Co-Expression Network Approach to Studying the Effects of Botulinum Neurotoxin-A

Kavitha Mukund⬡, Samuel R. Ward⬡, Richard L. Lieber⬡, and Shankar Subramaniam⬡

Abstract—Botulinum Neurotoxin A (BoNT-A) is a potent neurotoxin with several clinical applications.The goal of this study was to utilize co-expression network theory to analyze temporal transcriptional data from skeletal muscle after BoNT-A treatment. Expression data for 2000 genes (extracted using a ranking heuristic) served as the basis for this analysis. Using weighted gene co-expression network analysis (WGCNA), we identified 19 co-expressed modules, further hierarchically clustered into five groups. Quantifying average expression and co-expression patterns across these groups revealed temporal aspects of muscle's response to BoNT-A. 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 highly ranked, significantly expressed genes—*Dclk1* and *Ostalpha*—within group 5 suggested possible mechanistic roles in recovery from BoNT-A induced atrophy. Phenotypic correlations of groups with titin and myosin protein content further emphasized the effect of BoNT-A on the sarcomeric contraction machinery in early phase of chemodenervation. In summary, our approach revealed a hierarchical functional response to BoNT-A induced paralysis with early metabolic and later ECM responses and identified putative biomarkers associated with chemodenervation. Additionally, our results provide an unbiased validation of the response documented in our previous work.

Index Terms—Co-expression networks, clustering, gene ranking, muscle, botox, timecourse, cross sectional temporal data

---

1 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 (NMJ) 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], [3], [4], [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], [7], [8], [9], [10]. Such studies have been crucial in determining the dose-response efficacy of BoNT-A, for varied therapeutic use. However, only a handful studies have focused on deciphering transcriptional control of BoNT-A in striated muscle [11], [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 (t = 1, 4, 12, and 52 weeks after injection). We recently analyzed the functional changes associated with mammalian skeletal muscle after BoNT-A treatment via differential analysis of genes, and interpreted our results in the framework of established physiological muscle networks [14]. 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 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], [19], [20] and multiple time point studies [21]. Employing WGCNA on our time course data, we extracted modules, clustered into groups, which were "temporally" co-expressed i.e, exhibiting significant regulation only at certain timepoints.

- *K. Mukund and S. Subramaniam are with the Department of Bioengineering, UC San Diego, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: {k1mukund, shankar}@ucsd.edu.*
- *S.R. Ward is with the Departments of Radiology, Orthopaedic Surgery, and Bioengineering, 9500 Gilman Drive, La Jolla, CA 92093. E-mail: s1ward@ucsd.edu.*
- *R.L. Lieber is with the Shirley Ryan AbilityLab, 355 East Erie Chicago, IL 60611. E-mail: rlieber@ric.org.*

We identified putative novel gene associations with chemodenervation in muscle and utilized network toplogy and transcriptional factor binding, to infer their possible roles within muscle.

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.

2 METHODS

2.1 Data Acquisition

Data utilized in this study is publicly available through the Gene Expression Omnibus accession GSE52350. A detailed description of data generation and pre-processing steps are available in the original publication [14].

Briefly, samples were obtained for 5 sets of rats (n = 4/set) that included the Tibialis anterior (TA) muscle of BoNT-A injected rats sampled 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). Probe collapsing and removal of outlier samples resulted in a reduced dataset 13751 genes across 19 arrays.

This reduced dataset of the original publication was utilized as the input data for our current analysis.

2.2 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*. 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].

A detailed approach is outlined in the Additional File 1, Note S1, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2017.2763949.

2.3 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.

We utilized 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 [26]. 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 percent of the total gene set was utilized as the "seed set" to build our co-expression network.

2.4 Co-Expression Network Generation and Modularity Detection

We utilized an open source, R/Bioconductor package called WGCNA [27] for co-expression network 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 (pooled data analysis results provided in Additional File 1, Note S2, available online). 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).

2.5 Visualization and Functional Enrichment Analysis

Network modules generated using WGCNA were exported and visualized in Cytoscape [28] 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 [29] and RDAVIDWebService [31] were 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.

Functional enrichment was also utilized to assess statistical significance of the co-expressed modules by comparing the chosen test statistic (here p value of the top term in annotation cluster, pVal) with that expected at random. Background distribution of randomized modules was generated with sizes equal to original modules identified. That is, a module was defined to be significantly co-expressed (p_{modSig}) if the number of times the observed parameter (pVal) for a given module, was smaller than permuted
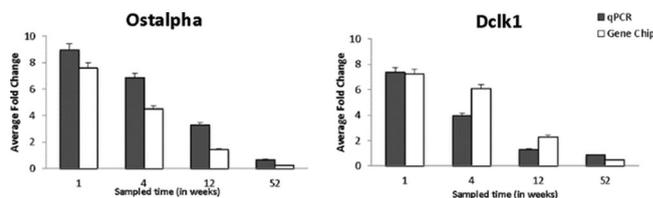
Fig. 1. Comparison of quantitative real-time PCR data with microarray expression data: Expression fold changes from GCRMA normalized data is shown in comparison with the real time PCR fold changes measured. The fold change was measured with respect saline injected rat control samples.

