RESEARCH ARTICLE

A data-driven methodology reveals novel myofiber clusters in older human muscles

Yotam Raz[1] | Erik B. van den Akker[1,2,3] | Tijmen Roest[1] | Muhammad Riaz[4] | Ondine van de Rest[5] | H. Eka D. Suchiman[1] | Nico Lakenberg[1] | Stefanie A. Stassen[6] | Maaike van Putten[4] | Edith J. M. Feskens[5] | Marcel J. T. Reinders[2,3] | Jelle Goeman[7] | Marian Beekman[1] | Vered Raz[4] | Pieternella Eline Slagboom[1]

[1]Section of Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands

[2]Leiden Computational Biology Center, Leiden University Medical Center, Leiden, the Netherlands

[3]The Delft Bioinformatics Lab, Delft University of Technology, Delft, the Netherlands

[4]Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands

[5]Division of Human Nutrition, Wageningen University & Research, Wageningen, the Netherlands

[6]Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands

[7]Department of Medical Statistics, Leiden University Medical Center, Leiden, the Netherlands

Correspondence
Erik B. van den Akker, Section of
Molecular Epidemiology, Leiden University
Medical Center, Leiden 2300RC, the
Netherlands.
Email: e.b.van_den_akker@lumc.nl

Abstract

Skeletal muscles control posture, mobility and strength, and influence whole-body metabolism. Muscles are built of different types of myofibers, each having specific metabolic, molecular, and contractile properties. Fiber classification is, therefore, regarded the key for understanding muscle biology, (patho-) physiology. The expression of three myosin heavy chain (MyHC) isoforms, MyHC-1, MyHC-2A, and MyHC-2X, marks myofibers in humans. Typically, myofiber classification is performed by an eye-based histological analysis. This classical approach is insufficient to capture complex fiber classes, expressing more than one MyHC-isoform. We, therefore, developed a methodological procedure for high-throughput characterization of myofibers on the basis of multiple isoforms. The mean fluorescence intensity of the three most abundant MyHC isoforms was measured per myofiber in muscle biopsies of 56 healthy elderly adults, and myofiber classes were identified using computational biology tools. Unsupervised clustering revealed the existence of six distinct myofiber clusters. A comparison with the visual assessment of myofibers using the same images showed that some of these myofiber clusters could not be detected or were frequently misclassified. The presence of these six clusters was

reinforced by RNA expressions levels of sarcomeric genes. In addition, one of the clusters, expressing all three MyHC isoforms, correlated with histological measures of muscle health. To conclude, this methodological procedure enables deep characterization of the complex muscle heterogeneity. This study opens opportunities to further investigate myofiber composition in comparative studies.

K E Y W O R D S

bioinformatics, clustering, fibertyping, human, muscle, muscle health, myofiber, myosin heavy chain, RNA-sequencing, sarcomere

1 | INTRODUCTION

Human skeletal muscles undergo profound metabolic changes throughout lifetime, affected by physical activity, physiology, and disease. Muscles are built of different types of myofibers, each having specific metabolic, molecular, and contractile properties. Thus, alterations in muscle function are reflected by changes in myofiber type composition. Fiber classification is, therefore, regarded the key for understanding processes affecting muscle biology, physiology, and pathophysiology.[1-6] For instance, muscle degeneration is often characterized by atrophy of especially fast-glycolytic myofibers,[7-9] while insulin resistance is associated with shifts from slow-oxidative to fast-glycolytic myofibers.[10] Conventionally, myofiber type is determined by colorimetric staining for metabolic enzyme activity, such as ATPase.[4,11] Thus, obtained measurements are then typically employed to assign fibers to one of two classes (positive and negative to the staining). For example, the expression of myosin heavy chain (MyHC) isoforms, assessed by immunohistochemistry,[9,12-14] is employed to classify myofibers as fast- or slow-twitch. However, visual classification of myofibers does not only oversimplify the natural occurring variation, but also is prone to introduce observational bias.[15] Hence, traditional methods for fiber typing, that is, the binary classification of fibers on the basis of the visual assessment of a single parameter, seems insufficient to capture the more subtle variations in myofiber composition that for instance occur during mild conditions or muscle aging.

Various studies have reported the presence of hybrid myofibers, that is, myofibers expressing two MyHC isoforms, in the context of aging.[16,17] Specifically, an increase in the numbers of hybrid myofibers was reported in murine aging.[9,18-20] Employing multiplexed immune-fluorescently labeled antibodies to simultaneously measure multiple MyHC isoforms together with a semi-automated method for high-throughput image quantification, we reported that the proportion of myofibers expressing two MyHC isoforms exceeds those expressing only a single MyHC isoform[19] This methodological procedure can capture a large number of myofibers allowing one to study the muscle composition at a much higher resolution using computational analyses. This opens the opportunity to describe myofibers based on multiple MyHC abundances, and moreover to study the continuous transitions in myofiber composition. As an example, we recently reported differences in myofiber compositions between two mouse muscles, which are related to differential muscle involvement in dystrophic mouse models and age.[21] Thus far, however, little is known about the biological implications of myofiber composition in human muscles. Since aging muscles have been reported to higher fiber transformations, we used muscles from elderly to develop a methodology for studying complex myofiber composition in humans.

Here we report a methodological procedure investigating myofiber composition in *vastus lateralis* derived from 56 healthy elderly human subjects (age: 63 ± 5 years). Obtained muscle biopsies were characterized by various histological staining and RNA expression profiles (summarized in Figure 1A). Procedures for high-throughput quantitative imaging were applied to quantify the mean fluorescence intensity (MFI) for each of three MyHC isoforms per myofiber. Subsequent computational biology analyses of the 16 939 obtained myofibers identified six distinct myofiber clusters (summarized in Figure 1B). Some of these myofiber clusters may be hard to delineate using traditional methods, however correlations with RNA expression profiles of sarcomeric genes, as well as, various muscle health parameters, suggesting their biological relevance.

