

RESEARCH ARTICLE

Heterogeneous muscle gene expression patterns in patients with massive rotator cuff tears

Michael C. Gibbons¹, Kathleen M. Fisch², Rajeswari Pichika³, Timothy Cheng³, Adam J. Engler¹, Simon Schenk³, John G. Lane³, Anshu Singh^{3,4}, Samuel R. Ward^{3,5*}

1 Department of Bioengineering, University of California San Diego, La Jolla, California, United States of America, **2** Department of Computational Biology, University of California San Diego, La Jolla, California, United States of America, **3** Department of Orthopedic Surgery, University of California San Diego, La Jolla, California, United States of America, **4** Department of Orthopedic Surgery, Kaiser Permanente, San Diego, La Jolla, California, United States of America, **5** Department of Radiology, University of California San Diego, La Jolla, California, United States of America

* s1ward@ucsd.edu



OPEN ACCESS

Citation: Gibbons MC, Fisch KM, Pichika R, Cheng T, Engler AJ, Schenk S, et al. (2018) Heterogeneous muscle gene expression patterns in patients with massive rotator cuff tears. *PLoS ONE* 13(1): e0190439. <https://doi.org/10.1371/journal.pone.0190439>

Editor: Chunfeng Zhao, Mayo Clinic Minnesota, UNITED STATES

Received: September 28, 2017

Accepted: December 14, 2017

Published: January 2, 2018

Copyright: © 2018 Gibbons et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All raw data are contained in the Supporting Information file.

Funding: This work was funded by the University of California, San Diego Frontiers of Innovation Program (M.C.G.), the Orthopedic Research and Education Foundation (T.C.), and the National Institute of Child Health and Human Development (S.R.W.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Detrimental changes in the composition and function of rotator cuff (RC) muscles are hallmarks of RC disease progression. Previous studies have demonstrated both atrophic and degenerative muscle loss in advanced RC disease. However, the relationship between gene expression and RC muscle pathology remains poorly defined, in large part due to a lack of studies correlating gene expression to tissue composition. Therefore, the purpose of this study was to determine how tissue composition relates to gene expression in muscle biopsies from patients undergoing reverse shoulder arthroplasty (RSA). Gene expression related to myogenesis, atrophy and cell death, adipogenesis and metabolism, inflammation, and fibrosis was measured in 40 RC muscle biopsies, including 31 biopsies from reverse shoulder arthroplasty (RSA) cases that had available histology data and 9 control biopsies from patients with intact RC tendons. After normalization to reference genes, linear regression was used to identify relationships between gene expression and tissue composition. Hierarchical clustering and principal component analysis (PCA) identified unique clusters, and fold-change analysis was used to determine significant differences in expression between clusters. We found that gene expression profiles were largely dependent on muscle presence, with muscle fraction being the only histological parameter that was significantly correlated to gene expression by linear regression. Similarly, samples with histologically-confirmed muscle distinctly segregated from samples without muscle. However, two sub-groups within the muscle-containing RSA biopsies suggest distinct phases of disease, with one group expressing markers of both atrophy and regeneration, and another group not significantly different from either control biopsies or biopsies lacking muscle. In conclusion, this study provides context for the interpretation of gene expression in heterogeneous and degenerating muscle, and provides further evidence for distinct stages of RC disease in humans.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The progressive and irreversible loss of rotator cuff (RC) muscle that occurs in RC disease is a vexing clinical challenge[1]. Despite advances in surgical tools and techniques, outcomes of RC repair are often unsatisfactory, especially for those with large tendon tears and chronic, advanced disease[1]. These suboptimal outcomes include tendon re-tear and persistent functional limitations[2], and occur in a significant number of cases[3]. Compositional changes on a macroscopic scale can in part explain these outcomes, as muscle volume is displaced by fat [4]. Changes at the tissue level may also be responsible for poor outcomes, as muscle fiber organization and force production are reduced with tear[5], and patients with the most severe RC disease (those undergoing RSA) demonstrate widespread muscle fiber degeneration[6]. To better understand the biological processes that govern muscle loss and fatty infiltration in RC disease, several studies have evaluated gene expression in human RC muscle[7, 8]. Here, we aim to address a key limitation of previous studies by correlating gene expression to histological biopsy composition, and provide potential interpretations of our findings as they relate to progression of RC disease.

Two previous studies of human RC muscle gene expression showed that when compared to small tears, large or massive tears generally exhibit depressed expression of key myogenic, adipogenic, and fibrotic genes along with high myostatin expression[7, 8], suggestive of an anti-myogenic disease process[9, 10]. However, a major limitation of these and many other molecular studies of heterogeneous tissues is the difficulty of reconciling gene expression values with changes in tissue composition, which could influence measured transcript abundance[11]. Given the gross changes in muscle composition observed across the spectrum of RC diseases [6, 12, 13], it is reasonable to hypothesize that gene expression changes are driven as much by the composition (e.g. muscle content) of the tissue as by changes in gene expression that occur within a given tissue type or cell population, a measurement which itself remains difficult[14]. Despite this common assumption, that gene expression patterns are influenced by changes in the underlying tissue composition, no previous study has included both gene expression and compositional data. Therefore, two major aims of this work were to 1) generate evidence to determine whether and to what extent tissue composition predicts gene expression patterns, and 2) use those findings to provide context for and caution against interpretation of gene expression data in the absence of compositional data.

Beyond the technical limitations of previous studies, we placed an emphasis on patients with advanced RC disease in this study, as these patients typically have the most severe muscle loss and the poorest outcomes among patients with cuff tears. We were particularly interested in genes and pathways involved in muscle atrophy and regeneration along with adipogenic and fibrotic genes, given the apparent irreversibility of muscle loss and fat and fibrotic tissue accumulation following chronic, massive RC tear. By providing insight into the relationship between gene expression and tissue composition, we hope to provide some perspective and context for previous studies while offering insight into the biological processes that govern the latter stages of RC disease.

Materials and methods

Patients

Twenty-three patients undergoing reverse total shoulder arthroplasty (RSA) were consented for RC muscle biopsy. All biopsies were performed with informed written consent under the approval of the UC San Diego IRB (study #090829). In order to specifically target muscle, obvious regions of tendon and fat were avoided and only samples that macroscopically appeared

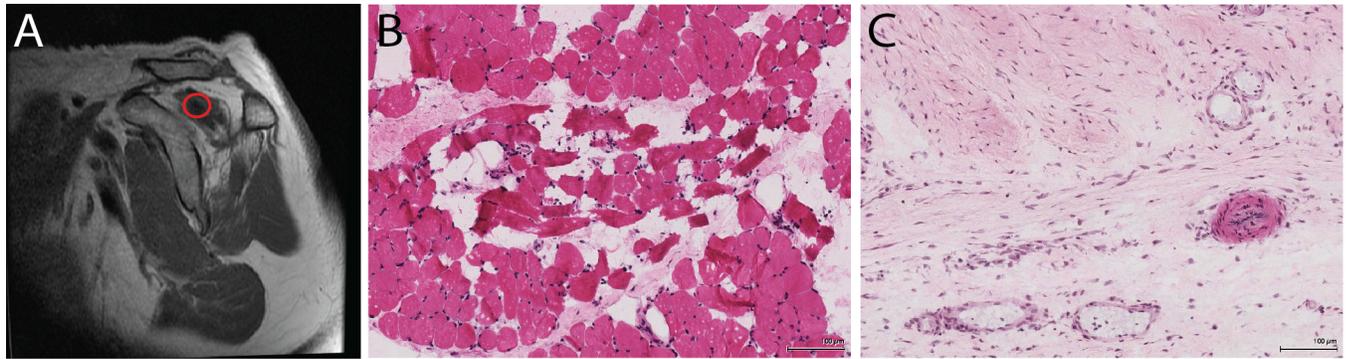


Fig 1. (A) MRI demonstrating the approximate biopsy region, where only regions of apparent muscle were targeted. (B) Representative H&E image of a muscle-containing biopsy, with high levels of muscle degeneration. (C) Representative H&E image of a biopsy that did not contain muscle, demonstrating high cellularity and presence of larger vascular structures.