pVal* from 500 randomized runs resulting in a $p_{modSig} \leq 0.1$ (Eq. 1, Additional File 2, available online). Top term and corresponding p-value for the randomized modules were fetched from DAVID using the RDAVIDWebService [30], a Biconductor/R package.

$$p_{modSig} = 1 - \frac{1}{500} \sum_{1}^{500} count(pVal_i^* > \quad pVal) \qquad (1)$$

2.6 Protein Interaction Network

Protein interactions with combined score > 0.2, for all genes within module M3* (group 5) were extracted from STRING [31] and imported into Cytoscape [28]. 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.

2.7 Identifying Transcription Factor Binding Sites

MotifDB, a library in R, was utilized to extract promoter sequences 1000 bp (1 kb) upstream of genes using rat build rn5 as reference [32]. Sequence based motif searches were also performed using the same package.

3 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).

3.1 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. Empirical Bayes statistic [26] was utilized to rank the list of 13,751 genes (see Methods). Top ~15 percent (2,000) of highly ranked genes identified, were chosen as "seed set" for network reconstruction.

To ascertain if genes contained this list played a role in treated muscle, we compared the 2000 genes with the
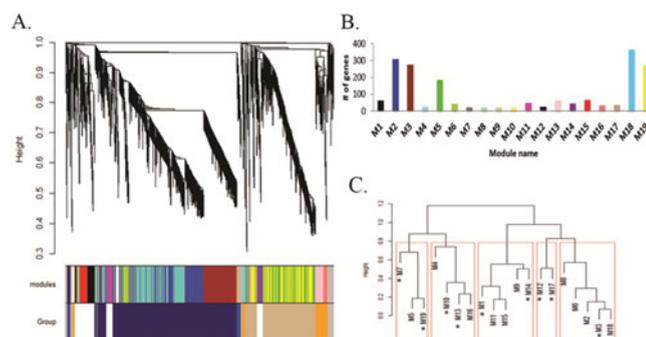
Fig. 2. A: Hierarchical clustering of 2000 genes from BoNT-A treated samples: The upper panel shows the clustering dendrogram with the middle and lower panels indicating the modules and their corresponding groups (designated by blue, white, dark blue, orange and tan colors) respectively. B: Size distribution of the modules identified where each color in the bar plot corresponds to the "modules" panel of Fig. 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 Fig. 2A. "*" indicates module significance identified via permuation testing

significantly differential genes (at any time point) identified in our earlier work [14]. An overlap of 61 percent was identified and 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, fewer genes associated with mitochondrial metabolism were identified in our current list, indicative of a fairly quick stabilization of mitochondrial metabolism gene expression across the time course. Interestingly, the two highest-ranking genes identified in the ranked list were *Ostalpha* and *Dclk1* with no currently evidenced role in skeletal muscle function (expression validated using qPCR, Fig. 1).

3.2 Reconstructing the BoNT-A Transcriptional Network

WGCNA [27] was utilized to reconstruct the BoNT-A transcriptional network. Initial adjacency matrix was 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 (Fig. 2A) with each module ranging in size between 20-360 genes (Fig. 2B). Module significance was assessed via permutation testing (Additional File 2, available online). 76 genes that did not belong to any module were excluded from further analysis for the purposes of this study (Fig. 2A, shown in light grey). The modules were re-clustered into five groups (Fig. 2C) based on their module Eigen-gene correlation (see Methods) namely group1 (M5, M7 and M19), group 2 (M4, M10, M13, and M16), group 3 (M1, M9, M11, M14, and M15), group 4 (M12, and M17), and group 5 (M2, M3, M6, M8, and M18).

3.3 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
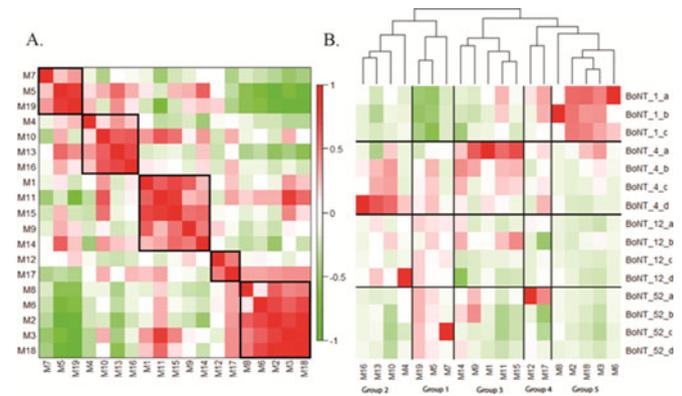
Fig. 3. A: Correlation heatmap of the module eigen genes: Each tile in the heatmap represents the scaled correlation between the module eigen genes (colored as per legend with red –correlated and green- anti-correlated). Groups are represented in black squares along the diagonal. High intra-group correlation is observed. B. Average expression heatmap for all modules identified in our network (colored as per legend with red –upregulated and green- downregulated)