2 | MATERIALS AND METHODS

2.1 | Study population

The current study was performed within the Growing Old Together (GOTO) study. Participants underwent a 13-weeks lifestyle intervention consisting of 12.5% caloric restriction and 12.5% increase in physical activity.[22] Exclusion criteria included individuals with known diabetes or fasting glucose > 7.0 mmol/L. In total 164 participants underwent the lifestyle intervention, of which 56 participants (23 females and 33 males) volunteered to donate muscle biopsies that were
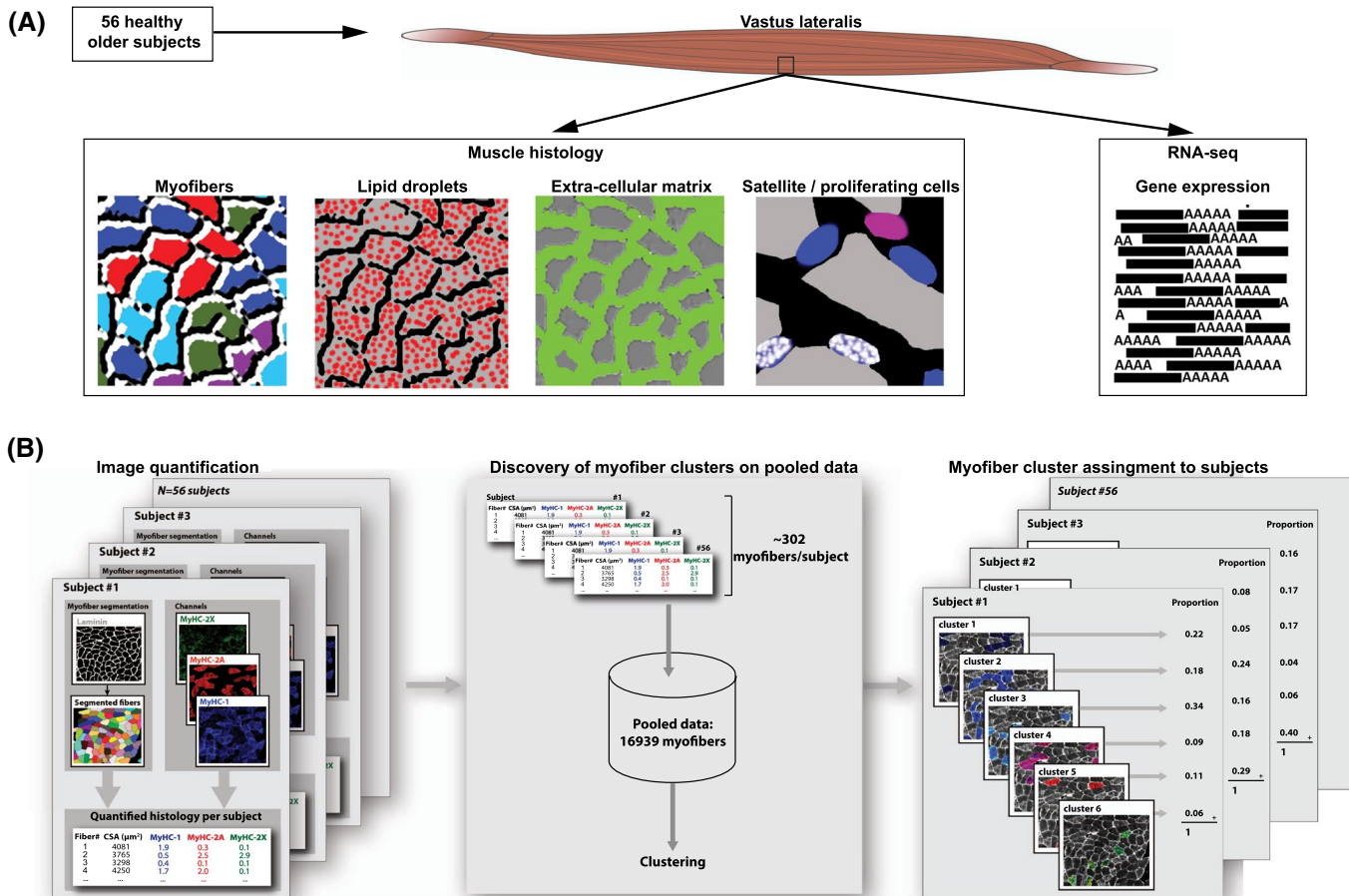
FIGURE 1 Studying myofiber composition with a data-driven procedure. A, A summary of the procedure that was used in this study: vastus lateralis muscle biopsies were collected from 56 healthy older subjects. From each muscle biopsy, cryosections were made for histological investigations and for gene-expression using RNAseq. B, A summary of the data-driven myofiber analysis procedure. Left panel: Human vastus lateralis muscle cross-sections were stained with a mixture of antibodies to MyHC-1, MyHC-2A, and MyHC-2X isoforms and laminin. Individual myofibers were segmented based on laminin, and from each myofiber mean fluorescent intensities (MFI) for each MyHC isoform and the cross-sectional surface area (CSA) were obtained. Myofiber data was collected per subject (N = 56). Middle panel: Myofiber data from all subjects was pooled (N = 16 393 myofibers), and clustering was performed on the pooled data. Right panel: Each myofibers was assigned to a myofiber cluster, per individual. The proportion of myofibers for each cluster in each subject was calculated for further analyses

of sufficient and quantity for histological evaluation. Clinical characteristics of the participants are summarized in Supporting Information Table S1. All participants of the GOTO study signed a written informed consent for participation in the trial. The trial was performed in accordance with all the local relevant guidelines and regulations and adheres to the declaration of Helsinki and its later amendments. The GOTO study was registered at the Dutch Trial Register (NTR3499, www.trial register.nl) and was approved by the medical ethical committee of the Leiden University Medical Center.

2.2 | Collection and sectioning of muscle biopsies

Participants came in the morning after at least 10 hours of fasting and consumed a standardized liquid meal (Nutridrink,

Nutricia Advanced Medical Nutrition, Zoetermeer, The Netherlands). Approximately 45 minutes after the standardized liquid meal, biopsies were collected from the *vastus lateralis* muscle (VL). Biopsies were collected on the lateral side of the upper leg, 10 cm cranial of the patella, using a 3 mm biopsy needle. The procedure was performed under local anesthesia. The obtained biomaterial was immediately frozen in liquid nitrogen and stored at −80°C prior to cryosectioning, as previously described.[23]

2.3 | RNA isolation

Muscle biopsies were homogenized with the Beadbug homogenizer (Sigma-Aldrich, St. Louis, MO, USA) using glass beads. Subsequently RNA was isolated with RNA-Bee (Tel-Test, Friendswood, TX, USA) and the NucleoSpin RNA XS

Kits (Macherey-Nagel, Düren, Germany) according to the manufacturers' instructions. RNA concentrations were determined on the NanoDrop 2000 (Thermo Fisher Scientific, Breda, The Netherlands), and RNA quality (all samples were of good quality with RNA Integrity Number > 7.0) was assessed on the Agilent 2100 Bioanalyzer system (Agilent Technologies, Amstelveen, The Netherlands).

2.4 | RNA library preparation, sequencing, and alignment

RNA-sequencing (RNA-seq) library preparation, sequencing, and RNA-seq data generation was performed by the Human Genotyping facility (HugeF) of the ErasmusMC, the Netherlands, as described in the protocols of the Biobank-Based Integrative Omics Studies (BIOS) consortium.[24] In short, libraries were made from total muscle RNA using Illumina TruSeq version 2 library preparation kits. With the Illumina HiSeq 2000 platform, paired-end sequencing reads (2 × 50-basepairs) were generated, with 10 pooled samples per lane. Data processing were performed using the in-house BIOPET Gentrap pipeline, as previously described.[25] In short, low quality trimming was performed using sickle version 1.200 ("se" "-t" "sanger"). Adapter clipping was performed using cutadapt version 1.1 ("-m" "25"). Reads were aligned to GRCh37, while masking common SNVs in the Dutch population[26] MAF > 0.01), using STAR version 2.3.0e ("—outSAMstrandField" "intronMotif" "—outSAMunmapped" "Within" "—outFilterMultimapN-max" "5" "—outFilterMismatchNmax" "8"). Sam to bam conversion and sorting was performed using Picard version 2.4.1. Read quantification was performed using htseq-count version 0.6.1p1 ("—format" "bam" "—order" "pos" "—stranded" "no") using Ensembl gene annotations version 71 for gene definitions. Sequencing resulted in an average of 36.7 million reads ($\pm$ 5.9 million reads) per sample, and 98.0% ($\pm$1.3%) was successfully mapped. one sample did not pass RNA-seq quality control steps and was, therefore, excluded from the respective analyses. The respective analyses with RNA-seq data were carried out with N = 55 samples. Genes with an expression under one count per million (lof 2) were excluded. For downstream analyses read counts were scaled using calcNormFactors of the edgeR package, logCPM transformed using voom of package limma. Structural muscle genes are defined as the gene ontology term "muscle contraction" (GO:0006936), excluding the "regulation of muscle contraction" (GO:0006937).

2.5 | Tissue histology

Cross-sectional cryosections (16 μm thick) of muscle samples were made from 56 subjects with the CM3050-S cryostat. Slides were stored at −20°C prior to staining. All stainings were carried out without fixation. Three antibodies, each specific to one particular mysosin heavy chain (MyHC) isoform, were combined with anti-laminin into a mixture, as detailed in Riaz et al.[19] In brief, cryosections were stained with rabbit anti-laminin (1:1000, Abcam, Cambridge, UK) and mouse anti-6H1 (1:5, Developmental Studies Hybridoma Bank (DSHB), Iowa City, Iowa, USA), detecting MyHC-2X. Subsequently, secondary antibodies goat anti-rabbit-conjugated-Alexa Fluor 647 (1:1000, Molecular Probes, Life Technologies) and goat anti-mouse-conjugated-Alexa Fluor 488 (1:1000, Molecular Probes, Life Technologies) were applied. Lastly, conjugated monoclonal antibodies were applied, BA-D5-conjugated-Alexa Fluor 350 (1:75, DSHB) and SC-71-conjugated-Alexa Fluor 549 (1:700, DSHB), detecting MyHC-1 and MyHC-2A, respectively. A representative image for each MyHC channel separately is shown in Figure S4. Confirming other studies,[13,19,27] we found no overlap between the MyHC isoforms.