<https://doi.org/10.1371/journal.pone.0190439.g001>

vascular and organized in fascicular structures were retained (red circle in Fig 1A). The supraspinatus was always biopsied if tissue meeting these criteria was present, otherwise the infraspinatus or teres minor were biopsied using the same criteria. In some cases both supraspinatus and infraspinatus were biopsied in order to increase the number of samples containing myofibers, leading to a total of 31 RSA biopsies, for which the histological data has been published [6]. Samples from patients with intact RC tendon were obtained arthroscopically from the superficial and lateral surface of the supraspinatus muscle during subacromial decompression surgeries (n = 9). After biopsy, samples were flash frozen in liquid nitrogen-cooled isopentane and stored at -80°C for future processing.

RNA isolation and quantitative PCR

For gene expression analysis, biopsy cross-sections weighing approximately 30–50 mg were cut from the biopsy center and homogenized in bead tubes (Navy, NextAdvance) with TRIzol (Ambion). RNeasy spin columns (Qiagen) were used to extract RNA using the manufacturer's protocol. One microgram of complimentary DNA (cDNA) was reverse transcribed with iScript cDNA Synthesis kits (Biorad). Quantitative PCR was carried out on custom plates on a BioRad CFX384 Touch qPCR analyzer for a panel of 42 genes associated with myogenic, atrophic, adipogenic, fibrotic and inflammatory pathways (Table 1), with cycle threshold determined using a SYBR green fluorophore. On-plate quality assessment was performed to assess gDNA contamination and RNA quality. Each sample was contained on a single plate, negating the need for inter-plate corrections.

Gene expression analysis

Raw cycle-threshold values (Ct values) were obtained from all samples and read into a qPCR expression set using the R Bioconductor package HTqPC, and were normalized using the delta-Ct normalization method to obtain gene expression values (*RPS18* and *ACTB* used as reference genes). Note that a maximum Ct of 39 was applied to all genes of interest to allow for statistical comparisons, and that lower values indicate higher expression in this method [15].

To determine the effect of tissue composition on muscle gene expression, linear regression of normalized gene expression and previously measured and reported histological parameters [6] was implemented. The following histological parameters from the previous study [6] were evaluated here: relative tissue fractions of muscle, connective tissue, and fat, along with inflammation (macrophage density) and vasculature (α -SMA+ vessel density and size). Coefficients

Table 1. Gene categories, linear regression, and PCA data.

Gene Category	Gene Name (Abbreviation)	p-value	r ²	1st PC	2nd PC
Muscle structure/myogenesis	Embryonic Myosin Heavy Chain (MYH3)	n.s.	—	0.194	0.077
Muscle structure/myogenesis	Myosin Heavy Chain—Type I (MYH1)	6.24E-05	0.41	0.108	0.211
Muscle structure/myogenesis	Insulin-like Growth Factor 1 (IGF1)	n.s.	—	0.106	-0.204
Muscle structure/myogenesis	Cysteine and Glycine Rich Protein 3 / Muscle LIM Protein (CSR3)	1.77E-07	0.602	0.131	0.194
Muscle structure/myogenesis	Ankyrin Repeat And SOCS Box Containing 15 (ASB15)	7.43E-10	0.726	0.144	0.200
Muscle structure/myogenesis	Ankyrin Repeat Domain 2-Stretch Responsive Muscle (ANKRD2)	7.88E-09	0.678	0.165	0.163
Muscle structure/myogenesis	Paired box 7 Transcription Factor (PAX7)	8.72E-09	0.676	0.168	0.185
Muscle structure/myogenesis	Myogenin/Myogenic Factor 4 (MYOG)	4.93E-08	0.635	0.154	0.177
Muscle structure/myogenesis	Myogenic Differentiation 1/Myogenic Factor 3 (MYOD1)	1.08E-07	0.615	0.128	0.172
Muscle structure/myogenesis	Myogenic Factor 5 (MYF5)	6.88E-07	0.564	0.160	0.153
Atrophy/Myogenic Inhibition	Myostatin/Growth Differentiation Factor 8 (MSTN)	1.28E-05	0.469	0.212	0.108
Atrophy/Myogenic Inhibition	Activin Receptor 2B (ACVR2B)	n.s.	—	0.217	0.002
Atrophy/Myogenic Inhibition	Tripartite Motif Containing 63/E3 Ubiquitin Ligase (TRIM63)	n.s.	—	0.169	0.144
Atrophy/Myogenic Inhibition	Forkhead Box O3 (FOXO3)	n.s.	—	0.233	-0.068
Atrophy/Myogenic Inhibition	F-box Protein 32/Atrogin-1/Muscle Atrophy Fbx32 (FBXO32)	3.23E-09	0.697	0.218	0.111
Atrophy/Myogenic Inhibition	Caspase-3 (CASP3)	n.s.	—	0.208	-0.127
Atrophy/Myogenic Inhibition	Caspase-1 (CASP1)	n.s.	—	0.175	-0.155
Metabolism	Protein Tyrosine Phosphatase, non-Receptor Type 4 (PTPN4)	n.s.	—	0.236	0.018
Metabolism	Mammalian Target of Rapamycin (MTOR)	n.s.	—	0.227	-0.055
Adipogenic	PPARG Coactivator 1 Alpha (PPARGC1A)	1.13E-06	0.549	0.209	0.098
Adipogenic	Peroxisome Proliferator-Activated Receptor Gamma (PPARG)	n.s.	—	0.200	-0.073
Adipogenic	Peroxisome Proliferator-Activated Receptor Delta (PPARD)	n.s.	—	0.223	-0.071
Adipogenic	Fatty Acid Binding Protein 4 (Adipocyte-Specific) (FABP4)	n.s.	—	0.201	0.007
Adipogenic	CCATT/Enhancer Binding Protein Alpha (CEBPA)	n.s.	—	0.193	-0.048
Adipogenic	Adiponectin (ADIPOQ)	n.s.	—	0.183	0.021
Adipogenic	Wnt Family Member 10B (WNT10B)	n.s.	—	-0.060	0.106
Inflammation	Tumor Necrosis Factor (TNF)	n.s.	—	0.119	-0.038
Inflammation	Interleukin-6 (IL6)	n.s.	—	0.093	-0.169
Inflammation	Interleukin-10 (IL10)	n.s.	—	0.047	-0.195
Inflammation	Interleukin-1 Beta (IL1B)	n.s.	—	0.051	-0.162
Fibrosis	Platelet-Derived Growth Factor Receptor Alpha (PDGFRA)	n.s.	—	0.140	-0.194
Fibrosis	Tissue Inhibitor of Metalloproteinase 3 (TIMP3)	n.s.	—	-0.177	0.128
Fibrosis	Tissue Inhibitor of Metalloproteinase 1 (TIMP1)	n.s.	—	0.093	-0.229
Fibrosis	Matrix Metalloproteinase 9 (MMP9)	n.s.	—	0.004	-0.154
Fibrosis	Matrix Metalloproteinase 3 (MMP3)	n.s.	—	-0.026	-0.148
Fibrosis	Matrix Metalloproteinase 1 (MMP1)	n.s.	—	-0.023	-0.135
Fibrosis	Lysyl Oxidase (LOX)	n.s.	—	0.070	-0.224
Fibrosis	Fibronectin 1 (FN1)	n.s.	—	-0.016	-0.219
Fibrosis	Connective Tissue Growth Factor (CTGF)	n.s.	—	0.113	-0.209
Fibrosis	Collagen Type III Alpha 1 Chain (COL3A1)	n.s.	—	0.012	-0.228
Fibrosis	Collagen Type I Alpha 1 Chain (COL1A1)	n.s.	—	0.011	-0.230
Fibrosis	Transforming Growth Factor Beta 1 (TGFB1)	n.s.	—	0.151	-0.217

Genes categorized by most relevant category. Coefficient of determination (r²) for normalized expression and muscle fraction calculated via histology. Gene weights for the first two principle components are reported. Note that directionality of gene weights indicates genes with opposing expression trends, and does not indicate positive or negative disease effects.

