMC = \frac{|\sum_{u \in M} kOut(u)|}{size(network) - size(M)} \quad \text{Eq. 7}$$

The SMC defined above describes the sum of absolute signed connectivity of connections of nodes within module M to the nodes outside module M, normalized to the size; higher SMC values indicate highly cohesive modules. SMC was subsequently used as the ranking criteria to find the most interesting modules and as the statistic for assessing module statistical significance.

4. Assessing statistical significance of differentially co-expressed modules

Statistical significance of the differentially co-expressed modules was assessed by generating a background distribution of correlation differences expected at random and the observed test statistic. In our case, we chose SMC as the statistic of interest as it represented tightly co-expressed modules. We randomized the labels of class and controls and obtained a randomized, signed and weighted difference network using our approach. Modules in the random networks were identified containing the same set of genes as in the original DCN.

This process was repeated for n permutations. The significance p-value was assessed by computing the SIC (for each module) and performing a permutation test against our background distribution. That is, a module was defined to be significantly differentially co-expressed if the number of times the observed

parameter (SIC) for a given module, exceeded the permuted SIC* from the n randomized runs was less than 0.05.

$$p = 1 - \frac{1}{n}\sum_{1}^{n} count(SMC_i^{*} < SMC)$$

6.3.2 Data acquisition and processing

The abundance of expression data on heart failure on GEO motivated our analysis to be based on publicly available data. Search terms with keywords "cardiomyopathy", "heart failure" and "homo sapiens" rendered 20 results for "expression profiling by array" from the Gene Expression Omnibus. We chose two large expression studies with accession (GSE5406 and GSE57338) where the samples comprised of left- ventricular tissue from patients with ICM, IDCM and non-failing controls. The pipeline followed for processing and obtaining the final reduced data sets for our analysis is outlined in Figure S6.1. Briefly raw .CEL files for GSE7338 were downloaded and preprocessed using library "Oligo" in Bioconductor/R. The pooled data was RMA normalized where present/absent calls were only made on probe sets with "main" annotation. Transcripts were retained if at least 50% of the probe sets were significant and present in at least 75% of the pooled samples (234/313 samples). All probes with missing entrez gene identifiers were excluded from this study. Similarly, the series matrix file of RMA normalized data was extracted for GSE5406 due to the non-availability of raw .CEL files. Multiple probes in each normalized dataset were accounted for using the "collapseRows" function in R's WGCNA library. ComBat [21] was used for cross-array normalization. Outlier samples were detected from each disease type as samples with average inter-sample correlation <2.5SDs below mean.

The following was performed to meaningfully downsize the list of 9464 for co-expression network analysis to limit network size. First, a Cyber-T test [22] identified genes with significant p-value between each condition and non-failing control. Genes in common between the two lists were retained as the set of genes to be extracted. The final resulting data used in our analysis contained 4363 genes across 166 ICM samples, 149 IDCM samples.

6.3.3 Enrichment Analysis, protein network interactions and visualization

Enrichment analysis was performed using ClueGO [23], a cystoscope plug-in using the most recent updates from Gene Ontology's biological process. Cytoscape [24] was utilized for all network visualization. String database [25] was used to extract the protein-protein interactions mentioned in Table 6.1.

6.3.4 Over-represented transcription factors and single tissue eQTLs

Pscan [57], an open source software for identification of over-represented transcription factor binding sites in co-expressed genes was utilized in our analysis. We retained default regions (-450+50) bp of the transcription start site using the Jaspar 2016 database for analysis. Transcription factors were considered to be over-represented if the associated p-value <0.01. All data for significant single tissue eQTL data for heart left ventricle tissue was downloaded via the GTeX consortium webpage [27].

6.4 Results and Discussion

6.4.1 Identifying differentially co-expressed modules between ischemic and idiopathic dilated cardiomyopathy

The approach described in this study attempts to identify modules from two condition data in an unsupervised and unbiased manner (Figure 6.1, detailed description presented in Methods). In contrast to existing differential co-expression methods, our treatment for clustering of signed networks allowed us to assimilate the influence of both positive and negative differences prior to clustering [18].

In the course of this analysis, we applied the above approach to identifying differentially co-expressed modules from ICM and IDCM. The gene expression data for ICM and IDCM samples was suitably processed (see Methods, Figure S6.1), resulting in two gene expression matrices containing 4363 genes across 166 ICM and 149 IDCM samples respectively. The differential co-expression network was generated by first computing the pearson correlation for each condition- ICM and IDCM (Eq. 1, see Methods). The correlation identified for each condition type was transformed to Fisher's z-score using equation 1 (see Methods). The difference in co-expression was computed as in Eq. 2 as follows

$$\delta = |\zeta_{ICM}| - |\zeta_{IDCM}|$$

The magnitude of correlation differences between the two conditions is captured via the scalar matrix δ, with a positive sign indicating a higher magnitude of correlation in ICM and vice versa for a negative sign. The adjacency matrix (A) was obtained my normalizing δ to the maximum. Hierarchical clustering on the dissimilarity matrix

(0.5(1+A)) with dynamic cut height [28] resulted in 28 modules with module sizes varying between 44 and 477 genes.

6.4.2 Gene module prioritization

Modularity detection within biological networks and its interpretation are often subjective, where not all modules extracted from a network are biologically meaningful. In order to prioritize the 28 modules identified for further analysis, we utilized a measure of module network topology- signed module centrality (SMC, Eq. 7). SMC provides an estimate of module cohesiveness (see Methods), allowing us to identify strongly co-expressed gene modules. Using permutation-based significance testing 11/28 modules had a higher than random SMC value ($p<0.05$) and were considered significant. Assessing the significance of the differentially co-expressed modules based on SMC ($p<0.05$) while utilizing signed edges allowed us to identify modules, which may be functionally more relevant in one condition relative to another.