The intramyocellular lipid droplets, sections were stained with 1μM of Nile Red (Sigma-Aldrich, Saint Louis, Missouri, USA) for 15 minutes at room temperature, as previously described.[28] To detect collagen, muscle satellite cells, and proliferating cells, immunostaining was performed. Immunostaining was performed with primary antibody goat anti-collagen1 (1:400, SouthernBiotech, Birmingham, Alabama, USA) and secondary rabbit anti-goat-conjugated-Alexa Fluor 488 (1:1000, Life Technologies, Eugene, Oregon, USA), primary antibody mouse anti-Pax7 for satellite cells (1:75, DSHB) detected by secondary goat anti-mouse-Alexa Fluor 594 (1:1000, Life Technologies) and a conjugated antibody rabbit anti-Ki67-conjugated-Alexa Fluor 647 for dividing cells (1:100, Cell Signaling Technology, Danvers, Massachusetts, USA), respectively. Nuclei were counterstained with 4',6-di-amidino-2-phenylindole (DAPI) (0.5 μg/mL, Sigma-Aldrich). Phosphate-buffered saline (PBS) containing 0.05% Tween (PBST) was used for washing sections between antibody incubations, 5% nonfat milk powder (FrieslandCampina, Amersfoort, The Netherlands) in PBST was used as a blocking agent. Slides were mounted with Aqua Poly-Mount (Polysciences Inc, Niles, Illinois, USA). Representative images are shown in Figure S5. During staining, imaging, and quantification, researchers were blinded for outcome measures.

2.6 | Imaging

Images were made with a DM5500 fluorescent microscope (Leica, Wetzlar, Germany) using LAS AF (Leica) software V2.3.6. For Nile Red and Pax7/Ki67 staining, a total of five representative fields with a 20x objective was obtained for each section. Nile Red was detected with the Y3 cube. DAPI was detected with the A4 cube, Pax7 with the TXR cube and Ki67 with the Y5 cube. For the collagen staining, three fields

per section were obtained with the 10x objective. For the fiber typing staining, images of the entire muscle section were made with a 5x objective, ensuring no overlap between fields. Laminin was detected with the Y5 cube, MyHC-2A with the TXR cube, MyHC-2X with the L5 cube, and MyHC-1 with the A4 cube. Fields obtained with the 20x objective each included an average of 29 ($\pm$10) muscle fibers, and fields obtained with the 5x objective included each an average of 184 ($\pm$56) muscle fibers. Representative images are shown in Figure S5. Images were taken from 56 subjects, unless other noted.

2.7 | Image quantification

Computational image quantification was carried out with STACKS macro V1.1 (adopted to human tissue) (previously described in Raz et al[21] and Riaz et al[19]). MyHC MFI measurement: individual fibers were automatically segmented based on laminin staining. Each image was visualized, and incorrect segmentations were manually removed. Subsequently the cross-sectional area (CSA) and corresponding MFI values for each MyHC isoform were obtained per myofiber. MFI values were corrected for background for each section and normalized for the fluorophore emission. CSAs in pixels were converted to μm^2. Intramyocellular lipid droplets area was measured by segmenting the Nile Red signal, as described previously.[19] Only 54 subjects were included as staining failed in two samples. A representative image is shown in Figure S5. The area of the extracellular matrix was measured from collagen-1 segmentation, as described previously.[19] Visual assessments of the Nile Red, collagen and cell counts were carried out by two independent researchers, after which the results were averaged prior to downstream analyses. Interclass correlations coefficients for these measurements varied from 0.80 to 0.91, signifying a strong agreement between the two researchers. A visual fiber typing was also performed: fibers were assigned to either MyHC-1, MyHC-1/2A hybrids, MyHC-2A, or fibers containing MyHC-2X. An average of 472 ($\pm$249) fibers were included per subject. The proportions of Pax7+ or Ki67+ nuclei were manually counted with an overlay image of the counterstained nuclei in respective channels. All manual counting were carried out by two independent investigators, and the intra-class correlation coefficients between the two investigators varied from .80 to .91 for the collagen area and the PAX7 and Ki67 counts. Representative images are shown in Figure S5. For statistical analyses, the proportions were calculated per subject.

2.8 | Data-driven myofiber composition

Data-driven myofiber composition is based on a density-based clustering approach on the joint distribution of the

three MyHC-isoform quantifications and is performed on the pooled measurements of all fibers in all samples. For this purpose, background-corrected and fluorophore-normalized MFI values (as done previously)[19] of myofibers with CSA $\geq$ 500 μm^2 (average: 302 $\pm$ 148 myofibers per sample) were first scaled per sample (without centering) and subsequently transformed (natural logarithm($x + 1$). 3D density estimates were computed using the Kernel smoothing (ks) package and, -for visualization purposes only-, were plotted using misc3d package, with an inner and outer density shell comprising the 25% and 55% most dense areas, respectively. Assignment of fibers to fiber clusters was performed using the mean-shift algorithm implemented in the LPCM package, a density-based clustering approach. After assignment of fibers to one of the six observed fiber clusters, proportions of the total fibers per fiber cluster were calculated for each individual. Assessments comparing CSA differences between myofiber clusters were stratified between males and females.

2.9 | Statistics

All statistical analyses were performed in R (version 3.3.2). Heatmaps were generated with R (heatmap.2 package). Boxplots, scatterplots, and bar charts were generated with Prism 6 (Graphpad Software, La Jolla, California, USA). Correlation and association analyses (Pearson correlation) with myofiber proportions were performed after applying a logit transformation using a 0.1% pseudo count. P-values < .05 were considered statistically significant for all performed analyses. A co-expression network of the structural genes was computed using the Pearson correlation. Positive correlations with $r >= .4$ were visualized using Gephi, yielding three distinct gene modules.

3 | RESULTS

3.1 | Multivariate identification of myofiber clusters

A flowchart summarizing the multivariate identification of myofiber clusters is provided in Figure 1B. A semi-automatic image processing procedure for measurement of the MFI for each MyHC isoform from each individual myofiber was previously reported in mouse muscles.[19,21] Here we applied the same procedure on human muscles that were stained for MyHC-2A, MyHC-2X, and MyHC-1. In total, 56 elderly VL muscle samples were used in this study (Figure 1B, step 1). On average, 302 myofibers were measured per subject, amounting to a total of 16 939 myofibers in the whole dataset. Although our image analysis and quantification are highly automated, we manually removed poorly segmented

myofibers. Yet, we cannot exclude that few myofibers were poorly segmented. For the data-driven detection of myofiber types, all fibers across all samples were pooled to maximize the power for detecting distinct myofiber types, ensuring that the obtained results apply to all samples in our study. An unsupervised clustering of the pooled dataset on the basis of every two MyHC isoforms combinations resulted in multiple clusters (Figure S1). Therefore, we next applied an unsupervised clustering of the pooled dataset with all three MyHC isoforms. This resulted in the discovery of distinct clusters of myofibers (Figure 1B, step 2). For statistical analysis of those myofiber clusters, for each subject we computed the proportions of myofibers that were assigned to each of the six discovered myofiber clusters (Figure 1B, step 3). The proportions of myofiber clusters per subject were used for the remaining analyses in this paper.