<https://doi.org/10.1371/journal.pone.0190439.t001>

of determination were calculated for linear relationships between expression values and histological parameters, and were considered significantly predictive when $r^2 > 0.2$ and Bonferroni-corrected p-values were statistically significant ($\alpha = 0.05$).

Unsupervised hierarchical clustering using Euclidean distance was then applied to the normalized expression values to determine the ability of gene expression patterns to differentiate muscle-containing samples from those without muscle, and to identify potential sub-clustering of muscle-containing samples. Where appropriate, compositional parameters were compared using a two-tailed t-test ($\alpha < 0.05$) to determine significant differences in average composition of the sub-group versus the remaining biopsy pool. Additionally, principle component analysis (PCA) was performed on the normalized gene expression values using the R package `prcomp` [16], in order to better appreciate sample clustering and to identify the genes with the largest effect on variability between samples. Subsequent differential gene expression sub-analyses were performed based on the groups identified by hierarchical clustering and confirmed by PCA.

Differential expression values (delta-delta-Ct) [15] were calculated with the `limmaCtData` wrapper in HTqPCR for the Bioconductor package `limma` using a moderated *t*-test [17]. Based on the cluster analysis and histological data available for the RSA biopsies, the intact comparisons described below include only intact biopsies that clustered with muscle-containing RSA samples. Differential expression values were computed for: 1) RSA biopsies with muscle present vs. without muscle, 2) pooled RSA biopsies versus the intact group, and 3) each main RSA cluster compared to the intact group. Differential expression values were also computed between clusters to determine if different sub-groups had significantly different gene expression. In all analyses, muscle content was included as a covariate to correct for the demonstrated effects of muscle content on expression profile, and genes with a Benjamini-Hochberg adjusted p-value < 0.05 were considered significantly differentially expressed. All raw data used in this study may be found in the Supporting Information (S1 File).

Results

As discussed in our previous publication [6], only 8/31 (26%) RSA samples contained histological evidence of skeletal muscle (Fig 1B). In samples lacking histologically identifiable muscle (Fig 1C), expression of muscle-specific genes was almost entirely absent (Fig 2, black dendrogram branches). Similarly, muscle content was significantly correlated to the expression of 11 muscle related genes (Table 1); there was no significant correlation between gene expression and any other histologic metric.

Hierarchical cluster analysis of normalized expression values (noting that with the delta-Ct method, greater expression is denoted by negative values) resulted in two primary clusters, with all muscle-containing RSA samples contained in one primary cluster that also included 5/9 intact samples (Fig 2). Within this muscle-containing cluster, samples were further segregated into two groups: a group with higher expression of myogenic, adipogenic, and metabolic genes and lower expression of fibrotic genes which will be referred to as the HIGH cluster (Fig 2, red), and a second group with lower expression of myogenic and atrophic genes and higher expression of fibrotic genes which will be referred to as the LOW cluster (Fig 2, orange). Based on the exclusivity of the main muscle cluster, only intact samples that co-segregated with muscle-containing RSA samples were used for differential expression comparisons between INTACT and RSA sub-clusters.

The strong effect of muscle presence and relative expression of muscle-related genes was observed in PCA as well. PCA analysis showed a high degree of separation between muscle-containing RSA samples and RSA biopsies without muscle (Fig 3). While the muscle-

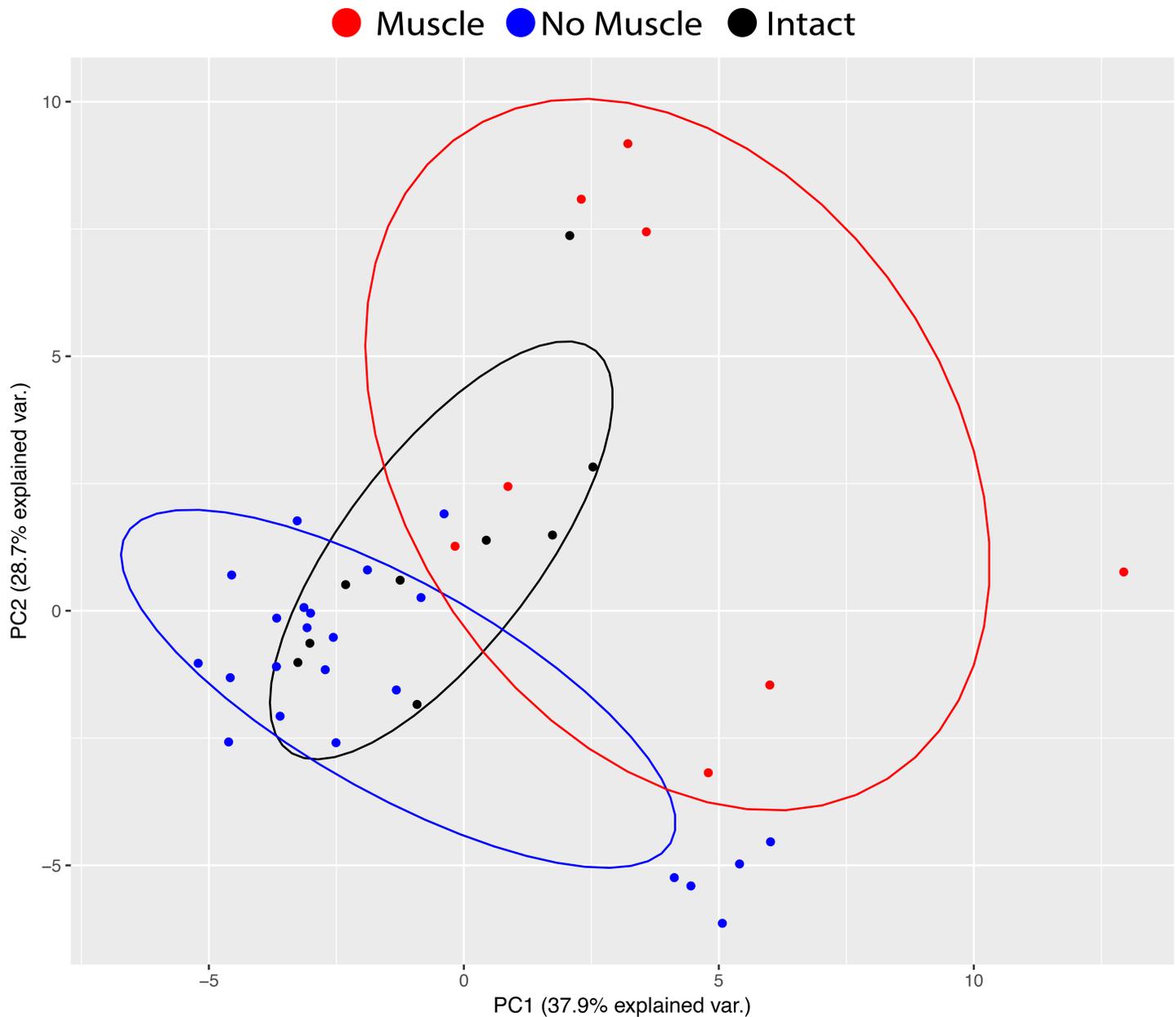


Fig 3. Principal component analysis employed to visualize variability between biopsies. Samples containing histological muscle are red, samples without muscle are blue, and controls are black. Of particular note are the cluster of blue samples in the lower center which correspond to the HI-FAT group in Fig 2, and the high variability in expression among the muscle-containing samples.