For all modules, we assessed node level connectivity parameters $kTotal_s$ (Eq. 4) and $kWithin_s$ (Eq. 5). Connectivity provides insight into the behavior of individual nodes within a signed differential co-expression network. It is important to realize that our connectivity is defined on a difference network, that is, the edge weights offer insight to differences in correlation between two networks and hence, a node with a high negative connectivity indicates stronger correlations in condition 2 between itself and its neighbors and vice versa for positive connectivity nodes (from Eq.2). Interestingly, all of the eleven modules also exhibited a negative size normalized kWithin (sum of

$kWithin_s$ of all genes within the module normalized to module size) suggesting a higher fraction of genes with strong negative edges.

6.4.3 Differentially co-expressed modules recapitulate aspects of disease pathogenesis

The basic premise of all co-expression studies is that co-expressed genes are likely co-regulated and changes in correlation structure captured via differential co-expression might be attributable to changes in the roles of certain genes within pathways across conditions. For instance, differential activation patterns for transcription factors and their downstream targets in healthy and disease tissue. We hypothesized that a functional enrichment analysis of the significant modules would reveal functional pathways that might be differentially regulated in patients with ICM and IDCM.

Functional enrichment, in general, identifies over represented categories of annotated genes in given gene set (modules) and is not specific to the condition under study. It is essential to interpret enrichment in the context of study. Table 6.1 outlines the top 5/11 modules, which exhibited functional categories highly relevant to cardiac muscle. For instance, module M21 was particularly enriched for processes associated with immune response such as lymphocyte activation, response to cytokine stimulus. Abnormalities in cellular and humoral immunomodulation have been previously recognized as being associated with IDCM; however, whether they are a cause or a consequence of the disease are uncertain. IDCM has been hypothesized to be an immune disorder; the presence of HLA class II antigens (such HLA-A, HLA-DRA, HLA-DRB4) [29].

Additionally, this module also contained genes associated with T-cell activation and aggregation such as CD3E/3G/48/74, SOD2, RHOH, RELB, NOD2, and EGR1.This activation might be a result of either interaction with foreign antigens or altered expression of HLA antigens in myocytes. A direct correlation between an increased activation of T-cells and the severity of both IDCM and ICM has been previously reported. However, the specificity of activation has been found to be different between the two pathologies [30].

Likewise, M4 was enriched for genes that are associated with control of the arterial blood pressure in the heart, particularly via angiotensin, and included genes such as ACE and CMA1. Angiotensin, a peptide hormone, part of the renin-angiotensin system is necessary for vasoconstriction and a subsequent positive regulation of blood pressure. It has been shown that CMA1, rather than angiotensin converting enzyme (ACE), is largely responsible for converting angiotensin I to the vasoactive peptide angiotensin II in the heart and vasculature [31]. DD gene variant of ACE has been suggested to contribute to the pathogenesis both types of cardiomyopathy [32]. It is interesting to observe that these genes were co-expressed with other cardiac markers known to be influenced by the angiotensin system such as ECE1 expressed in the endocardium and myocardium, serving as a potential regulatory site for the production of the active vasoconstrictor peptide Endothelin [33]; the cardiac ion channels KCNA5 and transporters ATP1A1 implicated in atrial fibrillation [34,35]; major insulin substrate of the heart IRS2 [36] implicated in an association between insulin resistance, hypertension, and cardiovascular disease and SERPINE-1/PAI-1 [37]. Other genes in

this module included known (cardiac) muscle markers such as CDKN1A, RYR3, FKBP5, SPP1 and MYC, whose presence in the module may point to the effect of the renin-angiotensin system on these markers in IDCM and ICM.

We further extracted the PPI [25] for the top 5 modules, as proteins interactions are often argued to offer a more deterministic view of the physical interactions occurring at a molecular level. Interestingly, all 5 modules (Table 6.1) exhibited a higher than random number of PPI ($p<<0.05$). This suggests a statistically non-random interaction between genes within each of the differentially co-expressed module at a protein level.

Thus, the relevant enrichment in the modules identified further supports the hypothesis that these co-expressed modules represent a group of genes that might be differentially co-regulated in ICM and IDCM.

6.4.4 Module hubs are strongly associated with known markers of heart failure due to DCM

Module hubs represent genes with high connectivity to other nodes within a module. In our study, we defined genes that have a *kWithin* in the upper and lower 5% quantile, as hubs (see Methods). 152 hub genes were identified in the 11 significant modules, with 100 of them being in the top 5 modules. The general notion is that the hub nodes represent genes that have significant roles to play in physiology. We hypothesized that since our modules capture differential co-expression between ICM and IDCM, the hubs might capture a gene set that is not particularly enriched for known markers associated with DCM, but a gene set that might be differentially co-regulated between ICM and IDCM.

A recent systematic meta-analysis of extant literature identified a list of 110 markers genes associated with heart failure after DCM (and not etiology of DCM per say) [11]. Of the 32 genes from this list present in our entire gene list, we identified 6 (~19%) genes in the top 5 modules including genes such as MYH6, TMEM43 and TXNRD2. Consistent with our hypothesis, we observed that these known markers were not identified as networks hubs but were strongly co-expressed with network hubs (Figure 6.2A). Several of the hubs identified among the significant modules represent genes that have been identified to be directly correlated with DCM and subsequent heart failure such as VCAM1 [38], IRS2 [36], SERPINE1/PAI-1 [37], BMP6 and ACVRL1 [39] (Table S6.1) .

We next asked the question if the hub genes contained variants in cis-eQTL within cardiac muscle, suggesting a possible role for these variants in dilated cardiomyopathy (ICM or IDCM). Expression quantitative trait loci (eQTL) analysis allows us to investigate how gene expression levels are affected by DNA variants with possible roles in disease. Utilizing publicly available data from heart left ventricle, we identified all genes in eQTL with known SNPs. Specifically, using the GTeX consortium's single tissue (heart left ventricle) significant eQTL data, we extracted SNPs that are in cis-eQTL with genes in the top 5 modules. We identified 15% of the hubs to have at least 1 SNP in eQTL ($p<<0.01$) in contrast to 13% of the genes that were not hubs having SNPs in eQTL (Figure 6.2B, Table 6.2). For instance, 3/8 hub genes identified in M18 (MTRR, CLEC4A and RPS27L) have SNPs that are in significant cis-

eQTL, within the heart left-ventricle. Of the total 7/179 significant cis-eQTL SNPs identified are either synonymous, missense or within the UTR (Table S6.2).