An unsupervised clustering of the pooled myofiber dataset using the MFI values from the three MyHC isoforms revealed six clusters, with the vast majority of myofibers to be assigned to one of the six clusters (Figure S2A). Myofibers assigned to the same clusters had highly similar MFI values, and each cluster was characterized by a distinct combination of MFIs of the MyHC isoforms (Figure 2A,B). The MFI characteristics of each cluster are shown in Figure S2B,C. The largest cluster (#1), consisted 28.6% of all fibers, characterized by high MFI values for all three MyHC isoforms (Figure 2B, cluster #1). Cluster #2 had 19.3% of all myofibers, characterized by high MFI for both MyHC-2X and in MyHC-1 but low in MyHC-2A (Figure 2B). Hence, nearly 50% of all myofibers were included in the first two clusters, both being characterized by a high MyHC-2X. Cluster #3, with 15.0% of all myofibers, had on average relatively high MyHC-2A MFI (Figure 2B), yet, the elongated shape of the cluster implies a wide range of MyHC-2A MFI values (Figure 2A). Cluster #4 (13.1%) was high in MyHC-1 MFI, and cluster #5 (13.1%) had low MFI for all three MyHC isoforms (Figure 2B). The smallest cluster (#6, 10.8%) contained both MyHC-2A and MyHC-1. In summary, two clusters were predominantly high
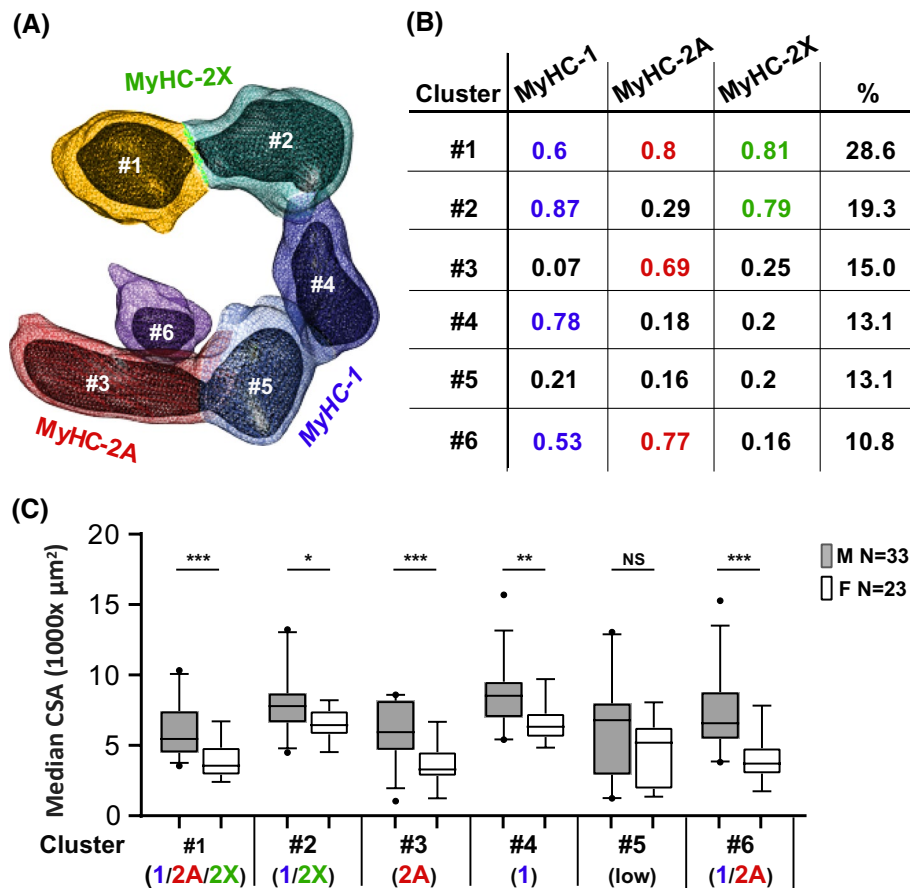
(A)

(B)

Cluster	MyHC-1	MyHC-2A	MyHC-2X	%
#1	0.6	0.8	0.81	28.6
#2	0.87	0.29	0.79	19.3
#3	0.07	0.69	0.25	15.0
#4	0.78	0.18	0.2	13.1
#5	0.21	0.16	0.2	13.1
#6	0.53	0.77	0.16	10.8

(C)

FIGURE 2 Human muscles are characterized by six main myofiber clusters. A, A 3D representation of the main six clusters that were obtained from unsupervised clustering. Per cluster, the inner (darker) and outer (lighter) density shells constitute the 25% or 55% of the myofibers, respectively. A darker density shell represents myofibers with the most similar MyHCs MFI combination. B, The table shows the mean MFI of each cluster for the three MyHC isoforms, and the proportion of myofibers assigned to each cluster in the pooled dataset. Highlighted in color are the higher expressed MyHC isoforms. C, Boxplots show the median cross-sectional area (CSA) (in um^2) per myofiber cluster on the x-axis (as called in Figure 2B). Plots are stratified for sex (male (M) is depicted in gray and female (F) in white). Boxes show the quartiles and whiskers show the 95% of the data. Outliers are signified with dots. Significance is noted with asterisks (NS: not significant, *$P < .05$, **$P < .01$, ***$P < .001$)

in only one MyHC isoform (#3 and #4), two clusters were high in two isoforms (#2, #6). The cluster low in all MyHC isoforms (#5) could also represent myofibers expressing a MyHC isoform that was not included in this study. The cluster high for all three MyHC isoforms (#6) was has not been unreported thus far. We then investigated a biological relevance for these clusters.

As myofiber types were reported to differ in size, that is, myofibers expressing MyHC-2A are smaller than those expressing MyHC-1 or MyHC-2X,[29] we assessed whether the data-driven clusters differ in the mean CSA. In accordance with the literature,[30] we found that males had larger CSA than females, for all myofiber clusters, except cluster #5 (Figure 2C). Moreover, the myofiber clusters containing MyHC-2A (#1, #3 and #6) had larger CSA and showed more prominent sex-related differences (Figure 2C). This suggests that expression of MyHC-2A could contribute to sex-related muscular differences.

3.2 | A visual myofiber typing does not capture all different myofiber clusters

As a visual assessment of myofiber typing is the current standard procedure to describe myofiber type composition, we then compared the results obtained with our new data-driven method with those obtained by a standard visual assessment. Whereas myofiber MFI was measured from each imaging channel separately, the eye-based assessment was made on the overlay image, otherwise myofiber hybrids could not be determined. By eye, four classes of fiber types could be distinguished: type-1 (blue), type-2A (red), type-1/2A hybrids (purple), and type-2A/hybrids 2X (yellow) (Figure 3A). To compare between the eye-based myofiber type scoring and data-driven clusters we investigated correlations between proportions of myofiber clusters and eye-based proportions. Additionally, in few overlay images we created a spatial localization of the data-driven clusters (an example is found in Figure 3A). We noted that myofibers in clusters #2, #4, and #5 appeared all as blue (MyHC-1 positive, Figure 3A), However, the #5 cluster showed as dark blue, whereas #2 and #4 were shown as light blue (Figure 3A). We did not distinguish between the light and dark hues and considered all blue myofibers as one type. Myofibers in cluster #2 or #1 appeared as red (MyHC-2A) or green/red (yellow in the overlay image) (MyHC-2A/2X) (Figure 3A). Correlations between the data-driven clusters and the eye-based myofiber groups revealed that the visual type-1 (blue) proportions highly correlated with the proportions of the data-driven MyHC-1 positive clusters sum (clusters #2, #4 and #5) (R^2: .68, Figure 3B). A detailed analysis showed that the visual MyHC-1 mostly correlated with clusters #2 (highest in MyHC-1 and also expresses MyHC-2X) (Figure

3C). Cluster #4 did not correlate with the visual blue cluster (Figure 2D), possibly because in few subjects the proportion of cluster #4 was very low, whereas the MyHC-1 positive fibers were very high in the visual assessment (Figure 3C-E, arrowhead). This implies that fiber typing by visual assessment does not capture the full spectrum of MyHC-1 expression variation. A low correlation was found between cluster #5 and the visual blue myofiber group. (Figure 3E), which could result from the fact that we combined light and dark blue in a single group or mixing black myofibers (negative to all MyHC isoforms) with dark blue.

The proportions of the visual type-1/2A group significantly correlated with the proportion in data-driven cluster #6 (MyHC-1/2A) (R^2: .10) (Figure 3F). No association was found between the visual type-2A and cluster #3 (high in MyHC-2A), as high proportions in the visual assessment got low proportions in the data-driven classification (Figure 3G). This could be explained by the wide MFI range in cluster #3 (Figure 3A), or misclassification of red and purple myofibers.

