

# High-Throughput Muscle Fiber Typing From RNA Sequencing Data

**Nikolay Oskolkov**

Lund University

**Malgorzata Santel**

Max Planck Institute for Evolutionary Anthropology

**Ola Ekström**

Lund University

**Gray J. Camp**

Max Planck Institute for Evolutionary Anthropology

**Eri Miyamoto-Mikami**

Juntendo University

**Kristoffer Ström**

Lund University

**Hemang M. Parikh**

University of South Florida

**Bilal Mir**

Lund University

**Dmytro Kryvokhyzha**

Lund University

**Mikko Lehtovirta**

Lund University

**Hiroyuki Kobayashi**

Tsukuba University Hospital

**Ryo Kakigi**

Josai International University

**Hisashi Naito**

Juntendo University

**Karl-Fredrik Eriksson**

Lund University

**Björn Nystedt**

National Bioinformatics Infrastructure Sweden, Uppsala University

**Noriyuki Fuku**

Juntendo University

**Barbara Treutlein**

Max Planck Institute for Evolutionary Anthropology

**Svante Pääbo**

Max Planck Institute for Evolutionary Anthropology

**Ola Hansson** (✉ [ola.hansson@med.lu.se](mailto:ola.hansson@med.lu.se))

Lund University

---

## Research Article

**Keywords:** muscle, expression, distribution, fiber

**Posted Date:** December 28th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1156564/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**BACKGROUND:** Skeletal muscle fiber type distribution has implications for human health, muscle function and performance. This knowledge has been gathered using labor-intensive and costly methodology that limited these studies. Here we present a method based on muscle tissue RNA sequencing data (totRNAseq) to estimate the distribution of skeletal muscle fiber types from frozen human samples, allowing for a larger number of individuals to be tested.

**METHODS:** By using single-nuclei RNA sequencing (snRNAseq) data as a reference, cluster expression signatures were produced by averaging gene expression of cluster gene markers and then applying these to totRNAseq data and inferring muscle fiber nuclei type via linear matrix decomposition. This estimate was then compared with fiber type distribution measured by ATPase staining or myosin heavy chain protein isoform distribution of 62 muscle samples in two independent cohorts (n = 39 and 22).

**RESULTS:** The correlation between the sequencing-based method and the other two were  $r_{\text{ATPas}} = 0.65$  [0.46 – 0.84], [95% CI] and  $r_{\text{myosin}} = 0.80$  [0.71 – 0.89], with  $p = 7.96 \times 10^{-6}$  and  $8.06 \times 10^{-6}$  respectively. The deconvolution inference of fiber type composition was accurate even for very low totRNAseq sequencing depths, *i.e.*, down to an average of ~5.000 paired-end reads.

**CONCLUSIONS:** This new method (<https://github.com/OlaHanssonLab/PredictFiberType>) consequently allows for measurement of fiber type distribution of a larger number of samples using totRNAseq in a cost and labor-efficient way. For the first time, it is now feasible to study the association between fiber type distribution and *e.g.* health outcomes in large well-powered studies.

## Introduction

Our bodies constitute to ~30 - 40% of skeletal muscle and it is the most abundant form of the three types of muscle, the others being smooth and cardiac. Skeletal muscle is composed of different fiber types (*i.e.*, muscle cell types), and the relative proportions of these types varies among muscles, locations within muscles, individuals, and the sex of individuals<sup>1-4</sup>. The oxidative and glycolytic potential and the contractile properties differ considerably between fiber types, with the mitochondria-rich slow-twitch fibers (Type I) having higher oxidative capacity, and fast-twitch fibers (Type IIa and Type IIx) having higher glycolytic capacity<sup>1</sup>. The proportions also change as people age, with Type II fibers being preferentially affected by sarcopenia<sup>5</sup>. Exercising skeletal muscle is a major site for catabolic metabolism of blood glucose and lipids and the metabolic characteristics of this tissue influence both the performance of elite athletes<sup>6,7</sup> and an individual's predisposition to disease, *e.g.* impairments in glucose and lipid transportation into myocytes can promote diabetes and atherosclerotic vascular disease<sup>8-10</sup>. Although environmental factors, such a training or sedentary behavior and aging, lead to adaptive alteration in capillary density and fiber composition (an increase in Type IIa vs Type IIx) in skeletal muscle, these features are partially under genetic control<sup>11,12</sup>. However, the extent to which transition of Type I to Type II fibers and vice-versa occurs remains uncertain<sup>6</sup>. Muscle tissues are usually classified according to the

predominant myosin heavy chain (MyHC) isoforms, but this is a simplified classification that disregards the large number of other proteins expressed in skeletal muscle<sup>2,13</sup>. It is thus not straightforward to estimate and compare muscle fiber types in humans, *e.g.*, owing to large heterogeneity and limitations of the methodology, and it is even more challenging in frozen, postmortem samples.