<https://doi.org/10.1371/journal.pone.0190439.g003>

differences in expression of any genes, though 8/9 pro-myogenic genes trended toward reduced expression (Fig 5).

These contrasting results highlight the need for a more detailed segregation and comparison of samples. Compared to INTACT, the HIGH group (red in Fig 2) demonstrated increased expression of both pro- and anti-myogenic genes, with reduced gene expression of three extracellular matrix proteins (Fig 6). However, in the LOW group (orange in Fig 2) no genes were significantly different compared to INTACT. Unsurprisingly, the HI-FAT group demonstrates increased expression of adipogenic genes and decreased muscle-related genes compared to

Fold Change Histological Muscle vs. No Muscle RSA Biopsies Only

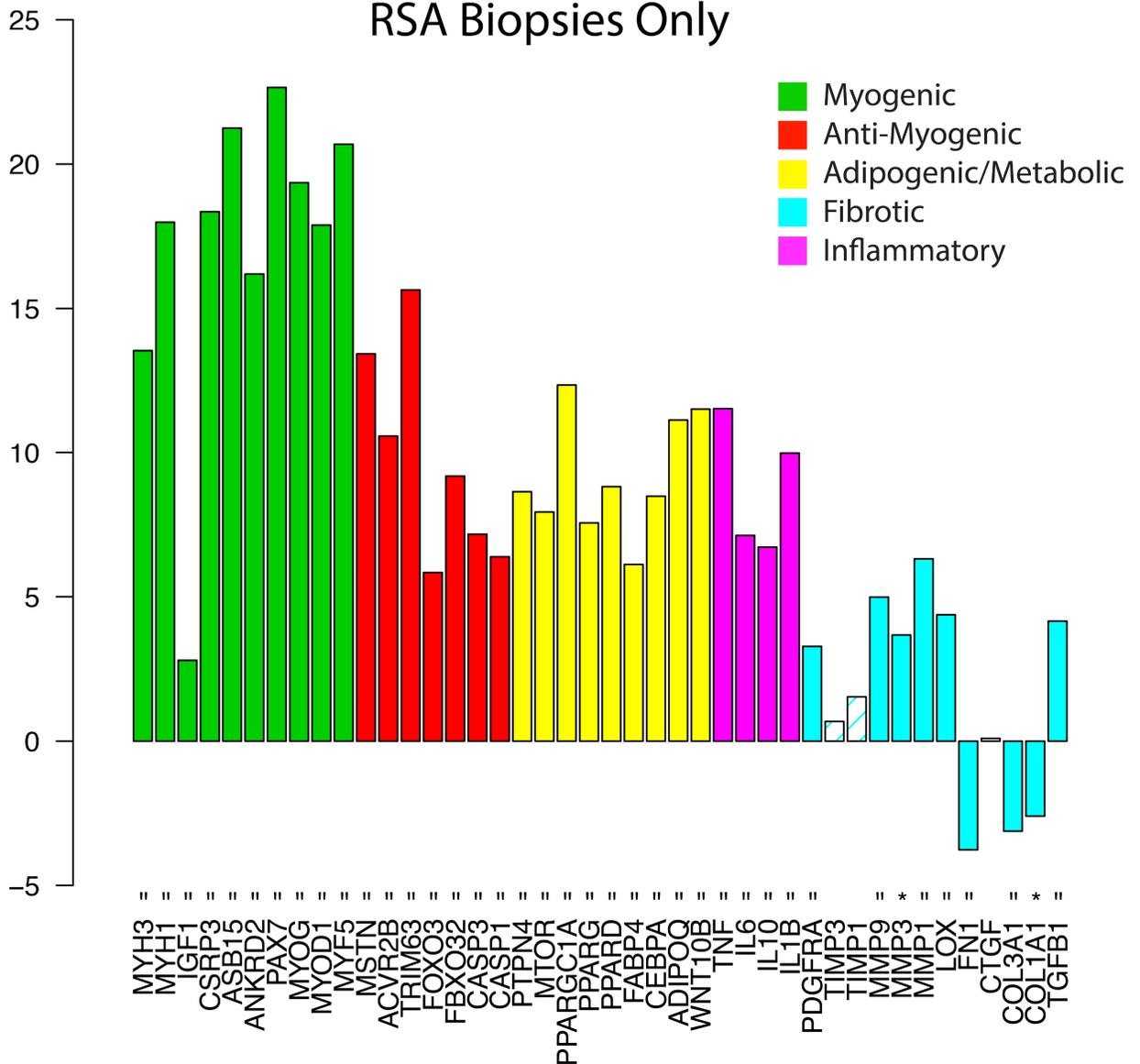


Fig 4. Fold change in expression between the RSA biopsies that contain muscle compared to those without muscle. Solid bars indicate significant up- or down-regulation ($p < 0.01$ and $p < 0.05$ indicated by ‘*’, and ‘**’, respectively). With muscle present, nearly all genes of interest are significantly differentially regulated, with increased expression of muscle- and fat-related genes and decreased expression of fibrosis-related genes.

<https://doi.org/10.1371/journal.pone.0190439.g004>

INTACT, while the NO-MUSCLE group shows significant reduction of both pro- and anti-myogenic genes.

Finally, fold-change expression was calculated between each sub-group (high muscle expression, low muscle expression, high fat content/expression, and the remaining samples without histological muscle or high fat). Compared to LOW group, the HIGH group demonstrated significantly increased expression of both pro- and anti-myogenic genes, with a similar but higher magnitude pattern observed between the HIGH and NO-MUSCLE groups (Fig 7A and 7B). This pattern was also repeated in comparing the HIGH and HI-FAT groups, though