6.4.5 Over-representation of transcription factor in differentially co-expressed modules

Next, we hypothesized that identifying the overrepresentation of transcription factors associated with the top 5 modules would point to a discrete set of transcription factors differentially influencing the pathogenesis of heart failure in DCM. -450+50bp of the transcription start site was used to identify over-represented transcription factors with an associated p-value of <0.01 for genes identified within the differentially co-expressed modules.

Previous studies have identified several transcription factors to be associated with heart failure arising from ICM and IDCM [40–43].. For instance, TFs associated with innate immune response such as STAT1, IRF family of transcription factors (IRF1, IRF2), NFATC2, NFKB1, and CREB1 to be over-represented in at least one differentially co-expressed module identified here. Additionally, EGR1 was identified as a transcription factor significantly over-represented in 3/5 modules with a mean z-score of 4.1 (Figure 6.3). EGR1 is a transcription factor containing zinc finger DNA-binding motifs and domains to both activate and repress transcription induced under various conditions (stress, inflammatory response etc.) to bring about transcriptional regulation of various signaling cascades vital to growth, differentiation and apoptosis. Studies have shown that the sustained overexpression of EGR1 induces rapid induction of apoptosis associated with the activation of caspases and collapse of the mitochondrial

membrane potential in DCM [44,45]. Interestingly, two other members of the early growth factor TF family - EGR2 and EGR3 that are posited to negatively regulate T-cell activation [46], were also identified as being over-represented in these modules. However, the exact roles of these master regulators in cardiac failure are yet to be realized.

It has been previously suggested that the cardiac muscle reverts to a developmental/fetal gene profile in DC [42,47]. Consistent with this observation, it was interesting to note that a majority of the modules also showed an over-representation of SP/KLF family of TFs (such as SP1, SP2, SP3, SP4, KLF4, KLF5, KLF14, KLF16), known to critically regulate a host of fundamental cell differentiation and developmental processes [48,49]. Kruppel-like factors (KLF) family of zinc-finger TFs are being increasingly investigated in human health, particularly in cardiovascular biology [50,51]. For instance, KLF5 has been previously suggested to modulate myocardial remodeling in cardiac fibroblasts, induced by angiotensin [52]. KLF15 on the other hand has been suggested to potently inhibit the transcriptional activity of activators MEF2 and GATA4 in cardiomyocyte hypertrophy [51], suggestive of its role in anti-hypertrophy of the heart. MEF2 and GATA4 TFs have been previously implicated in DCM [53,54]. Of the KLFs which repress transcription, ECE1 rapidly downregulates the expression of KLF 3/5/11 [55].

While the precise roles for each of the specificity proteins SP 1/2/3/4, are yet to be elucidated in cardiac research, the SP family of TFs are known to affect a variety of cardiac functions within the adult heart. For instance, SP1 is involved in the regulation

of IGF1 in cardiac muscle cells [56], SP1 binding sites have also been identified on the human cardiac alpha-actin promoter sites[57]. The presence of several SP-1 binding sites on the collagen-1 promoter have suggested a preferential role for SP-1 in collagen 1 activation in DCM [58].

Thus dynamic regulation of expression by multiple members of the SP/KLF in cardiac myocytes suggests that, as a family, they are actively involved in regulating hypertrophy and apoptosis, further implying a potential for further research in delineating their roles in ICM and IDCM.

6.5 Conclusion

This study describes an approach to differential co-expression analysis using signed co-expression networks, subsequently applied to discerning mechanisms underlying ischemic and idiopathic dilated cardiomyopathy - two leading causes of human heart failure. Functional analysis of the differentially co-expressed modules, identified utilizing the defined approach, corresponded well with current understanding of aberrant processes underlying disease pathogenesis. Topological assessment of the hubs identified within these modules point to a strong association with known markers of disease. Additionally, these hubs were also identified to be enriched for variants in cis-eQTL within the heart left ventricle. Our analysis confirmed the action of previously implicated transcription factors such as STAT1, CREB1 in the pathogenesis of heart failure after DC. Our results also suggest a differential regulation of the targets of SP/KLF family of transcriptional factors in ICM and IDCM. Further research is

however needed to assess the precise role and differences in regulation by these transcription factors in the pathogenesis of ICM and IDCM.

In summary, we provide a scalable and unsupervised approach to identifying differential co-regulation from case-control studies using signed co-expression networks. Our results not only corroborate existing mechanisms underlying pathogenesis of dilated cardiomyopathy, but also provide insight into probable mechanistic differences.

6.6 Acknowledgements

The content of chapter 6 is a modified presentation of material being prepared for submission, currently titled “Mechanisms underlying ischemic and idiopathic dilated cardiomyopathy utilizing signed differential co-expression network.” by Mukund K, Subramaniam S. The dissertation author is the primary author for this material.

6.7 Tables

Table 6.1: Functional enrichment of differentially co-expressed module between ICM and IDCM
This table provides the functional enrichment of five modules with significant SMC ($p<0.05$) and high enrichment p-values ($p<0.05$), identified as being differentially co-expressed between ICM and IDCM. The observed and expected number of protein-protein interactions (PPI) as identified in the String v10.1 database is also provided along with their associated p-value.