Last, no association was found between the visual type-2A/2X group and cluster #1 (high in all three MyHC isoforms) (Figure 3H). This could imply that the visual type-2A/2X group was misclassified. As expected, our analysis showed that no or low correlations were mostly found for the hybrid myofibers, suggesting that misclassification of myofibers is more prone to occur for hybrids. Overall, this analysis shows that a visual assessment of fiber groups cannot capture all myofiber clusters and a classification derived from this assignment is probably inaccurate as compared to a data-driven myofiber assignment.

3.3 | Myofibers clusters are associated with distinct gene expression profiles

In order to assess if molecular signatures of the muscle biopsy are echoed in those myofiber clusters, we investigated the correlation between the mRNA expression levels of sarcomeric genes and the proportions of the myofiber clusters. As RNA-seq data on muscle tissue also captures cells, other than myofibers,[31] we focused our analyses on sarcomeric genes (N = 71) that were expressed (>1 count per million) in our data. Those genes, including the genes encoding for MyHC-1, MyHC-2A, and MyHC-2X, mark muscle contractile properties.

A co-expression analysis on the 71 sarcomeric genes identified three distinct gene modules. (Figure 4A). The green module was enriched for fast-twitch genes, such as MYH1 (MyHC-2X), MYH2 (MyHC-2A), MYH4, and TNNi2, whereas the blue module was enriched for slow-twitched genes, including MYH7 (MyHC-1, TNNi3, and TNNT1). A third less tightly intercorrelated module, the orange module,
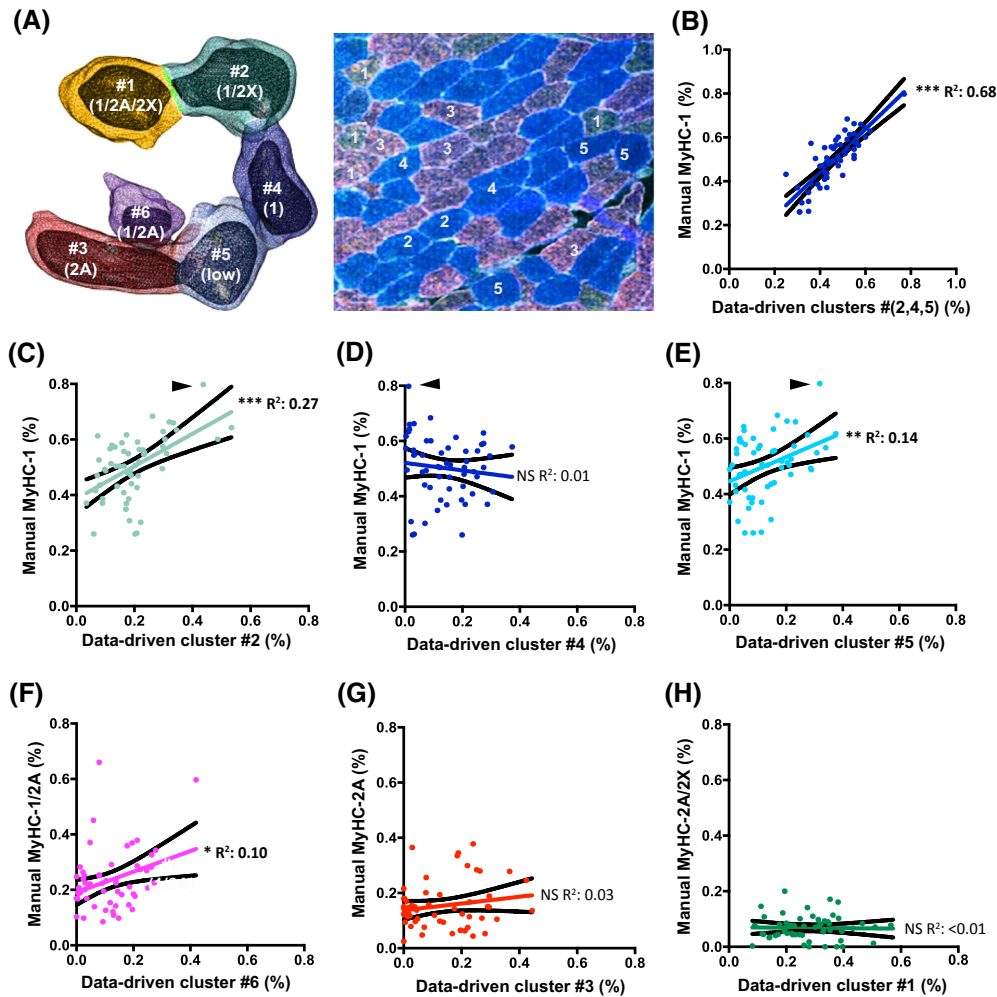
FIGURE 3 Associations between myofiber clusters and visual myofiber typing. A, A 3D clustering highlighting the clusters that were assessed in panels B-D. A MyHC-stained overlay image of a muscle cross-section with the assignment to the clusters (blue: MyHC-1, red: MyHC-2A, green: MyHC-2X, white: laminin). B, A correlation between proportions for clusters #(2,4,5) (identified in panel A as MyHC-1) and the visual MyHC-1 proportions. C-E, Correlations between myofiber proportions in clusters #(2,4,5) with visual MyHC-1. The arrowhead denotes the same individual. F, Correlation between proportions of cluster #6 with visual assigned proportions of MyHC-1/2A hybrids. G, Correlation between proportions of cluster #3 with visual assigned proportions of MyHC-2A. H, Correlations between proportions of cluster #1 with visual assigned proportions of MyHC-2A/2X hybrids. Associations are between the proportions in data-driven fiber type clustering and the corresponding visual proportions. Each dot represents a subject (N = 56). The colored line denotes the linear regression and the black lines represent 95% confidence interval of the linear regression. Statistical significance: NS: $P \geq .05$, *$P < .05$, **$P < .01$, ***$P < .001$. Goodness of fit is noted with the R^2 of the regression

was enriched for non-myogenic myosin Gene-Ontology term (Figure 4A).

Out of the 71 sarcomeric genes, 37 (52%) showed at least one significant correlation ($P <= .05$) with the proportion of a myofiber cluster. As expected, genes with the higher expression levels were found to have higher significant correlations (Figure S3). We observed that the majority of genes in the blue or green networks showed a significant association with one or more of the myofiber clusters (Figure 4A). In contrast, only nine genes in the orange network were found to be associated with myofiber clusters (Figure 4A).

As expected, genes from the same module exhibited similar correlation patterns with proportions of myofiber clusters