Fold Change Pooled RSA vs. INTACT

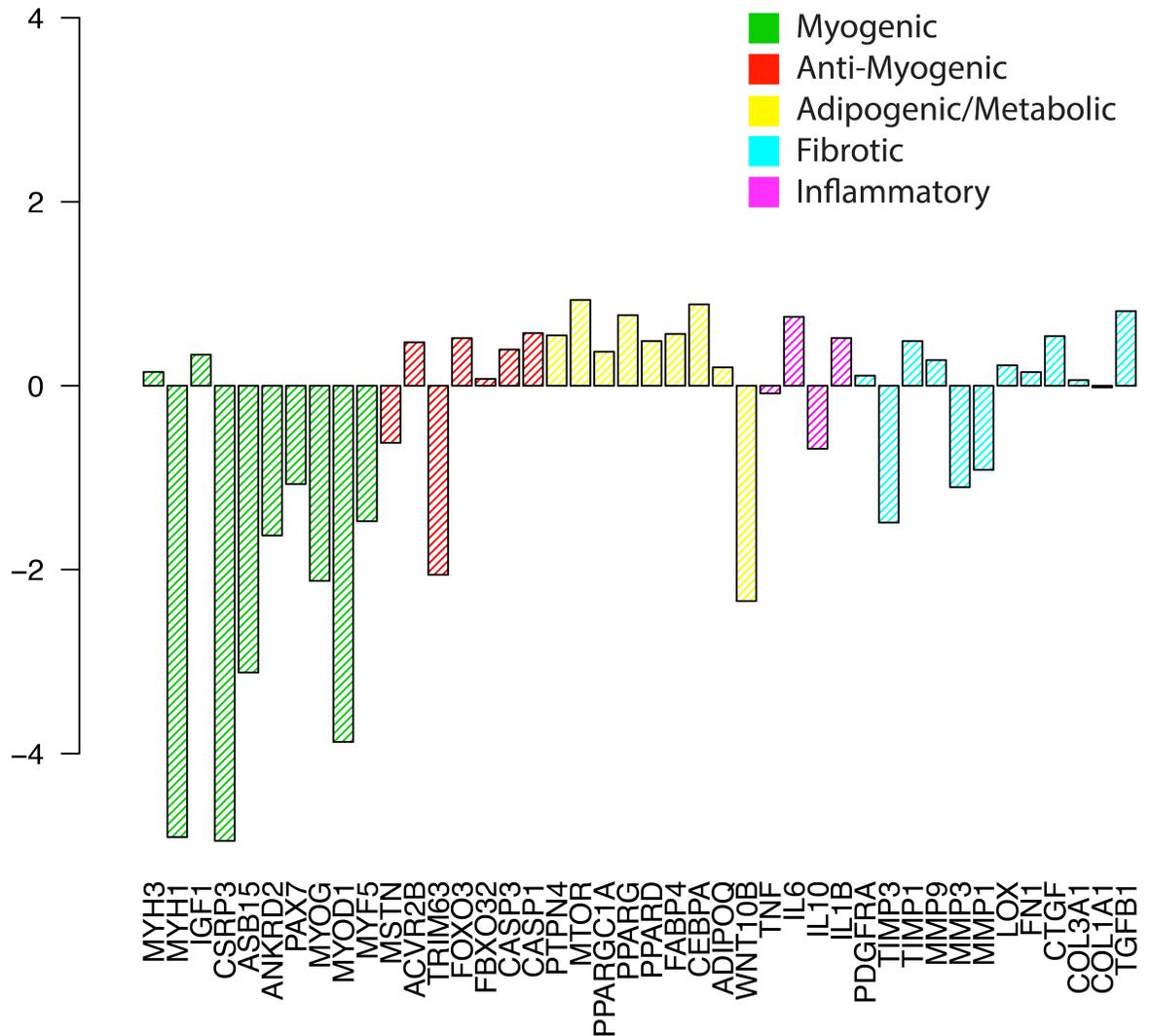


Fig 5. Fold change in expression between pooled RSA biopsies and controls. As a single pool, RSA biopsies are not significantly different from controls, though expression of pro-myogenic genes trended down while atrophic, adipogenic, and fibrotic genes trended up.

<https://doi.org/10.1371/journal.pone.0190439.g005>

additionally the HIGH group showed significantly decreased ECM-related gene expression (Fig 7C). Similar to the analysis relative to INTACT, the LOW group did not show significant differences in expression of any genes compared to either the HI-FAT or NO-MUSCLE groups (Fig 7D and 7E). Compared to the NO-MUSCLE group, the HI-FAT group demonstrated increased expression of both adipogenic and fibrotic genes.

Discussion

This study highlights the importance of understanding tissue composition, and in particular muscle content, when evaluating gene expression in human muscle biopsies taken from patients with chronic, degenerative disease. Even when screened at the time of surgery and harvested to avoid non-muscle tissue, biopsies are highly heterogeneous and often do not

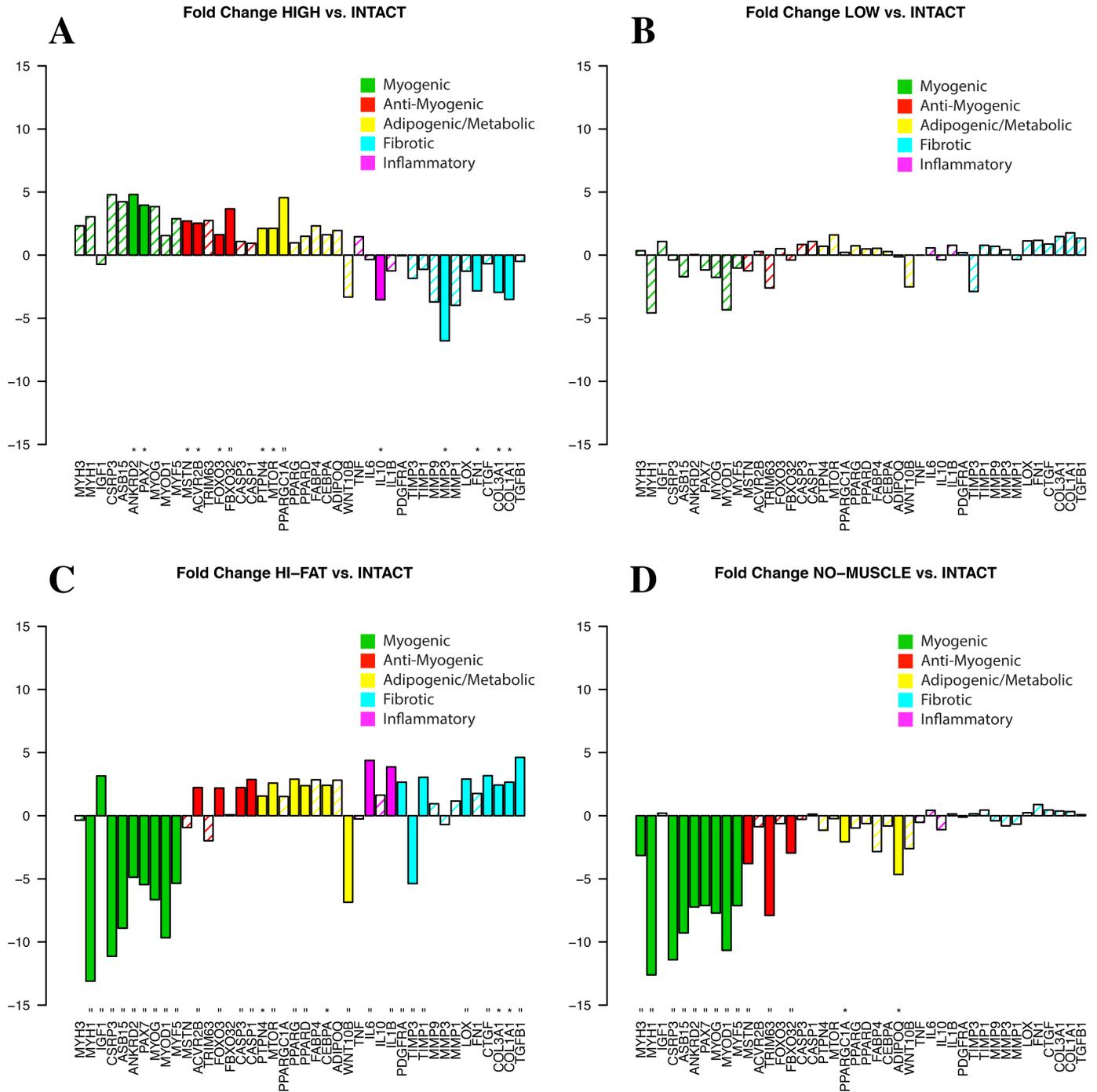


Fig 6. Fold changes in expression relative to INTACT for (A) HIGH muscle group (red in Fig 2), (B) LOW muscle group (orange in Fig 2), (C) HI-FAT group (purple in Fig 2), and (D) NO-MUSCLE group (black in Fig 2). Solid bars indicate significant up- or down-regulation ($p < 0.01$ and $p < 0.05$ indicated by ‘=’, and ‘**’, respectively).

<https://doi.org/10.1371/journal.pone.0190439.g006>

contain any muscle fibers[6]. This poses significant problems in the context of studying how and why muscle is irreversibly lost in RC disease[1] (and diseases with similar chronic and degenerative muscle loss, such as lumbar spine pathology[18] and muscular dystrophy[19–21]). When biopsies are taken from the muscle belly and appear to be muscle both via non-

Fig 7. Fold changes in expression between (A) HIGH and LOW expression muscle groups, (B) HIGH and NO-MUSCLE groups, (C) HIGH muscle and HI-FAT groups, (D) LOW muscle and HI-FAT groups, (E) LOW muscle and NO-MUSCLE groups, and (F) HI-FAT and NO-MUSCLE groups. Solid bars indicate significant up- or down-regulation ($p < 0.01$ and $p < 0.05$ indicated by ‘ ’, and ‘ * ’, respectively).