In addition to cells from differing types of muscle fibers, a biopsy sample also contains cells such as fibroblasts, endothelial cells, adipocytes, smooth muscle cells, and neuron-associated Schwann cells. Gene expression analysis of skeletal muscle is thus rendered difficult by the complexity of this tissue. Although sequencing of transcriptomes from muscle biopsies may still provide a perspective on functional differences between individuals, targeted RNA sequencing from isolated cells provides the opportunity to reveal differences specific to the different cell populations<sup>14</sup>. Previous studies investigating skeletal muscle have applied single-cell RNA sequencing (scRNAseq) to extracted mono-nucleated cells<sup>15-20</sup>. These studies have for example described a complex landscape of different cell types, *e.g.*, two different populations of muscle progenitor cells<sup>15</sup>, provided detailed knowledge concerning muscle regeneration<sup>19</sup> and muscle disease etiology<sup>20</sup>. A challenge for single-cell genomic studies in muscle and other solid tissues is to capture cell types that are difficult to isolate in suspension<sup>21</sup>, *e.g.*, muscle fibers have not been sequenced in the above mentioned scRNAseq studies. However, one study has sequenced poly-nucleated primary muscle fibers in mice and found that myofiber types predominantly express either slow or one of the fast isoforms of MyHC proteins, while only a small proportion of hybrid fibers can express more than one MyHC<sup>22</sup>. However, no study has investigated poly-nucleated primary muscle fibers from humans.

## Results

Here we present a method to estimate the proportion of skeletal muscle fiber types using only muscle tissue RNA sequencing (totRNAseq) data that could be used on frozen samples. The method is based on snRNAseq information of one human individual and then evaluated in two independent larger totRNAseq data sets of human skeletal muscle (the Muscle SATellite cell study, MSAT, n = 39 and the Juntendo Muscle Study, JMS, n = 23).

### *A novel muscle fiber type prediction model derived from single-nuclei RNA sequencing*

After filtering and quality control (see material and methods for details) the snRNAseq data set consisted of 2402 nuclei, with 15416 expressed genes. On average 200 genes were found per nuclei. A likely reason for the relatively low average gene-per-nuclei count is that the analysis is optimized for single-cell, rather than single-nuclei data and hence only reads mapping to exonic regions are counted. Three major clusters of nuclei were identified using graph-based clustering built on Louvain modularity optimization<sup>23</sup>. For visualization of the nuclei populations, t-distributed stochastic neighbor embedding (tSNE) non-linear dimension reduction was applied (Fig 1a). Examples of marker genes for the three clusters are shown in figure 1b. A separation of Type I and Type II fiber nuclei is clearly observed, *i.e.*, gene markers of Type I and Type II fibers (*e.g.*, *ATP2A2*, *TPM3*, *MYH7B* versus *ATP2A1*, *TNNT3*, *MYH2*) display

a distinct expression pattern in different clusters (Fig 1c). The identity of the remaining cluster is not clear, but likely does not consist of muscle cell nuclei, *e.g.*, indicated by low expression of the muscle marker *TTN*, while *MEG3* is found to be highly expressed (Fig 1c).

### Muscle fiber typing using total RNA sequencing from frozen samples

To evaluate the efficacy of estimating proportions of Type I versus Type II fiber nuclei in muscle samples from totRNAseq data, a deconvolution analysis<sup>24</sup> was performed in a data set consisting of 39 human subjects, *i.e.* the MSAT study (Table 1). Briefly, by using the snRNAseq data as a reference, cluster expression signatures were produced by averaging gene expression of cluster gene markers and then applying these to the totRNA MSAT data set by inferring muscle fiber nuclei type via linear matrix decomposition<sup>24</sup>, and then finally compare this estimate with the fiber type proportions measured by ATPase staining of the same muscle samples. The correlation between the estimated proportions of muscle fiber nuclei types from totRNAseq data and muscle fiber types from ATPase staining was  $r = 0.65$  [0.46 – 0.84], [95% CI] at  $p_{\text{spearman}} = 7.96 \times 10^{-6}$ ,  $n = 39$  (Fig 2a). The same deconvolution analysis was performed on a second dataset consisting of 23 human subjects from Japan, *i.e.*, the JMS (Table 2). The correlation between the estimated proportions of muscle fiber nuclei types from totRNAseq data and muscle fiber types measured by MyHC protein isoform distribution in the JMS was  $r = 0.80$  [0.71 – 0.89], [95% CI] at  $p_{\text{spearman}} = 8.06 \times 10^{-6}$ ,  $n = 22$  (Fig 2b). One individual was excluded due to low sequencing quality.