(Figure 4B). We employed hierarchical clustering of the "gene expression—myofiber cluster" correlations to group genes on the basis of similar associations with proportions of myofiber clusters and identified three groups. Group I mostly consisted of fast-twitch genes (green module, Figure 4B). As expected, expression of group I genes positively correlated with cluster #3 proportions (high in MyHC-2A), and inversely correlated with cluster #2 proportions (high in MyHC-1 and MyHC-2X) (Figure 4B). Genes of group II were all part of the orange module, and were positively correlated with cluster #1 proportions (high in all three MyHC isoforms) and cluster #2 proportions (high in MyHC-1 and MyHC-2X), and inversely correlated with cluster #3 proportions (high in MyHC-2A)
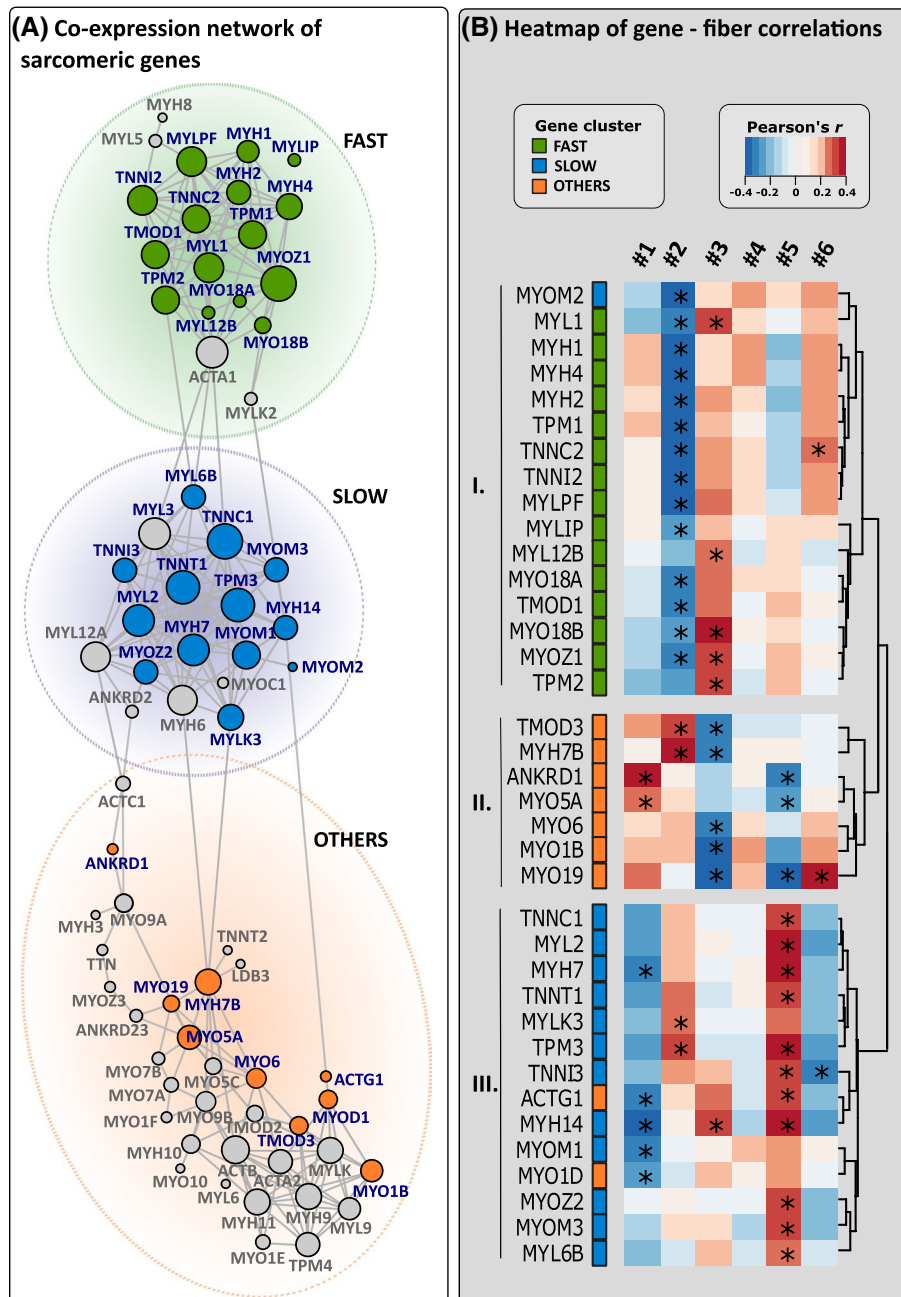
FIGURE 4 mRNA expression of sarcomere genes reflects fiber proportions. A, A co-expression network analysis on 71 sarcomere genes revealed three gene modules. The size of a gene in the co-expression network is proportional to their connectivity degree, (the number of connections with other genes). A connection is drawn for Pearson correlations > .5. Genes are assigned to a cluster on the basis of a modularity clustering. Gene-modules were computed and are indicated with large circles. The green and blue gene-modules are enriched by fast or slow-twitch genes, respectively. The orange module is enriched by other/non-myogenic genes. Genes colored in gray did not have significant correlations with any of the proportions of myofiber clusters (in panel B). B, A heatmap of correlations between gene expressions (columns) and proportions of myofiber clusters (columns) across 55 participants. Only genes with at least one significant correlation with the proportion of myofiber cluster (*P* < .05, denoted with an asterisk) are shown. Hierarchical clustering of correlations is depicted on the right side. Genes were annotated based on the gene modules found in panel A. Based on the dendrogram three groups of genes were identified that have a similar pattern of correlations with myofiber cluster proportions

and cluster #5 proportions (low in all three MyHC isoforms) (Figure 4B). Group III mostly included slow-twitch genes (blue module, Figure 4B), and were positively correlated with cluster #5 proportions (low in all three MyHC) and

inversely correlated with cluster #1 proportions (high in all three MyHC isoforms). Especially the many positive correlations with cluster #5 proportions is striking when considering that myofibers in cluster #5 have a low protein expression.

In view of the positive correlation with the expression of especially slow-twitch genes, this could suggest that cluster #5 myofibers are myofibers in transition to become "slow" myofibers.

Whereas four out of six identified myofiber clusters exhibited significant correlations with genes from one or multiple modules, none or little correlations were observed with proportions in cluster #4 (high in MyHC-1) or cluster #6 (high in MyHC-1 and MyHC-2A). Collectively, this correlation analysis suggests that four of the data-driven myofiber clusters are marked by distinct sarcomere gene expression signatures, suggesting that variations in the proportions of the discovered clusters are accompanied by specific transcriptional changes.

Collectively, this correlation analysis suggests that four of the data-driven myofiber clusters are marked by distinct sarcomere gene expression signatures, suggesting that variations in the proportions of the discovered clusters are accompanied by specific transcriptional changes.

3.4 | Correlations of myofiber clusters with muscle health measures

To further assess a possible biological relevance for cluster #1 (high expression for all three MyHCs), we investigated

a correlation between its proportions and histological measures of aging-associated muscle health. Extracellular matrix (ECM) thickening and fatty infiltration by the accumulation of lipid droplets[28,32] are known to increase in aging. In addition, the ability for regeneration following injury marks healthy muscles.[33] This can be assessed by the presence of active satellite cells (marked by the combination of the satellite cell marker, PAX7, and the cell division marker, Ki67[34]). Therefore, we performed multiple histological stainings and correlated the outcomes with myofiber cluster proportions. Accordingly, we found that cluster #1, (high in all three MyHC isoforms) had significant correlations with three out of the four assessed muscle health parameters (Figure 5A), namely it was positively correlated with Collagen-1 area and negatively correlated with percentage of PAX7- or Ki67-positive nuclei (Figure 5B,C). Hence, our results suggest that variations in the proportions of cluster #1 (high in all three MyHC isoforms) might be key for studying the heterogeneity in muscle health.

4 | DISCUSSION

Contraction properties of skeletal muscles are predominantly determined by myofiber types. Traditionally, studies into
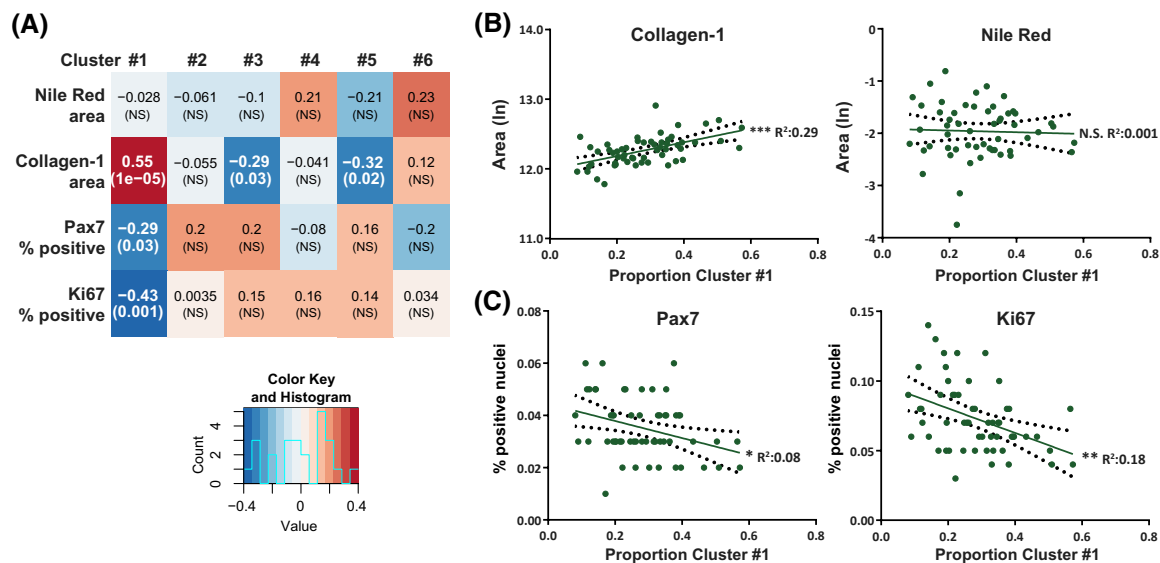