<https://doi.org/10.1371/journal.pone.0190439.g007>

invasive imaging and direct visualization, confirmation of muscle content, let alone broad analysis of biopsy composition, is rarely performed. This data highlights the potential pitfalls of omitting compositional data, where the conclusions reached with and without the context of histology are often drastically different.

The first broad aim of this study, to determine whether and to what extent the composition of a biopsy predicts its measured gene expression, demonstrated mixed results. Perhaps unsurprisingly, linear regression did demonstrate significant predictive capacity of muscle content for 10/17 genes expected to be expressed either solely or primarily by muscle fibers or muscle stem cells. However, the same analysis did not find any significant relationship between fat or fibrous tissue content and any gene family, suggesting that gene expression levels are not necessarily reflective of underlying tissue composition, as is often assumed. At a high level, this suggests that even when accounting for tissue composition, the mechanical and biochemical milieu of the torn RC effect gene expression in multiple cell and tissue types relevant to RC disease progression[22] independent of their relative quantity.

When taken as a single pool, there are no significant differences in the expression profile between RSA and INTACT biopsies. Alone, this finding could either be interpreted as RSA biopsies lacking a distinct gene expression signature compared to ‘healthy’ muscle, or as the INTACT biopsies possessing a ‘pathologic’ expression pattern (as the INTACT samples, while lacking tendon tears, nonetheless came from patients with shoulder pathology). With a more liberal approach, the trends (though insignificant) toward diminished signaling from muscle-related genes and pro-myogenic genes in RSA biopsies could be interpreted as diminished muscle maintenance overall, and indeed would align with previous studies[7, 8]. On the surface this does explain the macroscopic phenomenon; myogenic genes are diminished and muscle is lost over time[23, 24]. But this analysis is based on a highly heterogeneous biopsy pool in which the majority of biopsies *do not contain any muscle*, and therefore would not be expected to express muscle-specific myogenic genes. Indeed, the lack of a significant signal in the pooled RSA analysis appears to be due to high variability across RSA samples, and it is clear that any conclusions reached from this particular analysis, without the context of histology would, at best, be right for the wrong (or at least incomplete) reasons.

A clearer picture begins to emerge when biopsy composition is taken into account. Despite no significant difference in muscle fraction, muscle-containing biopsies segregate into two distinct groups (HIGH and LOW). Compared to INTACT, the HIGH group demonstrated increased expression of both atrophic/anti-myogenic and, to a lesser extent, pro-myogenic genes. This suggests that the HIGH biopsies are still responding to the atrophy pressure of mechanical unloading [25], and are mounting a regenerative response to the muscle fiber damage and degeneration observed histologically[6, 26–28]. Importantly, this interpretation directly refutes the conclusion reached from the pooled analysis, suggesting that continued muscle loss may be attributed to some combination of muscle atrophy and an inability of regenerative processes to match the rate of muscle degeneration, rather than a general reduction in myogenesis.

In contrast, the LOW group did not have any differentially expressed genes compared to INTACT, though myogenic gene expression trended downward. To better understand this finding, we looked to the sub-group comparisons. Surprisingly, the low-expression muscle group did not differentially express any genes compared to either the high fat or no muscle groups, but had increased expression of inflammatory and fibrotic markers compared to the

high-expression muscle group. Two competing explanations may account for this finding. One possibility is that this group is more 'terminally degenerated', evidenced by its similarity to the high-fat and no-muscle sub-groups. In this case, the similarity of expression to the INTACT group is indicative of pathology—the muscle is mechanically unloaded and contains degenerating fibers, but is not expressing the atrophic genes that are expressed in otherwise healthy unloaded muscle[25, 29, 30], nor is it mounting a normal regenerative response to muscle fiber damage[26, 27, 31–33]. However it is also possible that, based on its similarity to INTACT, the LOW group is in a less advanced stage of disease. Given the difficulty of resolving the time course of disease clinically, a definitive answer to this question from any clinical data set is unlikely; to satisfactorily determine the course of gene expression changes over time will likely require an animal model that adequately represents human disease.

Together, the segregation and interpretation of expression results for the HIGH and LOW groups provide further evidence for the existence of a biological spectrum of RC disease[6], though our understanding of how disease progresses through different stages, how these biological changes relate to clinical scoring systems[4, 34], and the processes that underlie the transition from active to terminal muscle degeneration, remains limited. While strategies have been developed for deconvolution of gene expression data, they rely on accurate understanding of both sample composition and expression in each individual cell population[35]. Even in healthy muscle this strategy would be complicated; fluorescence assisted cell sorting (FACS) could isolate and assay the subset of mononuclear cells expected to be found in the muscle, but there is often debate regarding the relationship between cell surface markers and specific cell populations[22, 36–38] and the isolation process itself may alter expression profiles in the cells of interest. Beyond that, the arguably most important cells, the multi-nucleated muscle fibers, would be lost in this process, as FACS cannot isolate myonuclei. In pathological muscle, where the relative cell fractions for known cell populations change dramatically and there remains the possibility of unexpected and potentially unidentified cell populations, this strategy may miss critical contributors to muscle pathology. Furthermore, though highly prevalent, the number of degenerating fibers in a given biopsy relative to the number of non-degenerating fibers at any single point in time is relatively low. One possible strategy to overcome this limitation is physical isolation of specific regions within the muscle biopsy using laser capture microscopy (LCM), which may allow for more targeted investigation of degenerative mechanisms without the confounding signal of surrounding fat, connective tissue, and non-degenerative muscle fibers. Indeed, LCM and associated downstream analytic techniques to hone in on mechanisms of muscle degeneration are a current focus in our lab.

Despite the fact that histological parameters other than muscle content are not linear predictors of gene expression, which may be explained by the chronicity of disease[23, 39] relative to the phase of active remodeling[40], these data nonetheless demonstrate the importance of interpreting gene expression data in the context of biopsy composition[11, 14]; without definitive evidence for the presence of muscle, there would be no basis for even the broad interpretations presented above. More generally, the disconnect between gene expression and histological findings (where even muscle presence and fat content are only modest predictors of expression levels and sample clustering) further complicates the interpretation of gene expression data not only in this and previous studies[7, 8], but throughout the RC disease literature, where complex interactions of the mechanical and biological environment have a significant impact on several of the tissue types central to RC muscle pathology[22].

A major remaining limitation of this work lies in our inability to histologically characterize our INTACT samples. Because the INTACT patients did have sufficient shoulder symptoms to warrant surgery, it is possible that these samples are not representative of normal, healthy muscle. Supporting this, 44% of INTACT biopsies were excluded from the differential expression

analysis based on cluster analysis, though it is important to note that trends in gene expression fold-change were not altered when normalized to the entire pool of INTACT samples. In the same vein, the relatively poor predictive ability of histological parameters outside of muscle content remains a limitation, and points to one of the overarching limitations of studying clinical samples, which is an inherent inability to resolve the time course of disease in the majority of patients[41]. To adequately address these limitations will require either development of a new animal model or validation of an existing model that adequately recapitulates both the atrophic and degenerative phases of human disease[22]. Such a model may then be employed to more accurately define the time course of disease, including the relationship between observed changes in gene expression as they relate to sample composition over time.