provided insight into the collective behavior of the modules within groups. We observed strong intra-group correlation (along the diagonal, Fig. 3A) while the inter-group correlation was fairly non-significant except for a strong anti-correlation between groups 1 and 5. Additionally we computed *average expression,* which captures the collective behavior of all co-expressed genes within a module (Fig. 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. 9 (∼50 percent) of the modules were identified to be statistically significant at $p_{modSig} \leq 0.1$ (represented by an asterisk (*) in module name in the following sections). 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 (Additional File 1: Fig. S1, available online).

3.3.1   Group 5

Modules of group 5 were broadly associated with contradictory functions of atrophy and cellular recovery after chemodenervation (enrichment p value <0.05, Additional File 2, available online).

In particular, module M3* showed 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* [35], [36], [37]. Absence of *Casp3* has been shown to protect muscle from denervation induced atrophy [35] while *Emb* and *Lrp4* may serve as retrograde signals for the recovery of NMJ. Other
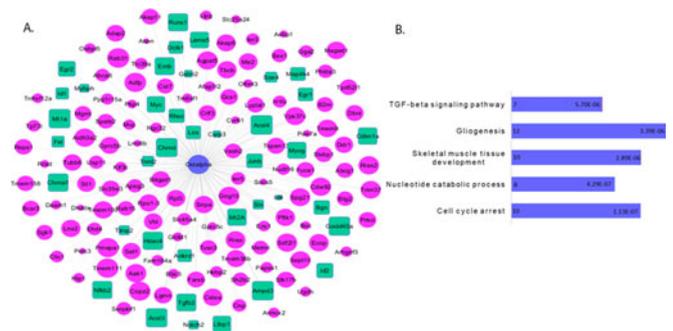
Fig. 4. Ostalpha co-expression sub network: A: Immediate neighbors of Ostalpha identified as being co-expressed in module M3*. Square nodes in green, indicate genes previously referenced in muscle literature, size of the node indicates the hub-ness (see Methods). B: The bar chart represents the relevant gene ontologies categories for this sub network identified with p-values and number of genes/category indicated within the plot.

pro-growth markers such as the Myogenic regulatory factors (MRFs)- *Myog* and *Myod1* [36], *Runx1* (sustains muscle under atrophic conditions [37]) 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* [38], [39].

A further advantage of co-expression network analysis over pairwise analysis of differential expression is that it allows to infer gene function based on network topology. [16]. Two highly ranked genes *Dclk1* and *Ostalpha*, with no evidenced role in treated or normal skeletal muscle, exhibited significant expression in our samples (Fig. 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 post-natally [40], [41]. 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 [42].

However, no prior evidence exists for the role of *Ostalpha* (a known bile transporter [43]) in skeletal muscle function. Immediate neighbors of *Ostalpha* within module M3* included several markers of striate muscle function including Emb, Chrnd Chrna, Cdkn1a, Mybph, Tnnt2, Ankrd1 (Fig. 4, Additional File 2, available online) Interestingly, functional analysis of first neighbors for Ostalpha revealed an over-representation of genes associated with TGF- beta signaling pathway, cell cycle arrest and response to DNA damage, further emphasizing a possible role for Ostalpha in response to muscle injury. A similar functional analysis of *Dclk1* gene neighborhood revealed an over-representation of genes associated with regulation of muscle adaptation upon inactivity or injury (Additional File 2, available online). *Casp3* for which *Dclk1* is a substrate in apoptotic neural cells [44], was also co-expressed in M3*.
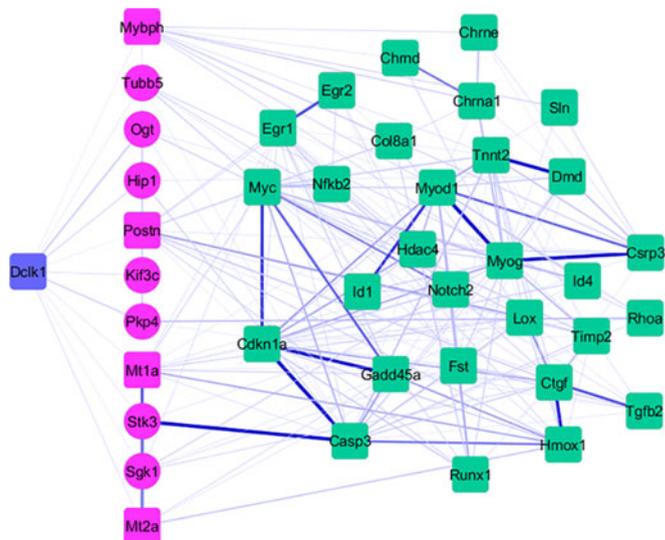
Fig. 5. Protein interaction map for Dclk1: Select protein interaction partners of Dclk1within M3 are shown here. Thickness and color of edges indicate the combined score for the interaction from STRING database. Nodes in pink are the 1-step neighbors for Dclk1 and nodes in green are its 2-step neighbors. Square nodes indicate genes previously referenced in muscle literature. A more complete network of interactions can be found in Additional File 1: Fig. S2.