Module name	**Module size**	**Top enrichment terms**	**p-value**	**Observed PPI**	**Expected PPI**	**p-value**
M21	244	defense response, cellular response to cytokine stimulus	<E-17	456	237	0
M19	192	blood circulation, circulatory system process	<E-08	119	90	2.10E-03
M4	112	regulation of angiotensin, membrane hyperpolarization	<E-04	50	33	4.57E-03
M3	328	regulation of vasoconstriction, organic anion transport	<E-05	327	244	2.54E-07
M18	68	serine family amino acid metabolic process, selenocysteine metabolic process	<E-11	79	26	0

Table 6.2: Hubs identified in top 5 modules with variants in cis-eQTL
This table represents the number of hub genes and variants that are in cis-eQTL as extracted from the GTeX consortium's cis-eQTL data for heart left ventricle with p<<0.0. A detailed list of SNPs and hubs are presented in Table S6.1

	Module size	**# of hubs**	**# of hubs with variants in cis-eQTL**	**#of associated SNPs**
M21	244	26	2	209
M18	68	8	3	178
M19	192	20	2	144
M3	328	34	7	334
M4	112	12	1	25

6.8 Figures

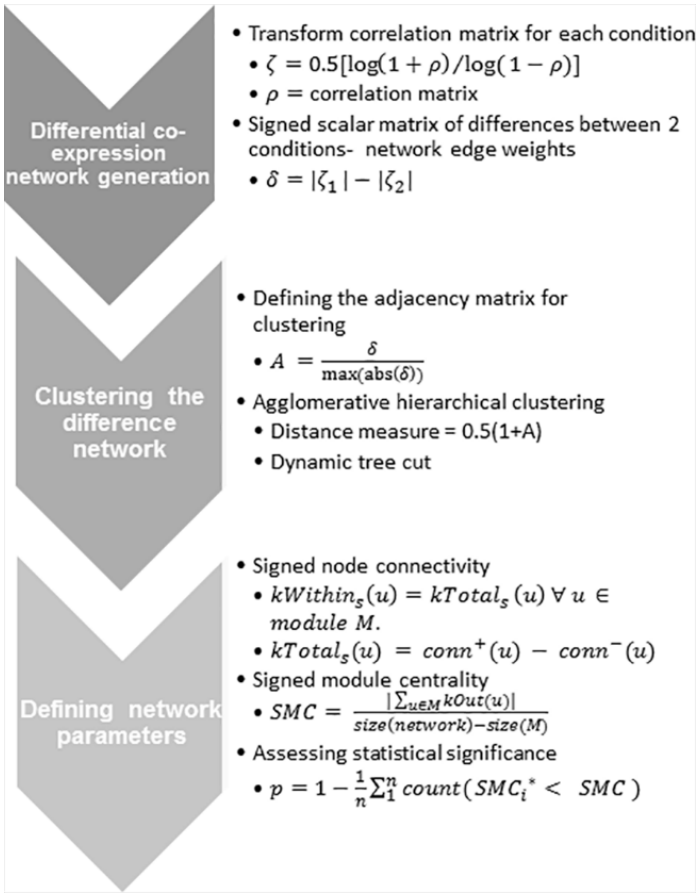

Figure 6.1: Workflow for clustering signed differential –co-expression network
The workflow presented here identifies the main steps involved in performing a differential co-expression analysis while retaining the sign of difference. Signed network parameters at node (connectivity) and module (signed module centrality) are also identified as part of this workflow.

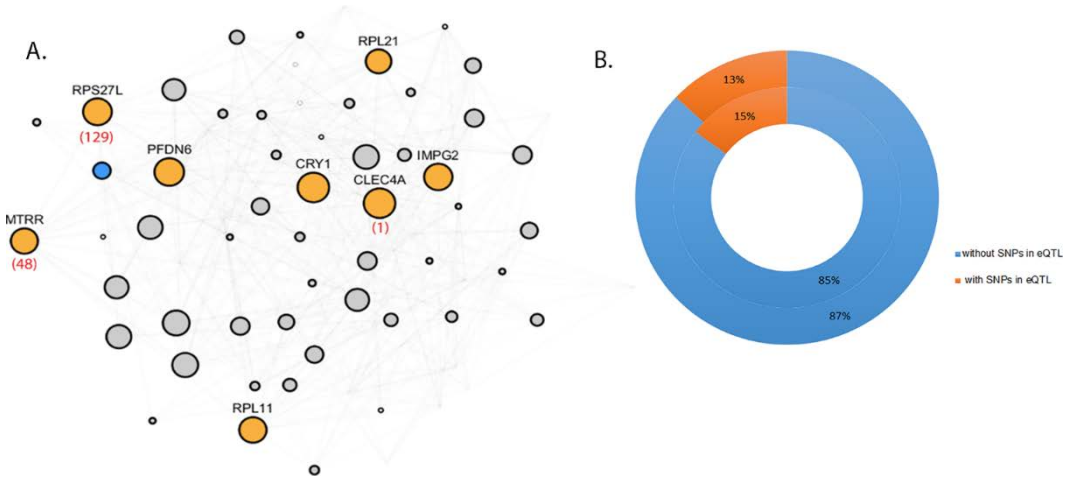

Figure 6.2: A. One of the top 5 modules is represented here (M18).
The size of the node is proportional to kWithin for the module. The orange nodes identified in this module are hubs, with the blue node indicating the node from the marker list of genes that are strongly associated with hub nodes as witnessed through the topological proximity of this node to the hub nodes. The number below the hubs in red indicate the number of SNPs in eQTL with the hub genes as identified through the GTeX project single tissue eQTL data for heart left ventricle. B. A radial graph giving the percentage of genes in the top 5 modules that are hubs (inner circle) and not hubs (outer circle) containing genes in eQTL.