Conclusion

As a group, pooled RSA muscle biopsies provide little insight into gene expression patterns in RC disease, with only trends toward down-regulation of muscle-related genes and up-regulation of a limited number of inflammatory and adipogenic genes. But when muscle content determined via histology is considered, much of the variation in RSA gene expression can be explained—samples containing muscle fibers are unanimously separated from samples without muscle, and further segregate into two categories: one with increased expression of both myogenic and atrophic genes indicative of a responsive muscle undergoing active muscle turnover[42, 43], and another with an expression pattern not significantly different from control samples or samples without muscle, indicating potentially diminished responsiveness to both atrophic and degenerative stimuli. These categories may help explain the variability in rehabilitation found in the patient population; while the first category may represent a patient subset with some potential to respond to intervention[44, 45], the remaining patients (both with and without histological muscle) may represent a terminally degenerated muscle with limited regenerative capacity[1, 46, 47]. Future work in this area should focus on two areas: combining histology-informed interpretations of data with better controlled time series data from clinically relevant animal models, and employing advanced biological tools including FACS and LCM to physically separate and assay specific cell types and degenerative muscle fibers to better characterize the molecular mechanisms that govern progression of muscle loss and fat accumulation. The former strategy will allow for improved understanding of the temporal and spatial relationship between gene expression and tissue composition generally. However, the latter, more targeted approach will allow for a deeper understanding of the processes that govern irreversible muscle loss in chronic musculoskeletal conditions, including a more definitive understanding of the phases that define RC disease. Only by separating the highly heterogeneous mix of cell and tissue types[11, 14] will the cellular and molecular processes that govern RC disease progression from reversible, atrophic muscle loss to terminal muscle degeneration be elucidated. Ultimately, understanding the relationships between gene expression, disease state, and patient outcomes will aid in identifying optimal interventions on a more individualized basis, which will in turn lead to improved patient outcomes.

Supporting information

S1 File. Study data. This file contains all of the raw PCR and histological data for this paper. The 'Raw Cq Data' tab contains Cq data, where blank cells represent undetected genes which were set to the maximum Ct of 39 in order to perform statistical analysis. All histology data used for regression can be found in the 'Histology Data' tab.
(XLSX)

Author Contributions

Conceptualization: Michael C. Gibbons, Simon Schenk, Anshu Singh, Samuel R. Ward.

Data curation: Michael C. Gibbons, Kathleen M. Fisch, Timothy Cheng.

Formal analysis: Michael C. Gibbons, Kathleen M. Fisch.

Funding acquisition: Timothy Cheng, Samuel R. Ward.

Investigation: Michael C. Gibbons, Rajeswari Pichika, Timothy Cheng, John G. Lane, Anshu Singh.

Methodology: Michael C. Gibbons, Adam J. Engler, Simon Schenk, John G. Lane, Anshu Singh, Samuel R. Ward.

Project administration: Michael C. Gibbons, Adam J. Engler, Samuel R. Ward.

Resources: Simon Schenk, John G. Lane, Anshu Singh, Samuel R. Ward.

Software: Michael C. Gibbons, Kathleen M. Fisch.

Supervision: Michael C. Gibbons, Simon Schenk, Samuel R. Ward.

Validation: Michael C. Gibbons.

Visualization: Michael C. Gibbons, Kathleen M. Fisch.

Writing – original draft: Michael C. Gibbons.

Writing – review & editing: Michael C. Gibbons, Adam J. Engler, Simon Schenk, Anshu Singh, Samuel R. Ward.

References

1. Gladstone JN, Bishop JY, Lo IK, Flatow EL. Fatty infiltration and atrophy of the rotator cuff do not improve after rotator cuff repair and correlate with poor functional outcome. *The American journal of sports medicine*. 2007; 35(5):719–28. <https://doi.org/10.1177/0363546506297539> PMID: 17337727
2. Bishop J, Klepps S, Lo IK, Bird J, Gladstone JN, Flatow EL. Cuff integrity after arthroscopic versus open rotator cuff repair: a prospective study. *Journal of shoulder and elbow surgery*. 2006; 15(3):290–9. <https://doi.org/10.1016/j.jse.2005.09.017> PMID: 16679227
3. Galatz LM, Ball CM, Teefey SA, Middleton WD, Yamaguchi K. The outcome and repair integrity of completely arthroscopically repaired large and massive rotator cuff tears. *The Journal of Bone & Joint Surgery*. 2004; 86(2):219–24.
4. Goutallier D, Postel J-M, Bernageau J, Lavau L, Voisin M-C. Fatty muscle degeneration in cuff ruptures: pre-and postoperative evaluation by CT scan. *Clinical orthopaedics and related research*. 1994; 304:78–83.
5. Mendias CL, Roche SM, Harning JA, Davis ME, Lynch EB, Enselman ERS, et al. Reduced muscle fiber force production and disrupted myofibril architecture in patients with chronic rotator cuff tears. *Journal of Shoulder and Elbow Surgery*. 2015; 24(1):111–9. <https://doi.org/10.1016/j.jse.2014.06.037> PMID: 25193488
6. Gibbons MC, Singh A, Anakwenze O, Cheng T, Pomerantz M, Schenk S, et al. Histological Evidence of Muscle Degeneration in Advanced Human Rotator Cuff Disease. *The Journal of Bone & Joint Surgery*. 2017; 99(3):190–9. <https://doi.org/10.2106/jbjs.16.00335> PMID: 28145949
7. Choo A, McCarthy M, Pichika R, Sato EJ, Lieber RL, Schenk S, et al. Muscle Gene Expression Patterns in Human Rotator Cuff Pathology. *The Journal of Bone & Joint Surgery*. 2014; 96(18):1558–65.
8. Shah SA, Kormpakis I, Cavinatto L, Killian ML, Thomopoulos S, Galatz LM. Rotator cuff muscle degeneration and tear severity related to myogenic, adipogenic, and atrophy genes in human muscle. *Journal of Orthopaedic Research*. 2017.
9. Reardon KA, Davis J, Kapsa RM, Choong P, Byrne E. Myostatin, insulin-like growth factor-1, and leukemia inhibitory factor mRNAs are upregulated in chronic human disuse muscle atrophy. *Muscle & nerve*. 2001; 24(7):893–9.