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* [45]. Several of these TFs are known to be active in skeletal muscle with an important role in atrophying skeletal muscle [38]. Interestingly, we observed that *Dclk1* contained binding sites for muscle relevant TFs *Myod1, Runx1* and *Stat3* [46, p. 3], [47] on the promoter sequence 1 kb upstream while *Ostalpha* contained binding sites for *Myc* [48] and *Myod1*.

We next mapped the genes of module M3* onto protein interactions networks to identify known protein interactions [31]. *Dclk1* displayed several genes associated with skeletal muscle in its 2-step neighborhood including TFs identified above. (Fig. 5, Additional File 1: Fig. S2, available online).

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. Though M18 marginally missed our significance threshold, it was found to be enriched for markers of oxidative stress and metal ion imbalance (e.g., Mt1a, Mt2a).

3.3.2 Group 1

Enrichment analysis of repressed modules from group 1(M5, M7* and M19*) revealed functions associated specifically with metabolism and cellular homeostasis. This is consistent with an expected reduction in energy requirements 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*. Strong co-expression

of *Aqp4* (hub, see Methods) [49] with other metabolic genes within this module suggested a role for *Aqp4* in influencing metabolic activity of the chemodenervated muscle. M5, which marginally missed our statistical significance threshold, 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 (Additional File 2, available online). Regulation of ion homeostasis and metabolism are tightly coupled in skeletal muscle under most conditions [50].

The presence of functionally associated modules (such as M19* and M5) within a single group further reinforced the idea that grouping modules could provide a better view of their functional identity from time course studies.

3.3.3 Groups 2 and 3

Among the remaining groups, group 2 and to an extent group 3- exhibit higher expression in samples from 4 weeks (Fig. 3B). Group 2 modules (particularly M10*, M13*) showed enrichment for genes functionally associated with extracellular matrix (ECM). ECM is a carbohydrate rich connective tissue surrounding 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 M13* 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 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* [51] and *Casq2* [52], 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 [53]. Onset of contraction requires mechanical coupling of dihydropyridine receptors with ryanodine receptors (RYRs). This module contains an embryonic isoform of RYR –*Ryr3* that is strongly co-expressed with 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].

Correlation between M10* and several group 3 modules (Fig. 3) suggested a functional link between the increased expression of ECM associated genes and immature isoforms, reflecting the state of muscle at 4 weeks. This is consistent with the idea of progression of BoNT-A treated muscle towards a "hybrid" state, more pronounced by week 4 [14].

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 appeared to reverse trends and showed an increased average expression at later time points (specially 52 weeks) suggesting recovery of metabolism in the muscle after 1week.

A.

	ISO Pre	ISO Post
M7	-0.32 (0.2)	0.41 (0.1)
M5	-0.33 (0.2)	0.5 (0.06)
M19	-0.17 (0.5)	0.87 (2e-04)

	ISO Pre	ISO Post
M8	-0.013 (1)	-0.37 (0.2)
M6	0.31 (0.3)	-0.47 (0.07)
M2	0.27 (0.4)	-0.61 (0.02)
M3	0.15 (0.6)	-0.52 (2e-06)
M18	0.24 (0.4)	-0.79 (9e-04)

	Titin 1 Rx	Titin 1 Contra	Titin 2 Rx	Titin 2 Contra
M7	0.59 (0.02)	0.62 (0.01)	0.3 (0.3)	0.3 (0.3)
M5	0.79 (5e-04)	0.82 (2e-04)	0.69 (0.005)	0.77 (9e-04)
M19	0.68 (0.005)	0.8 (4e-04)	0.52 (0.05)	0.73 (0.002)

	Titin 1 Rx	Titin 1 Contra	Titin 2 Rx	Titin 2 Contra
M8	-0.32 (0.2)	-0.45 (0.09)	-0.38 (0.2)	-0.24 (0.4)
M6	-0.74 (0.002)	-0.76 (9e-04)	-0.66 (0.007)	-0.84 (8e-05)
M2	-0.82 (2e-04)	-0.87 (7e-05)	-0.73 (0.01)	-0.81 (3e-04)
M3	-0.75 (0.005)	-0.85 (7e-05)	-0.56 (0.03)	-0.77 (8e-04)
M18	-0.82 (2e-04)	-0.88 (2e-05)	-0.69 (0.01)	-0.85 (5e-05)

B.