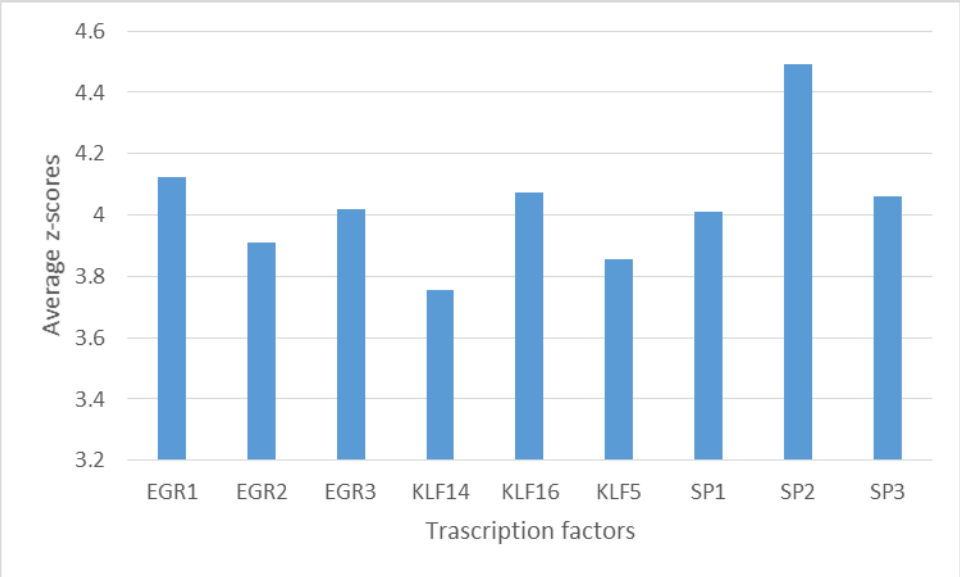

Figure 6.3: Z-scores of over- represented transcription factors in 5 top modules
The average z-score of select over-represented transcription factors extracted for the top 5 modules using the Pscan algorithm are presented here.

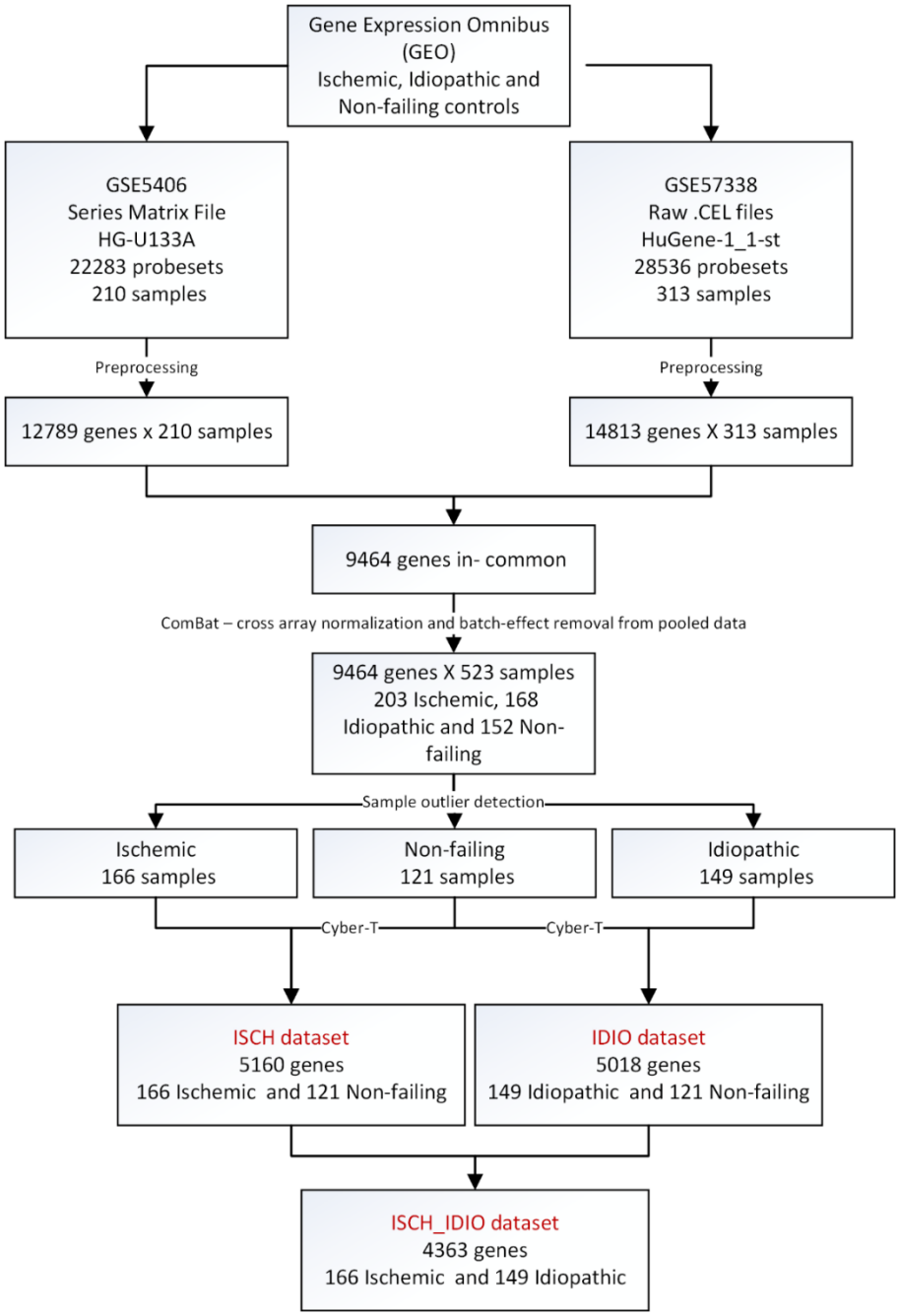

Figure S6.1: A detailed workflow for processing expression data from GEO
This workflow was adopted to extract and process data corresponding to ICM and IDCM samples from GEO prior to analysis.

6.9 References

1. Maron BJ, Towbin JA, Thiene G, Antzelevitch C, Corrado D, Arnett D, Moss AJ, Seidman CE, Young JB. Contemporary definitions and classification of the cardiomyopathies an American heart association scientific statement from the council on clinical cardiology, heart failure and transplantation committee; quality of care and outcomes research and functional genomics and translational biology interdisciplinary working groups; and council on epidemiology and prevention. Circulation. 2006;113:1807–16.

2. Hershberger RE, Hedges DJ, Morales A. Dilated cardiomyopathy: the complexity of a diverse genetic architecture. Nat. Rev. Cardiol. 2013;10:531–47.