10. Zimmers TA, Davies MV, Koniaris LG, Haynes P, Esqueda AF, Tomkinson KN, et al. Induction of cachexia in mice by systemically administered myostatin. *Science*. 2002; 296(5572):1486–8. <https://doi.org/10.1126/science.1069525> PMID: 12029139
11. Rodriguez-Gonzalez FG, Mustafa DA, Mostert B, Sieuwerts AM. The challenge of gene expression profiling in heterogeneous clinical samples. *Methods*. 2013; 59(1):47–58. <https://doi.org/10.1016/j.ymeth.2012.05.005> PMID: 22652627
12. Steinbacher P, Tauber M, Kogler S, Stoiber W, Resch H, Sanger A. Effects of rotator cuff ruptures on the cellular and intracellular composition of the human supraspinatus muscle. *Tissue and Cell*. 2010; 42(1):37–41. <https://doi.org/10.1016/j.tice.2009.07.001> PMID: 19709709
13. Lundgreen K, Lian B, Engebretsen L, Scott A. Lower muscle regenerative potential in full-thickness supraspinatus tears compared to partial-thickness tears. *Acta orthopaedica*. 2013; 84(6):565–70. <https://doi.org/10.3109/17453674.2013.858289> PMID: 24171689
14. Jung J, Jung H. Methods to analyze cell type-specific gene expression profiles from heterogeneous cell populations. *Animal Cells and Systems*. 2016; 20(3):113–7.
15. Dvinge H, Bertone P. HTqPCR: high-throughput analysis and visualization of quantitative real-time PCR data in R. *Bioinformatics*. 2009; 25(24):3325–6. <https://doi.org/10.1093/bioinformatics/btp578> PMID: 19808880
16. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. 2014.
17. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015:gkv007.
18. Shahidi B, Hubbard JC, Gibbons MC, Ruoss S, Zlomislic V, Allen RT, et al. Lumbar multifidus muscle degenerates in individuals with chronic degenerative lumbar spine pathology. *Journal of Orthopaedic Research*. 2017.
19. Petrof BJ. The molecular basis of activity-induced muscle injury in Duchenne muscular dystrophy. *Molecular and cellular biochemistry*. 1998; 179(1–2):111–24. PMID: 9543354
20. Haslett JN, Sanoudou D, Kho AT, Bennett RR, Greenberg SA, Kohane IS, et al. Gene expression comparison of biopsies from Duchenne muscular dystrophy (DMD) and normal skeletal muscle. *Proceedings of the National Academy of Sciences*. 2002; 99(23):15000–5.
21. Wren TA, Bluml S, Tseng-Ong L, Gilsanz V. Three-point technique of fat quantification of muscle tissue as a marker of disease progression in Duchenne muscular dystrophy: preliminary study. *American Journal of Roentgenology*. 2008; 190(1):W8–W12. <https://doi.org/10.2214/AJR.07.2732> PMID: 18094282
22. Gibbons MC, Singh A, Engler AJ, Ward SR. The Role of Mechanobiology in Progression of Rotator Cuff Muscle Atrophy and Degeneration. *Journal of Orthopaedic Research*. 2017.
23. Hebert-Davies J, Teefey SA, Steger-May K, Chamberlain AM, Middleton W, Robinson K, et al. Progression of Fatty Muscle Degeneration in Atraumatic Rotator Cuff Tears. *JBJS*. 2017; 99(10):832–9.
24. Sakuma K, Yamaguchi A. Molecular mechanisms in aging and current strategies to counteract sarcopenia. *Current Aging Science*. 2010; 3(2):90–101. PMID: 20158492
25. Bonaldo P, Sandri M. Cellular and molecular mechanisms of muscle atrophy. *Disease models & mechanisms*. 2013; 6(1):25–39.
26. Charge SB, Rudnicki MA. Cellular and molecular regulation of muscle regeneration. *Physiological reviews*. 2004; 84(1):209–38. <https://doi.org/10.1152/physrev.00019.2003> PMID: 14715915
27. Ciciliot S, Schiaffino S. Regeneration of mammalian skeletal muscle: basic mechanisms and clinical implications. *Current pharmaceutical design*. 2010; 16(8):906–14. PMID: 20041823
28. Goetsch SC, Hawke TJ, Gallardo TD, Richardson JA, Garry DJ. Transcriptional profiling and regulation of the extracellular matrix during muscle regeneration. *Physiological genomics*. 2003; 14(3):261–71. <https://doi.org/10.1152/physiolgenomics.00056.2003> PMID: 12799472
29. Bialek P, Morris CA, Parkington J, Andre MS, Owens J, Yaworsky P, et al. Distinct protein degradation profiles are induced by different disuse models of skeletal muscle atrophy. *Physiological genomics*. 2011.
30. Bodine SC, Latres E, Baumhueter S, Lai VK-M, Nunez L, Clarke BA, et al. Identification of ubiquitin ligases required for skeletal muscle atrophy. *Science*. 2001; 294(5547):1704–8. <https://doi.org/10.1126/science.1065874> PMID: 11679633
31. Czerwinska AM, Streminska W, Ciemerych MA, Grabowska I. Mouse gastrocnemius muscle regeneration after mechanical or cardiotoxin injury. *Folia Histochem Cytobiol*. 2012; 50(1):144–53. <https://doi.org/10.2478/18710> PMID: 22532150
32. Lepper C, Partridge TA, Fan C-M. An absolute requirement for Pax7-positive satellite cells in acute injury-induced skeletal muscle regeneration. *Development*. 2011; 138(17):3639–46. <https://doi.org/10.1242/dev.067595> PMID: 21828092

33. Wallace GQ, McNally EM. Mechanisms of muscle degeneration, regeneration, and repair in the muscular dystrophies. *Annual review of physiology*. 2009; 71:37–57. <https://doi.org/10.1146/annurev.physiol.010908.163216> PMID: 18808326
34. Fuchs B, Weishaupt D, Zanetti M, Hodler J, Gerber C. Fatty degeneration of the muscles of the rotator cuff: assessment by computed tomography versus magnetic resonance imaging. *Journal of Shoulder and Elbow Surgery*. 1999; 8(6):599–605. PMID: 10633896
35. Shen-Orr SS, Tibshirani R, Khatri P, Bodian DL, Staedtler F, Perry NM, et al. Cell type-specific gene expression differences in complex tissues. *Nature methods*. 2010; 7(4):287–9. <https://doi.org/10.1038/nmeth.1439> PMID: 20208531
36. Porpiglia E, Samusik N, Van Ho AT, Cosgrove BD, Mai T, Davis KL, et al. High-resolution myogenic lineage mapping by single-cell mass cytometry. *Nature Cell Biology*. 2017.
37. Rajkumar VS, Howell K, Csiszar K, Denton CP, Black CM, Abraham DJ. Shared expression of phenotypic markers in systemic sclerosis indicates a convergence of pericytes and fibroblasts to a myofibroblast lineage in fibrosis. *Arthritis research & therapy*. 2005; 7(5):R1113.
38. Mackey AL, Kjaer M, Charifi N, Henriksson J, Bojsen-Moller J, Holm L, et al. Assessment of satellite cell number and activity status in human skeletal muscle biopsies. *Muscle & nerve*. 2009; 40(3):455–65.
39. Essman JA, Bell RH, Askew M. Full-Thickness Rotator-Cuff Tear: An Analysis of Results. *Clinical orthopaedics and related research*. 1991; 265:170–7.
40. Jun J-I, Lau LF. Cellular senescence controls fibrosis in wound healing. *Aging (Albany NY)*. 2010; 2(9):627–31.
41. Bartolozzi A, Andreychik D, Ahmad S. Determinants of outcome in the treatment of rotator cuff disease. *Clinical orthopaedics and related research*. 1994; 308:90–7.
42. Raue U, Slivka D, Jemiolo B, Hollon C, Trappe S. Myogenic gene expression at rest and after a bout of resistance exercise in young (18–30 yr) and old (80–89 yr) women. *Journal of Applied Physiology*. 2006; 101(1):53–9. <https://doi.org/10.1152/jappphysiol.01616.2005> PMID: 16601301
43. Kim J-s, Kosek DJ, Petrella JK, Cross JM, Bamman MM. Resting and load-induced levels of myogenic gene transcripts differ between older adults with demonstrable sarcopenia and young men and women. *Journal of applied physiology*. 2005; 99(6):2149–58. <https://doi.org/10.1152/jappphysiol.00513.2005> PMID: 16051712
44. Jones SW, Hill RJ, Krasney PA, O'CONNOR B, Peirce N, Greenhaff PL. Disuse atrophy and exercise rehabilitation in humans profoundly affects the expression of genes associated with the regulation of skeletal muscle mass. *The FASEB journal*. 2004; 18(9):1025–7. <https://doi.org/10.1096/fj.03-1228fje> PMID: 15084522
45. Butt U, Rashid M, Temperley D, Crank S, Birch A, Freemont A, et al. Muscle regeneration following repair of the rotator cuff. *Bone Joint J*. 2016; 98(10):1389–94. <https://doi.org/10.1302/0301-620X.98B10.37231> PMID: 27694594
46. Warren GL, Summan M, Gao X, Chapman R, Hulderman T, Simeonova PP. Mechanisms of skeletal muscle injury and repair revealed by gene expression studies in mouse models. *The Journal of physiology*. 2007; 582(2):825–41.
47. Yan Z, Choi S, Liu X, Zhang M, Schageman JJ, Lee SY, et al. Highly coordinated gene regulation in mouse skeletal muscle regeneration. *Journal of Biological Chemistry*. 2003; 278(10):8826–36. <https://doi.org/10.1074/jbc.M209879200> PMID: 12477723