	Myosin IIA Rx	Myosin IIA Contra	Myosin IIB Rx	Myosin IIB Contra
M1	0.26 (0.3)	-0.59 (0.04)	-0.25 (0.4)	0.43 (0.1)
M11	0.58 (0.02)	-0.34 (0.2)	-0.25 (0.4)	0.62 (0.02)
M15	0.69 (0.004)	-0.48 (0.07)	-0.58 (0.04)	0.41 (0.1)
M9	-0.022 (0.9)	-0.65 (0.009)	-0.4 (0.4)	0.17 (0.5)
M14	0.01 (0.1)	-0.44 (0.1)	-0.5 (0.06)	0.26 (0.4)

Fig. 6. Correlation between 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-1 (top) and 5 (bottom). 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.

3.4 Assessment of Phenotype to Module Correlation

Correlation analysis was performed on all modules, irrespective of their statistical significance, with an aim to understand the overarching effects of transcriptional regulation on the phenotype. The following sections provide a discussion of the correlation identified between each phenotype measured and gene groups (modules) identified in our network.

Isometric Contraction Modules of group 5 correlated positively with isometric contraction before injection (ISO Pre, Fig. 6A) while being significantly negatively correlated ($p < 0.05$) with isometric contraction measured after injection (ISO Post, Fig. 6A) with opposing patterns exhibited by group 1 genes. This suggested active and opposing roles for genes associated with groups 1 and 5 in the impaired contractility, particularly at 1 week after injection (Additional File 1: Fig. S3A, Fig. S4, available online). 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.

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 [55]. 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 (Fig. 6A, Additional File 1: Fig. S3B, available online). 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 [56]. Additionally, 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.

Myosin Heavy Chain (Fig. 6B). 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. 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 (Additional File 1: Fig. S3C, Fig. S5, available online).

4 CONCLUSION

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 utilized a data driven, co-expression network theoretic approach to analyze the transcriptional response of skeletal muscle secondary to the primary insult of BoNT-A at the NMJ. Clustering the co-expression network for 2000 genes across treated samples resulted in 19 modules that were subsequently clustered into 5 groups.

Co-expression of transcription factors *Myod1*, and *Runx1* with highly expressed genes *Dclk1* and *Ostalpha* combined with the presence of putative binding sites for these transcription factors on *Dclk1* and *Ostalpha*, pointed to their potential roles in skeletal muscle recovery after chemodenervation.

Expression patterns revealed dramatic regulation of metabolism (group 1) and processes associated with muscle trophicity (group 5) in samples from week 1. 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. Presence of functionally associated modules such as M19* and M5; M3* and M18 within relevant groups, and their association with samples from particular time points further reinforced the ability of this approach to identify functionally cohesive groups from time course studies.

Correlation of phenotypic data (titin and myosin protein content) with gene groups provided insight into the dynamics of the contractile proteins over time. In summary, this 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 *Dclk1* and *Ostalpha*, setting the stage for further experiments with implications for clinical use of BoNT-A.

REFERENCES

[1] L. L. Simpson, "Identification of the major steps in botulinum toxin action," *Annu. Rev. Pharmacol. Toxicol.*, vol. 44, pp. 167–193, 2004.

[2] J. O. Dolly, G. W. Lawrence, J. Meng, J. Wang, and S. V. Ovsepian, "Neuro-exocytosis: Botulinum toxins as inhibitory probes and versatile therapeutics," *Curr. Opin. Pharmacol.*, vol. 9, no. 3, pp. 326–335, 2009.

[3] R. L. Rosales and A. S. Chua-Yap, "Evidence-based systematic review on the efficacy and safety of botulinum toxin-A therapy in post-stroke spasticity," *J. Neural Transm.*, vol. 115, no. 4, pp. 617–623, 2008.

[4] A. M. Lang, "Focused review: Botulinum toxin type A therapy in chronic pain disorders," *Arch. Phys. Med. Rehabil.*, vol. 84, pp. S69–S73, 2003.

[5] P. G. Foran, et al., "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.*, vol. 278, no. 2, pp. 1363–1371, 2003.

[6] J. B. Hulst, V. B. Minamoto, M. B. Lim, S. N. Bremner, S. R. Ward, and R. L. Lieber, "Systematic test of neurotoxin dose and volume on muscle function in a rat model," *Muscle Nerve*, vol. 49, no. 5, pp. 709–715, 2014.

[7] J.-M. Gracies, "Physiological effects of botulinum toxin in spasticity," *Mov. Disord.*, vol. 19, no. S8, pp. S120–S128, 2004.

[8] F. A. Meunier, G. Schiavo, and J. Molgó, "Botulinum neurotoxins: From paralysis to recovery of functional neuromuscular transmission," *J. Physiol.-Paris*, vol. 96, no. 1, pp. 105–113, 2002.

[9] H.-M. Lee, J.-J. J. Chen, Y.-N. Wu, Y.-L. Wang, S.-C. Huang, and M. Piotrkiewicz, "Time course analysis of the effects of botulinum toxin type a on elbow spasticity based on biomechanic and electromyographic parameters," *Arch. Phys. Med. Rehabil.*, vol. 89, no. 4, pp. 692–699, 2008.