3. Hazebroek M, Dennert R, Heymans S. Idiopathic dilated cardiomyopathy: possible triggers and treatment strategies. Neth. Heart J. 2012;20:332–5.

4. Hershberger RE, Morales A, Siegfried JD. Clinical and genetic issues in dilated cardiomyopathy: a review for genetics professionals. Genet. Med. 2010;12:655–67.

5. Mantziari L, Ziakas A, Ventoulis I, Kamperidis V, Lilis L, Katsiki N, Karavasiliadou S, Kiraklidis K, Pliakos C, Gemitzis K, Karvounis H. Differences in Clinical Presentation and Findings between Idiopathic Dilated and Ischaemic Cardiomyopathy in an Unselected Population of Heart Failure Patients. Open Cardiovasc. Med. J. 2012;6:98.

6. Kittleson MM, Minhas KM, Irizarry RA, Shui QY, Edness G, Breton E, Conte JV, Tomaselli G, Garcia JG, Hare JM. Gene expression analysis of ischemic and nonischemic cardiomyopathy: shared and distinct genes in the development of heart failure. Physiol. Genomics. 2005;21:299–307.

7. Azuaje FJ, Dewey FE, Brutsaert DL, Devaux Y, Ashley EA, Wagner DR. Systems-based approaches to cardiovascular biomarker discovery. Circ. Cardiovasc. Genet. 2012;5:360–7.

8. Villard E, Perret C, Gary F, Proust C, Dilanian G, Hengstenberg C, Ruppert V, Arbustini E, Wichter T, Germain M, Dubourg O. A genome-wide association study identifies two loci associated with heart failure due to dilated cardiomyopathy. Eur. Heart J. 2011;32:1065–76.

9. Meder B, Rühle F, Weis T, Homuth G, Keller A, Franke J, Peil B, Bermejo JL, Frese K, Huge A. A genome-wide association study identifies 6p21 as novel risk locus for dilated cardiomyopathy. Eur. Heart J. 2014;35:1069–77.

10. McNally EM, Barefield DY, Puckelwartz MJ. The genetic landscape of cardiomyopathy and its role in heart failure. Cell Metab. 2015;21:174–82.

11. Harakalova M, Kummeling G, Sammani A, Linschoten M, Baas AF, van der Smagt J, Doevendans PA, van Tintelen JP, Dooijes D, Mokry M, Asselbergs FW. A systematic analysis of genetic dilated cardiomyopathy reveals numerous ubiquitously expressed and muscle-specific genes. Eur. J. Heart Fail. 2015;17:484–93.

12. Kittleson MM, Shui QY, Irizarry RA, Minhas KM, Edness G, Conte JV, Parmigiani G, Miller LW, Chen Y, Hall JL, Garcia JG. Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy. Circulation. 2004;110:3444–51.

13. Molina-Navarro MM, Triviño JC, Martínez-Dolz L, Lago F, González-Juanatey JR, Portolés M, Rivera M. Functional networks of nucleocytoplasmic transport-related genes differentiate ischemic and dilated cardiomyopathies. A new therapeutic opportunity. PloS One. 2014;9:e104709.

14. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. 2005;4:1128.

15. de la Fuente A. From “differential expression”to “differential networking”–identification of dysfunctional regulatory networks in diseases. Trends Genet. 2010;26:326–33.

16. Tesson B, Breitling R, Jansen R. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinformatics. 2010;11:497.

17. Amar D, Safer H, Shamir R. Dissection of Regulatory Networks that Are Altered in Disease via Differential Co-expression. PLoS Comput. Biol. 2013;9:e1002955.

18. Leskovec J, Huttenlocher D, Kleinberg J. Signed networks in social media. Proc. 28th Int. Conf. Hum. Factors Comput. Syst. ACM; 2010. p. 1361–70.

19. Fisher RA. Statistical methods for research workers. Springer New York. 1992; 66-70.

20. Everett MG, Borgatti SP. The centrality of groups and classes. J. Math. Sociol. 1999;23:181–201.

21. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118–27.

22. Kayala MA, Baldi P. Cyber-T web server: differential analysis of high-throughput data. Nucleic Acids Res. 2012;40:W553–9.

23. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pagès F, Trajanoski Z, Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25:1091–3.

24. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498.

25. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013;41:D808–15.

26. Zambelli F, Pesole G, Pavesi G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. Nucleic Acids Res. 2009;37:W247–52.

27. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Kyung H, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis T, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalin A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struewing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Little AR, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J, Moore HF. The genotype-tissue expression (GTEx) project. Nat. Genet. 2013;45:580–5.

28. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics. 2008;24:719–20.

29. Anderson JL, Carlquist JF, Lutz JR, DeWitt CW, Hammond EH. HLA A, B and DR typing in idiopathic dilated cardiomyopathy: a search for immune response factors. Am. J. Cardiol. 1984;53:1326–30.

30. Fukunaga T, Soejima H, Irie A, Sugamura K, Oe Y, Tanaka T, Nagayoshi Y, Kaikita K, Sugiyama S, Yoshimura M, Nishimura Y, Ogawa H. Relation between CD4+ T-cell activation and severity of chronic heart failure secondary to ischemic or idiopathic dilated cardiomyopathy. Am. J. Cardiol. 2007;100:483–8.

31. Akasu M, Urata H, Kinoshita A, Sasaguri M, Ideishi M, Arakawa K. Differences in tissue angiotensin II–forming pathways by species and organs in vitro. Hypertension. 1998;32:514–20.

32. Raynolds M, Bristow M, Bush E, Abraham W, Lowes B, Zisman L, Taft CS, Perryman MB. Angiotensin-converting enzyme DD genotype in patients with ischaemic or idiopathic dilated cardiomyopathy. The Lancet. 1993;342:1073–5.

33. Zolk O, Quattek J, Sitzler G, Schrader T, Nickenig G, Schnabel P, Shimada K, Takahashi M, Böhm M. Expression of endothelin-1, endothelin-converting enzyme, and endothelin receptors in chronic heart failure. Circulation. 1999;99:2118–23.