FIGURE 5 Associations between proportions in the largest cluster (#1; high for all three MyHC isoforms) with features of muscle health. A, Heatmap representing the Pearson correlations between proportions of myofiber clusters (columns) and histological parameters of muscle health (rows). Red represents positive correlations and blue negative correlations. The statistical significance of the correlation (p-value) is shown in brackets. B, Scatter plots of correlations with proportions of Cluster #1 and features of muscle fibrosis. Left panel: a correlation between collagen-1 mean fluorescence intensity (ln) and the proportion of myofibers in cluster #1 (N = 56 subjects). Right panel: a correlation between Nile Red mean fluorescence intensity (ln) and the proportion of myofibers in cluster #1 (N = 54 subjects). Each dot represents a subject (N = 56). The green solid line denotes the linear regression and the black dotted lines represent 95% confidence interval of the linear regression. Statistical significance: NS: $P \geq .05$, $*P < .05$, $**P < .01$, $***P < .001$. Goodness of fit is noted with the R^2 of the regression. C, Scatter plots of correlations with proportions of Cluster #1 and features of muscle regeneration. Left panel: a correlation between the proportion of Pax7 positive nuclei and the proportion of myofibers in cluster #1 (N = 56 subjects). Right panel: a correlation between the proportion of Ki67 positive nuclei and the proportion of myofibers in cluster #1 (N = 56 subjects)

decline of muscle health among older people focus on two myofiber types, fast- and slow- twitch, which are marked by the expression of MyHC fast or slow MyHC isoforms. In this methodological paper, we applied a high-throughput image quantification of MyHC isoforms MFI. We identified six distinct myofiber clusters in VL muscles derived of 56 healthy elderly subjects. We found that the proportions of identified myofiber clusters are associated with molecular features, suggesting their biological significance. Specifically, we report on two novel clusters in muscles for healthy elderly: a cluster high in all three MyHC isoforms (cluster #1) and a cluster low in all three MyHC isoforms (cluster #5), that potentially could contribute to a better understanding of muscle biology.

The first novel cluster (#1) was defined by a relatively high protein expression of all three MyHC isoforms. Strikingly, cluster #1 was also the most abundant myofiber cluster detected in our study of healthy elderly subjects (28,6%), implying that it either was missed in previous studies, that is, due to the limitations of visualization assessments, or that these fibers are specific to the healthy elderly. Hence, it remains to be determined whether this cluster can be found in the muscles of younger individuals. Hybrid myofibers, that is, myofibers expressing more than one MyHC isoform, have been reported to play a role in denervation and/or regeneration conditions.[20,35] Accordingly, we found that cluster #1 correlated with several measures of muscle health. Higher proportions of cluster #1 correlated with reduced proportions of satellite cells, reduced proportions of proliferating cells and a thickening of the extracellular matrix. As none of the other myofiber clusters showed such overwhelming correlations with histological measures of muscle health, this suggests that variations in the proportion of cluster #1 might be key to studying the heterogeneity in muscle health.

The second novel cluster (#5) was defined by a relatively low protein expression of all three MyHC isoforms. Despite the low MyHC protein levels, we found high correlations with mRNA expression of genes encoding for slow-twitch fibers. This could suggest that these are myofibers in transitions to become "slow" myofibers. It is notable that myofiber clusters with a high protein expression levels of one, two or three MyHC isoforms showed, in general, less correlations with mRNA expression levels. Although in young animals a correlation between mRNA and protein levels of sarcomeric genes,[19,36] in elderly and in a muscle aging mouse model only limited correlation was reported between mRNA and protein levels.[37,38] Abundance of sarcolemma proteins results, in part, by reduced proteasomal activity.[8] Proteasomal activity differs between myofiber types,[39] and is reduced during aging.[19,40]

Myofiber composition seems more complex in human than in mouse muscle. Myofiber transitions by the method applied here were monitored in a mouse model for muscle aging and a model for limb girdle.[19,21] The pattern of

clusters differed between mouse diaphragm and quadriceps.[21] Myofiber composition seems more complex in human than in mouse muscle. Myofiber transitions by the method applied here were previously monitored in a mouse model for muscle aging and a model for limb girdle.[19,21] Whereas the pattern of clusters differed between mouse diaphragm and quadriceps,[21] at most two major myofiber clusters were identified using any combination of two MyHC isoforms in mouse.[21] In contrast, we observe multiple clearly defined clusters in human data in this setting. This may point to fundamental differences in muscle contraction capacity between similar muscles of different species. Accordingly, fiber excursions in human and mouse 26 limb muscle groups were previously shown to be significantly different for 22 muscles.[41] This, in part, can explain the low correlations we found for some of the myofiber clusters.

A possible limitation of our study relates to the use of statistical correlations for suggesting a biological meaning or importance of the discovered myofiber clusters. Ideally, future studies in muscular disorders or interventions, complemented by biomechanical measurements, are performed to independently reinforce the function of the identified myofiber clusters. Moreover, it remains to be studied whether all identified myofiber clusters contribute to muscle contraction and function and whether this refined fiber typing can assist in understanding heterogeneity in muscle health at older age, in monitoring interventions aimed at increasing muscle health and their relation to overall health. Nevertheless, we believe that a data-driven analysis of myofiber clusters, as opposed to the traditional visual-based single isoform analysis, opens novel and exciting opportunities to improve our understanding of muscle biology.

To conclude, we show that a computational biology analysis using a large myofiber dataset, compiled from a relatively large number of elderly leads to better understanding of the complex muscle heterogeneity. The presence of the six distinct myofiber clusters is reinforced by correlations with mRNA expression levels and with histological features of muscle health. Moreover, we show that data-driven analysis is more accurate in capturing the heterogeneity of myofibers. Together we underline the viewpoint that data-driven analysis of myofiber clusters opens opportunities for an improved understanding of muscle biology. This methodological study opens avenues for further comparative studies in muscle aging and muscle diseases.

CONFLICT OF INTEREST

All authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Y. Raz, E.B. van den Akker, V. Raz, M. Beekman, E.J.M. Feskens, and P.E. Slagboom, Conceptualization; Y. Raz, E.B. van den Akker, M.J.T. Reinders, V. Raz, M. Riaz, and J. Goeman, Methodology; E.B. van den Akker, Software; Y. Raz and E.B. van den Akker, Formal Analysis; Y. Raz, T. Roest, and H.E.D. Suchiman, Investigation; S.A. Stefanie, V. Raz, O. van de Rest, and M. Riaz, Resources; Y. Raz, E.B. van den Akker, T. Roest, and M. Beekman, Data Curation; Y. Raz, Writing – Original Draft; E.B. van den Akker, V. Raz, M. Beekman, and P.E. Slagboom, Writing – Review & Editing; Y. Raz and E.B. van den Akker, Visualization; M. Beekman and P.E. Slagboom, Supervision; P.E. Slagboom and E.J.M. Feskens, Funding Acquisition.