[10] C. G. Garner, A. Straube, T. N. Witt, T. Gasser, and W. H. Oertel, "Time course of distant effects of local injections of botulinum toxin," *Mov. Disord.*, vol. 8, no. 1, pp. 33–37, 1993.

[11] J. Ma, et al., "Time course of recovery of juvenile skeletal muscle after botulinum toxin A injection: An animal model study," *Amer. J. Phys. Med. Rehabil.*, vol. 83, no. 10, pp. 774–780, 2004.

[12] J. Ma, et al., "Gene expression of nAChR, SNAP-25 and GAP-43 in skeletal muscles following botulinum toxin A injection: A study in rats," *J. Orthop. Res.*, vol. 23, no. 2, pp. 302–309, 2005.

[13] J. Shen, et al., "How muscles recover from paresis and atrophy after intramuscular injection of botulinum toxin A: study in juvenile rats," *J. Orthop. Res.*, vol. 24, no. 5, pp. 1128–1135, 2006.

[14] K. Mukund, M. Mathewson, V. Minamoto, S. R. Ward, S. Subramaniam, and R. L. Lieber, "Systems analysis of transcriptional data provides insights into muscle's biological response to botulinum toxin," *Muscle Nerve*, vol. 50, no. 5, pp. 744–758, 2014.

[15] A. de la Fuente, "From 'differential expression'to 'differential networking'–identification of dysfunctional regulatory networks in diseases," *Trends Genet.*, vol. 26, no. 7, pp. 326–333, 2010.

[16] R. R. Nayak, M. Kearns, R. S. Spielman, and V. G. Cheung, "Coexpression network based on natural variation in human gene expression reveals gene interactions and functions," *Genome Res.*, vol. 19, no. 11, pp. 1953–1962, 2009.

[17] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," *Stat. Appl. Genet. Mol. Biol.*, vol. 4, no. 1, 2005, Art. no. 1128.

[18] M. Maschietto, et al., "Co-expression network of neural-differentiation genes shows specific pattern in schizophrenia," *BMC Med. Genomics*, vol. 8, no. 1, 2015, Art. no. 23.

[19] B. E. Haas, et al., "Adipose co-expression networks across finns and Mexicans identify novel triglyceride-associated genes," *BMC Med. Genomics*, vol. 5, no. 1, 2012, Art. no. 61.

[20] J. A. Miller, M. C. Oldham, and D. H. Geschwind, "A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging," *J. Neurosci.*, vol. 28, no. 6, 2008, Art. no. 1410.

[21] X. Chang, S. Liu, Y.-T. Yu, Y.-X. Li, and Y.-Y. Li, "Identifying modules of coexpressed transcript units and their organization of Saccharopolyspora erythraea from time series gene expression profiles," *PLoS One*, vol. 5, no. 8, 2010, Art. no. e12126.

[22] I. Piec, A. Listrat, J. Alliot, C. Chambon, R. G. Taylor, and D. Bechet, "Differential proteome analysis of aging in rat skeletal muscle," *FASEB J.*, vol. 19, no. 9, pp. 1143–1145, 2005.

[23] M. N. Goodman, S. M. Dluz, M. A. McElaney, E. Belur, and N. B. Ruderman, "Glucose uptake and insulin sensitivity in rat muscle: Changes during 3-96 weeks of age," *Amer. J. Physiol.-Endocrinol. Metab.*, vol. 244, no. 1, pp. E93–E100, 1983.

[24] T. D. Schmittgen and K. J. Livak, "Analyzing real-time PCR data by the comparative CT method," *Nature Protoc.*, vol. 3, no. 6, pp. 1101–1108, 2008.

[25] S. R. Ward, V. B. Minamoto, K. P. Suzuki, J. B. Hulst, S. N. Bremner, and R. L. Lieber, "Recovery of rat muscle size but not function more than 1 year after a single botulinum toxin injection," *Muscle Nerve*, May 2017, PMID: 28556093.

[26] Y. Chuan Tai and T. P. Speed, "On gene ranking using replicated microarray time course data," *Biometrics*, vol. 65, no. 1, pp. 40–51, 2009.

[27] P. Langfelder and S. Horvath, "WGCNA: An R package for weighted correlation network analysis," *BMC Bioinf.*, vol. 9, no. 1, p. 559, 2008.

[28] P. Shannon, et al., "Cytoscape: A software environment for integrated models of biomolecular interaction networks," *Genome Res.*, vol. 13, no. 11, 2003, Art. no. 2498.

[29] G. Bindea, et al., "ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks," *Bioinformatics*, vol. 25, no. 8, pp. 1091–1093, 2009.

[30] C. Fresno and E. A. Fernández, "RDAVIDWebService: A versatile R interface to DAVID," *Bioinformatics*, vol. 9, no. 21, pp. 2810–2811, 2013.