34. BORLAK J, THUM T. Hallmarks of ion channel gene expression in end-stage heart failure. FASEB J. 2003;17:1592–608.

35. Lu G, Xu S, Peng L, Huang Z, Wang Y, Gao X. Angiotensin II upregulates Kv1. 5 expression through ROS-dependent transforming growth factor-beta1 and extracellular signal-regulated kinase 1/2 signalings in neonatal rat atrial myocytes. Biochem. Biophys. Res. Commun. 2014;454:410–6.

36. Velloso LA, Folli F, Sun XJ, White MF, Saad M, Kahn CR. Cross-talk between the insulin and angiotensin signaling systems. Proc. Natl. Acad. Sci. 1996;93:12490–5.

37. Chen H-C, Bouchie JL, Perez AS, Clermont AC, Izumo S, Hampe J, Feener EP. Role of the angiotensin AT1 receptor in rat aortic and cardiac PAI-1 gene expression. Arterioscler. Thromb. Vasc. Biol. 2000;20:2297–302.

38. Tousoulis D, Homaei H, Ahmed N, Asimakopoulos G, Zouridakis E, Toutouzas P, Davies GJ. Increased plasma adhesion molecule levels in patients with heart failure who have ischemic heart disease and dilated cardiomyopathy. Am. Heart J. 2001;141:277–80.

39. Dobaczewski M, Chen W, Frangogiannis NG. Transforming growth factor (TGF)-β signaling in cardiac remodeling. J. Mol. Cell. Cardiol. 2011;51:600–6.

40. Satoh J, Tabunoki H. A Comprehensive Profile of ChIP-Seq-Based STAT1 Target Genes Suggests the Complexity of STAT1-Mediated Gene Regulatory Mechanisms. Gene Regul. Syst. Biol. 2013;7:41–56.

41. Ng DC, dos Remedios CG, Bogoyevitch MA. Activation of signal transducer and activator of transcription (STAT) pathways in failing human hearts. Cardiovasc. Res. 2003;57:333–46.

42. Oka T, Xu J, Molkentin JD. Re-employment of developmental transcription factors in adult heart disease. Elsevier; 2007. p. 117–31.

43. Hannenhalli S, Putt ME, Gilmore JM, Wang J, Parmacek MS, Epstein JA, Morrisey EE, Margulies KB, Cappola TP. Transcriptional genomics associates FOX transcription factors with human heart failure. Circulation. 2006;114:1269–76.

44. Zins K, Pomyje J, Hofer E, Abraham D, Lucas T, Aharinejad S. Egr-1 upregulates Siva-1 expression and induces cardiac fibroblast apoptosis. Int. J. Mol. Sci. 2014;15:1538–53.

45. Lucas T, Kovatchki D, Abraham D, Schaefer R, Hofer E, Aharinejad S. Overexpression of Egr-1 is associated with dilated cardiomyopathy and induces cardiac cell apoptosis. FASEB J. 2007;21:A13.

46. Safford M, Collins S, Lutz MA, Allen A, Huang C-T, Kowalski J, Blackford A, Horton MR, Drake C, Schwartz RH, Powell JD. Egr-2 and Egr-3 are negative regulators of T cell activation. Nat. Immunol. 2005;6:472–80.

47. Taegtmeyer H, Sen S, Vela D. Return to the fetal gene program. Ann. N. Y. Acad. Sci. 2010;1188:191–8.

48. Suske G. The Sp-family of transcription factors. Gene. 1999;238:291–300.

49. Suske G, Bruford E, Philipsen S. Mammalian SP/KLF transcription factors: bring in the family. Genomics. 2005;85:551–6.

50. Feinberg MW, Lin Z, Fisch S, Jain MK. An emerging role for Krüppel-like factors in vascular biology. Trends Cardiovasc. Med. 2004;14:241–6.

51. Fisch S, Gray S, Heymans S, Haldar SM, Wang B, Pfister O, Cui L, Kumar A, Lin Z, Sen-Banerjee S, Das H, Petersen CA, Mende U, Burleigh BA, Zhu Y, Pinto YM, Liao R, Jain MK. Kruppel-like factor 15 is a regulator of cardiomyocyte hypertrophy. Proc. Natl. Acad. Sci. 2007;104:7074–9.

52. Shindo T, Manabe I, Fukushima Y, Tobe K, Aizawa K, Miyamoto S, Kawai-Kowase K, Moriyama N, Imai Y, Kawakami H, Nishimatsu H, Ishikawa T, Suzuki T, Morita H, Maemura K, Sata M, Hirata Y, Komukai M, Kagechika H, Kadowaki T, Kurabayashi M, Nagai R. Krüppel-like zinc-finger transcription factor KLF5/BTEB2 is a target for angiotensin II signaling and an essential regulator of cardiovascular remodeling. Nat. Med. 2002;8:856–63.

53. Li R-G, Li L, Qiu X-B, Yuan F, Xu L, Li X, Xu YJ, Jiang WF, Jiang JQ, Liu X, Fang WY, Zhang M, Peng LY, Qu XK, Yang YQ. GATA4 loss-of-function mutation underlies familial dilated cardiomyopathy. Biochem. Biophys. Res. Commun. 2013;439:591–6.

54. Xu J, Gong NL, Bodi I, Aronow BJ, Backx PH, Molkentin JD. Myocyte enhancer factors 2A and 2C induce dilated cardiomyopathy in transgenic mice. J. Biol. Chem. 2006;281:9152–62.

55. Cullingford TE, Butler MJ, Marshall AK, Sugden PH, Clerk A. Differential regulation of Krüppel-like factor family transcription factor expression in neonatal rat cardiac myocytes: effects of endothelin-1, oxidative stress and cytokines. Biochim. Biophys. Acta BBA-Mol. Cell Res. 2008;1783:1229–36.

56. Li T, Chen Y-H, Liu T-J, Jia J, Hampson S, Shan Y-X, Kibler D, Wang PH. Using DNA microarray to identify Sp1 as a transcriptional regulatory element of insulin-like growth factor 1 in cardiac muscle cells. Circ. Res. 2003;93:1202–9.