REFERENCES

1. Boros-Hatfaludy S, Fekete G, Apor P. Metabolic enzyme activity patterns in muscle biopsy samples in different athletes. *Eur J Appl Physiol*. 1986;55:334-338.
2. Engel WK. Selective and nonselective susceptibility of muscle fiber types: a new approach to human neuromuscular diseases. *JAMA Neurol*. 1970;22:97-117.
3. Gehlert S, Weber S, Weidmann B, et al. Cycling exercise-induced myofiber transitions in skeletal muscle depend on basal fiber type distribution. *Eur J Appl Physiol*. 2012;112:2393-2402.
4. Giacomelli F, Bibbiani C, Bergamini E, Pellegrino C. Two ATPases in the sarcoplasmic reticulum of skeletal muscle fibres. *Nature*. 1967;213:679-682.
5. Midrio M, Danieli-Betto D, Megighian A, Velussi C, Catani C, Carraro U. Slow-to-fast transformation of denervated soleus muscle of the rat, in the presence of an antifibrillatory drug. *Pflügers Archiv*. 1992;420:446-450.
6. Ogata T, Yamasaki Y. Ultra-high-resolution scanning electron microscopy of mitochondria and sarcoplasmic reticulum arrangement in human red, white, and intermediate muscle fibers. *Anat Rec*. 1997;248:214-223.
7. Ciciliot S, Rossi AC, Dyar KA, Blaauw B, Schiaffino S. Muscle type and fiber type specificity in muscle wasting. *Int J Biochem Cell Biol*. 2013;45:2191-2199.
8. Nigro V, Piluso G. Spectrum of muscular dystrophies associated with sarcolemmal-protein genetic defects. *Biochem Biophys Acta*. 2015;1852:585-593.
9. Purves-Smith FM, Sgarioto N, Hepple RT. Fiber typing in aging muscle. *Exerc Sport Sci Rev*. 2014;42:45-52.
10. Marin P, Andersson B, Krotkiewski M, Bjorntorp P. Muscle fiber composition and capillary density in women and men with NIDDM. *Diabetes Care*. 1994;17:382-386.
11. Engel WK. Diagnostic histochemistry and clinical-pathological testings as molecular pathways to pathogenesis and treatment of the ageing neuromuscular system: a personal view. *Biochim Biophys Acta - Mol Basis Dis*. 2015;1852:563-584.
12. Gregorevic P, Meznarich NA, Blankinship MJ, Crawford RW, Chamberlain JS. Fluorophore-labeled myosin-specific antibodies simplify muscle-fiber phenotyping. *Muscle Nerve*. 2008;37:104-106.
13. Sawano S, Komiya Y, Ichitsubo R, et al. A one-step immunostaining method to visualize rodent muscle fiber type within a single specimen. *PLoS ONE*. 2016;11:e0166080-e0166080.
14. Schiaffino S, Reggiani C. Myosin isoforms in mammalian skeletal muscle. *J Appl Physiol*. 1994;77:493-501.
15. Simoneau JA, Bouchard C. Genetic determinism of fiber type proportion in human skeletal muscle. *FASEB J*. 1995;9:1091-1095.
16. Klitgaard H, Mantoni M, Schiaffino S, et al. Function, morphology and protein expression of ageing skeletal muscle: a cross-sectional study of elderly men with different training backgrounds. *Acta Physiol Scand*. 1990;140:41-54.
17. Murgia M, Toniolo L, Nagaraj N, et al. Single muscle fiber proteomics reveals fiber-type-specific features of human muscle aging. *Cell Rep*. 2017;19:2396-2409.
18. Andersen JL, Gruschy-Knudsen T, Sandri C, Larsson L, Schiaffino S. Bed rest increases the amount of mismatched fibers in human skeletal muscle. *J Appl Physiol*. 1999;86:455-460.
19. Riaz M, Raz Y, van Putten M, et al. PABPN1-dependent mRNA processing induces muscle wasting. *PLoS Genet*. 2016;12:e1006031.
20. Rowan SL, Rygiel K, Purves-Smith FM, Solbak NM, Turnbull DM, Hepple RT. Denervation causes fiber atrophy and myosin heavy chain co-expression in senescent skeletal muscle. *PLoS ONE*. 2012;7:e29082.
21. Raz V, Raz Y, van de Vijver D, Bindellini D, van Putten M, van den Akker EB. High-throughput data-driven analysis of myofiber composition reveals muscle-specific disease and age-associated patterns. *FASEB J*. 2019;33:4046-4053.
22. Schutte BA, van den Akker EB, Deelen J, et al. The effect of standardized food intake on the association between BMI and (1) H-NMR metabolites. *Sci Rep*. 2016;6:38980.
23. Tiku V, Jain C, Raz Y, et al. Small nucleoli are a cellular hallmark of longevity. *Nat Commun*. 2016;8:16083.
24. Bonder MJ, Luijk R, Pital-Zhernakova D, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet*. 2017;49(1):131-138.
25. Zhernakova DV, Deelen P, Vermaat M, et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet*. 2017;49:139-145.
26. Francioli LC, Menelaou A, Pulit SL, et al. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet*. 2014;46:818.
27. Bloemberg D, Quadrilatero J. Rapid determination of myosin heavy chain expression in rat, mouse, and human skeletal muscle using multicolor immunofluorescence analysis. *PLoS ONE*. 2012;7:e35273-e35273.
28. Raz Y, Henseler JF, Kolk A, et al. Patterns of age-associated degeneration differ in shoulder muscles. *Front Aging Neurosci*. 2015;7:236.
29. Lexell J. Human aging, muscle mass, and fiber type composition. *J Gerontol A Biol Sci Med Sci*. 1995;50(Spec No):11-16.

30. Frontera WR, Hughes VA, Fielding RA, Fiatarone MA, Evans WJ, Roubenoff R. Aging of skeletal muscle: a 12-yr longitudinal study. *J Appl Physiol*. 2000;88:1321-1326.

31. Giordani L, He GJ, Negroni E, et al. High-dimensional single-cell cartography reveals novel skeletal muscle-resident cell populations. *Mol Cell*. 2019;74:609-621.e606.

32. Kragstrup TW, Kjaer M, Mackey AL. Structural, biochemical, cellular, and functional changes in skeletal muscle extracellular matrix with aging. *Scand J Med Sci Sports*. 2011;21:749-757.

33. Kim J, Braun T. Skeletal muscle stem cells for muscle regeneration. *Methods Mol Biol (Clifton, NJ)*. 2014;1213:245-253.

34. Mackey AL, Kjaer M, Charifi N, et al. Assessment of satellite cell number and activity status in human skeletal muscle biopsies. *Muscle Nerve*. 2009;40:455-465.

35. Andersen JL, Terzis G, Kryger A. Increase in the degree of coexpression of myosin heavy chain isoforms in skeletal muscle fibers of the very old. *Muscle Nerve*. 1999;22:449-454.

36. Chemello F, Grespi F, Zulian A, et al. Transcriptomic analysis of single isolated myofibers identifies miR-27a-3p and miR-142-3p as regulators of metabolism in skeletal muscle. *Cell Rep*. 2019;26:3784-3797.e3788.

37. Olie CS, Riaz M, Konietzny R, et al. Deacetylation inhibition reverses PABPN1-dependent muscle wasting. *iScience*. 2019;12:318-332.

38. Robinson MM, Dasari S, Konopka AR, et al. Enhanced protein translation underlies improved metabolic and physical adaptations to different exercise training modes in young and old humans. *Cell Metab*. 2017;25:581-592.

39. Raz V, Raz Y, Paniagua-Soriano G, et al. Proteasomal activity-based probes mark protein homeostasis in muscles. *J Cachexia Sarcopenia Muscle*. 2017;8:798-807.

40. Strucksberg KH, Tangavelou K, Schroder R, Clemen CS. Proteasomal activity in skeletal muscle: a matter of assay design, muscle type, and age. *Anal Biochem*. 2010;399:225-229.

41. Hu X, Charles JP, Akay T, Hutchinson JR, Blemker SS. Are mice good models for human neuromuscular disease? Comparing muscle excursions in walking between mice and humans. *Skelet Muscle*. 2017;7:26.

42. Raz V, Riaz M, Tatum Z, Kielbasa SM, t Hoen, PAC. The distinct transcriptomes of slow and fast adult muscles are delineated by noncoding RNAs. *FASEB J*. 2018;32:1579-1590.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.