[31] A. Franceschini, et al., "STRING v9. 1: Protein-protein interaction networks, with increased coverage and integration," *Nucleic Acids Res.*, vol. 41, no. D1, pp. D808–D815, 2013.

[32] P. Shannon, "MotifDb: An annotated collection of protein-DNA binding sequence motifs," R Package Version 1.6.0, 2014.

[33] N. Kim, et al., "Lrp4 is a receptor for Agrin and forms a complex with MuSK," *Cell*, vol. 135, no. 2, pp. 334–342, 2008.

[34] E. Lain, et al., "A novel role for embigin to promote sprouting of motor nerve terminals at the neuromuscular junction," *J. Biol. Chem.*, vol. 284, no. 13, pp. 8930–8939, 2009.

[35] P. J. Plant, J. R. Bain, J. E. Correa, M. Woo, and J. Batt, "Absence of caspase-3 protects against denervation-induced skeletal muscle atrophy," *J. Appl. Physiol.*, vol. 107, no. 1, pp. 224–234, 2009.

[36] S. B. Chargé and M. A. Rudnicki, "Cellular and molecular regulation of muscle regeneration," *Physiol. Rev.*, vol. 84, no. 1, pp. 209–238, 2004.

[37] X. Wang, et al., "Runx1 prevents wasting, myofibrillar disorganization, and autophagy of skeletal muscle," *Sci. Signal.*, vol. 19, no. 14, 2005, Art. no. 1715.

[38] S. H. Lecker, et al., "Multiple types of skeletal muscle atrophy involve a common program of changes in gene expression," *FASEB J.*, vol. 18, no. 1, pp. 39–51, 2004.

[39] R. W. Jackman and S. C. Kandarian, "The molecular basis of skeletal muscle atrophy," *Amer. J. Physiol.-Cell Physiol.*, vol. 287, no. 4, pp. C834–C843, 2004.

[40] T. A. Deuel, J. S. Liu, J. C. Corbo, S.-Y. Yoo, L. B. Rorke-Adams, and C. A. Walsh, "Genetic interactions between doublecortin and doublecortin-like kinase in neuronal migration and axon outgrowth," *Neuron*, vol. 49, no. 1, pp. 41–53, 2006.

[41] T. Shu, et al., "Doublecortin-like kinase controls neurogenesis by regulating mitotic spindles and M phase progression," *Neuron*, vol. 49, no. 1, pp. 25–39, 2006.

[42] G. Pallafacchina, et al., "An adult tissue-specific stem cell in its niche: A gene profiling analysis of < i> in vivoquiescent and activated muscle satellite cells," *Stem Cell Res.*, vol. 4, no. 2, pp. 77–91, 2010.

[43] N. Ballatori, N. Li, F. Fang, J. L. Boyer, W. V. Christian, and C. L. Hammond, "OST alpha-OST beta: A key membrane transporter of bile acids and conjugated steroids," *Front. Biosci. J. Virtual Libr.*, vol. 14, 2009, Art. no. 2829.

[44] G. J. Schenk, et al., "A potential role for calcium/calmodulin-dependent protein kinase-related peptide in neuronal apoptosis: in vivo and in vitro evidence," *Eur. J. Neurosci.*, vol. 26, no. 12, pp. 3411–3420, 2007.

[45] H.-M. Zhang, et al., "AnimalTFDB: A comprehensive animal transcription factor database," *Nucleic Acids Res.*, vol. 40, no. D1, pp. D144–D149, 2012.

[46] Y. Yang, et al., "STAT3 induces muscle stem cell differentiation by interaction with myoD," *Cytokine*, vol. 46, no. 1, pp. 137–141, 2009.

[47] K. Kami and E. Senba, "In vivo activation of STAT3 signaling in satellite cells and myofibers in regenerating rat skeletal muscles," *J. Histochem. Cytochem.*, vol. 50, no. 12, pp. 1579–1589, 2002.

[48] E. A. Veal and M. J. Jackson, "C-myc is expressed in mouse skeletal muscle nuclei during post-natal maturation," *Int. J. Biochem. Cell Biol.*, vol. 30, no. 7, pp. 811–821, 1998.

[49] D. Basco, G. P. Nicchia, A. D'Alessandro, L. Zolla, M. Svelto, and A. Frigeri, "Absence of aquaporin-4 in skeletal muscle alters proteins involved in bioenergetic pathways and calcium handling," *PloS One*, vol. 6, no. 4, 2011, Art. no. e19225.

[50] D. A. Hood, I. Irrcher, V. Ljubicic, and A.-M. Joseph, "Coordination of metabolic plasticity in skeletal muscle," *J. Exp. Biol.*, vol. 209, no. 12, pp. 2265–2275, 2006.

[51] Y. Ji, et al., "Disruption of a single copy of the SERCA2 gene results in altered Ca2+ homeostasis and cardiomyocyte function," *J. Biol. Chem.*, vol. 275, no. 48, pp. 38073–38080, 2000.