57. Gustafson T, Kedes L. Identification of multiple proteins that interact with functional regions of the human cardiac alpha-actin promoter. Mol. Cell. Biol. 1989;9:3269–83.

58. Pauschinger M, Knopf D, Petschauer S, Doerner A, Poller W, Schwimmbeck PL, Kühl U, Schultheiss HP. Dilated cardiomyopathy is associated with significant changes in collagen type I/III ratio. Circulation. 1999;99:2750–6.

Chapter 7- Summary and significance of findings

7.1 Summary of findings

Systems biology, as outlined in Chapter 1 is extensively utilized to discern mechanisms of disease, and increasingly so in muscle research. The methods described through the course of this dissertation illustrated the power of systems biology, particularly co-expression network theory, to meaningfully extract functional mechanisms in striated muscle under various conditions.

The analysis in Chapter 2 provided the first global and categorical assessment of transcriptional changes occurring across 1 year in mammalian skeletal muscle after treatment with Botulinum neurotoxin A (BoNT-A). Utilizing previously described physiological networks of muscle, systems-level analysis of mechanistic changes associated with treatment revealed dramatic regulation of several functional pathways at 1 week post-treatment. Muscle reverting to activation of fetal/immature isoforms indicated a possible role in muscle recovery. Transcriptional regulation associated with atrophy and fibrosis suggested transient extracellular effects in the early time points after BoNT-A injection. Derangement of ECM and fibrillar components was witnessed to occur by 4 weeks post- treatment.

Utilizing a data driven, network theoretic approach, we reassessed the effects of BoNT-A in Chapter 3. Clustering and re-grouping of co-regulated gene modules revealed dramatic regulation of metabolism and processes associated with muscle trophicity in samples from 1 week after treatment. Two putative marker genes Dclk1 and Ostalpha with potential role in skeletal muscle recovery, were identified. Consistent with findings of chapter 2, transcriptional recovery of muscle to the pre-treatment state

was witnessed beyond 12 weeks. Phenotypic correlation analysis in both chapters provided insight into the lag between transcriptional and structural response; probably a function of the length of time required for a neuromuscular unit to recover from a period of denervation-induced atrophy and fibrosis.

Chapters 4, 5 and 6 focused on deciphering the functional changes associated with human diseases affecting muscle. Particularly, in chapter 4, we illustrated the application of using preservation statistics to detecting modules functionally associated with dysregulated pathways in duchenne muscular dystrophy (DMD), as exemplified by the inflammatory module D2. This approach enabled identifying putative biomarkers, such as ACP5 identified within module D2, to be likely associated with the progression of DMD.

The quantitative framework proposed in chapter 5 for disease similarity used in conjunction with protein interaction data allowed for identification of commonly dysregulated pathways across a variety of muscle diseases. Likewise, the knowledge of how diseases functionally relate to each other using muscle functional modules provided invaluable insight into mechanistic differences in muscle, as witnessed through the differences identified in ALS and CP. Incorporating drug data into our quantitative framework allowed us to infer opportunities for exploring drug repurposing as option for treating diseases of the muscle.

In chapter 6, we described an approach to signed differential co-expression network analysis allowing us to identify groups of genes that are differentially co-regulated between conditions. In addition to the approach of identifying differential co-

regulation using signed networks, this chapter focused on utilizing the approach to discern mechanisms underlying ischemic (ICM) and idiopathic dilated cardiomyopathy (IDCM) - two leading causes of human heart failure. Our results not only corroborate existing mechanisms underlying pathogenesis of dilated cardiomyopathy, but also provide insight into probable mechanistic differences. The results of our analysis revealed possible differential action of SP/KLF family of transcription factors on targets regulated in ischemic and idiopathic dilated cardiomyopathy.

7.2 Significance of findings and future directions

During the course of this dissertation, we efficiently utilized and integrated publicly available high-throughput data from muscle, while harnessing the power of several existing tools from bioinformatics and systems biology. Furthermore, this thesis focused on designing and employing approaches to co-expression network analysis in an effort to elucidate the complex interactions underlying muscle pathologies.

Specifically, our results from chapters 2 and 3 indicated that at a molecular level, the effects of BoNT-A on muscle were relatively rapid, with most transcripts returning to control level by 12 weeks after treatment. This is consistent with use of the term “reversible chemodenervation”, with reference to the action of BoNT-A. Though no long-term transcriptional abnormalities were observed, our analysis of atrophy and fibrotic pathways suggested that further studies are necessary to determine optimal intervals for BoNT-A treatment from both a biological and physiological point-of-view.

The co-expression theoretic approaches adopted here enabled identification of putative biomarkers associated with muscle pathology. For instance, Dclk1 and Ostalpha were hypothesized to affect recovery of skeletal muscle after treatment with BoNT-A (chapter 3); while ACP5 was identified as a biomarker for progression of Duchenne muscular dystrophy (chapter 4) - setting the stage for further experiments for investigating the role of these biomarkers in the respective patho-mechanisms.

The methods described in chapters 4 and 6 provide a scalable and unsupervised approach to identifying differential co-regulation from two condition studies enabling the identification of probable mechanistic differences between conditions. The methods described in these two chapters can be utilized for exploratory analysis of dysregulated mechanisms as identified through case-control studies.

Chapter 5 demonstrated the value of an integrated approach with data obtained from various high-throughput sources (transcriptomic, protein interaction and drug-target) for synergistic identification of mechanisms shared among muscle diseases, which may or may not share clinical similarities. Furthermore, this study highlighted opportunities for therapeutic advancements for treating muscle diseases. Future studies incorporating various other forms of high-throughput data including epigenetic markers and genome wide association studies will reveal deeper insights into the disease associations and therapeutic opportunities.

In summary, the techniques and approaches developed as part of this dissertation, have lent themselves to delineating the complex system of interactions

underlying muscle pathologies; providing opportunities for drug development and personalized treatments to improve patient outcome.