[52] S. Viatchenko-Karpinski, et al., "Abnormal calcium signaling and sudden cardiac death associated with mutation of calsequestrin," *Circ. Res.*, vol. 94, no. 4, pp. 471–477, 2004.

[53] S. Kojic, D. Radojkovic, and G. Faulkner, "Muscle ankyrin repeat proteins: their role in striated muscle function in health and disease," *Crit. Rev. Clin. Lab. Sci.*, vol. 48, no. 5–6, pp. 269–294, 2011.

[54] C. W. Ward and G. G. Rodney, "Does a lack of RyR3 make mammalian skeletal muscle EC coupling a 'spark-less' affair?,'" *J. Physiol.*, vol. 586, no. 2, pp. 313–314, 2008.

[55] A. Engel and C. Franzini-Armstrong, *Myology: Basic and Clinical.* New York, NY, USA: McGraw-Hill, Medical Pub. Division, 2004.

[56] V. B. Minamoto, et al., "Increased efficacy and decreased systemic-effects of botulinum toxin A injection after active or passive muscle manipulation," *Dev. Med. Child Neurol.*, vol. 49, no. 12, pp. 907–914, 2007.

Kavitha Mukund received the BE degree from Vishvesvariah University, India, in 2006, the MS degree in biomedical informatics from Arizona State University, in 2009, and the PhD degree in bioinformatics and systems biology from University of California at San Diego, in 2016, advised by Dr. Shankar Subramaniam. She is a Helios scholar for 2009. She is currently working as a Postdoctoral Researcher in the Subramaniam lab. Her research interests include biological networks, systems biology and medicine, skeletal muscle, and big-data analysis.

Samuel R. Ward received the BS degree in physical therapy, in 1997, the PhD degree in biomechanics, in 2003, and completed the postdoctoral fellowship in muscle physiology in 2006. He has been with UC San Diego for 13 years. He is a deputy editor for the *Physical Therapy Journal* and has received numerous publication and professional association awards, including the Kappa Delta Award from the American Academy of Orthoapedic Surgeons in 2013. He was a core director for the National Skeletal Muscle Research Center (NIH R24) and San Diego Skeletal Muscle Research Centers (NIH P30). Currently, he is a Professor and Vice Chair of Research in Orthoapedic Surgery at UC San Diego. He is a clinician-scientist whose research focuses on musculoskeletal design and plasticity. He has published 100 peer-reviewed manuscripts and is currently funded by NIH and DoD. The current focus of his laboratory is on understanding chronic muscle atrophy and cell death in mature adults, with an emphasis on the use of stem cells and regenerative medicine approaches to treat these disease processes.

Richard L. Lieber received the PhD degree in biophysics from University of California, Davis, in 1983 and the MBA degree from University of California, San Diego, in 2013. He joined the faculty with the University of California, San Diego, in 1985 where he spent 30 years rising to the rank of Professor and Vice Chair of the Department of Orthopaedic Surgery. In recognition of the clinical impact of his basic science studies, He has been honored by the American Academy of Orthopaedic Surgeons (Kappa Delta Award; twice), the American Bone and Joint Surgeons (Nicolas Andry Award) the American College of Sports Medicine (fellow), and the Council for the International Exchange of Scholars (Fulbright Fellowship) and the American Society for Biomechanics (Borelli Award). He was also recently awarded the Senior Research Career Scientist from the Department of Veterans Affairs. He has published more than 250 peer reviewed journal articles. He is currently Chief Scientific Officer and Senior Vice President at the Rehabilitation Institute of Chicago focused on developing state-of-the-art approaches to understanding muscle contractures that result from cerebral palsy, stroke, and spinal cord injury
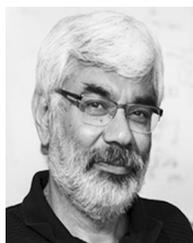
Shankar Subramaniam received the BS and MS degrees from Osmania University, Hyderabad, India, and the PhD degree in chemistry from the Indian Institute of Technology Kanpur, Kanpur, India. From 1991 to 1999, he was a Professor of Biophysics, Biochemistry, Molecular and Integrative Physiology, Chemical Engineering and Electrical and Computer Engineering with the University of Illinois, Urbana-Champaign, Urbana-Champaign, Illinois, prior to moving to the University of California, San Diego (UCSD), La Jolla, California. Currently, he is the Joan and Irwin Jacobs Endowed Chair in Bioengineering and Systems Biology at UCSD and a Professor of Bioengineering, Bioinformatics, and Systems Biology, Cellular and Molecular Medicine, Chemistry and Biochemistry, and Nanoengineering. He was the founding director of the bioinformatics and systems biology program at UCSD. He has authored more than 200 articles. Research in his laboratory spans several areas of bioinformatics, systems biology, and medicine. He is on the editorial board of several journals

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.