To test the possibility of implementing this method for fiber typing of a large number of samples, we estimated the needed minimal sequencing depth of totRNAseq data that accurately would infer skeletal muscle fiber type composition. Increasing the number of randomly selected reads were removed from the totRNAseq data of the 39 samples in the MSAT study. The initial average sequencing depth of 35 million paired-end (PE) reads were ‘down-sampled’ by this method, and Spearman correlations (Fig 2c) and mean square errors (Fig 2d) between the ATPase and totRNAseq predicted fractions of Type I fibers were calculated. The accuracy of deconvolution inference of fiber type composition was similar to the 35 million PE reads level even at very low sequencing depths, *i.e.*, down to an of average ~5.000 PE reads (Fig2c).

## Discussion

Here we present a method to estimate the proportion of skeletal muscle fiber types from frozen samples allowing for a larger number of samples to be measured in a standardized, cost and labor-efficient way. Skeletal muscle fiber type distribution is a determinant of physical performance<sup>25-28</sup> and overall health<sup>29-33</sup>, and is highly heritable in humans<sup>11,12</sup>. For example, a reduced proportion of oxidative slow-twitch Type I fibers is associated with lower insulin sensitivity in diabetic muscle<sup>29-32</sup> and muscle atrophy, *e.g.*, age-related sarcopenia is progressing in a fiber type-specific manner<sup>33</sup>. Recently, it has also been shown that skeletal muscle response and recovery from exercise training is dependent on fiber type

composition and is thus an important factor to consider in the development of individualized training advice<sup>34-36</sup>.

However, many of the above-mentioned associations with fiber type distribution are based on low-powered studies. One of the limiting factors is that the methodology for determining fiber type distribution is labor-intensive and thus relatively expensive. The method presented here allows for a larger number of samples to be analyzed using totRNAseq in a standardized and relatively fast way. The estimated fiber type proportion measured with the new totRNAseq based method is comparable to the ATPase based measurements, but the MyHC based method generally suggests lower Type I fiber proportions. With this new method, the distribution of Type I versus Type II fibers can be estimated, but it cannot distinguish between Type IIa and Type IIx fibers. However, this is achievable with only a small amount of muscle tissue (~5-10 mg), *e.g.*, from sampling with the minimally invasive micro biopsy technique<sup>37</sup>. The deconvolution inference of fiber type composition was accurate even for very low sequencing depths, *i.e.*, down to an average of ~5,000 PE reads. This means that a shallow-coverage totRNAseq experiment (or targeted RNAseq) will be sufficient to accurately estimate skeletal muscle fiber type composition at a low cost per sample (<1 dollar/sample). This new method consequently allows for the measurement of fiber type distribution of a larger number of samples and for the first time it is feasible to study the connection between fiber type distribution and health with well-powered studies. It can also be used for estimating fiber type distribution in public repositories of totRNAseq data (*e.g.*, Genotype-Tissue Expression project) and perform *in silico* analyses of fiber type associations. In conclusion, totRNAseq can efficiently be used to estimate skeletal muscle fiber type distributions of frozen samples.

## Materials And Methods

### snRNAseq data generation

For nuclei isolation from frozen tissue, all following steps were performed on ice with precooled buffers and centrifugation steps were performed at 4 °C. Tissue was disrupted and nuclei liberated through dounce homogenization in ice-cold homogenization buffer (0.32 M sucrose, 3 mM CaCl<sub>2</sub>, 2 mM magnesium acetate, 0.1 mM EDTA, 1 mM DTT, 10 mg/ml BSA, 10 mM Tris-HCL, protease inhibitors (Sigma-Aldrich)) in the presence of 0.1% NP40. The nuclei suspensions were sequentially passed through 40, 30 and 20-µm cell filters (Miltenyi Biotec) and centrifuged at 1000g for 10 min. The pellet was resuspended in 1% RNase-free BSA in PBS, and stained using DAPI (1:1,000, BD Pharmingen). Intact single nuclei were sorted in bulk using the DAPI-positive event population, at single cell sort precision and using a 100 µm nozzle (BD FACS AriaIII) into PBS / 1% BSA. Nuclei were counted and loaded on a 10x Chromium microfluidic chip, aiming for the maximum possible number of nuclei to be targeted obtained from the sorting. Single-nucleus experiments were performed using the 10x Genomics Single Cell 3' kit v.2 to encapsulate nuclei along with barcode tagged beads, generate and amplify cDNA and to generate sequencing libraries. Each pooled library was barcoded using i7 barcodes provided by 10x Genomics.

cDNA and sequencing library quality and quantity were determined using Agilent's High Sensitivity DNA Assay. Libraries were pooled and sequenced in 150-bp paired-end mode on Illumina's HiSeq platform.

### snRNAseq data processing and analysis

Post-processing pipeline Cell Ranger (<https://support.10xgenomics.com/single-cell-gene-expression/software/>) provided by 10X Genomics was used for demultiplexing, alignment, filtering, barcode and Unique Molecular Identifier (UMI) counting. The pipeline produced files in FASTQ and BAM formats, as well as the matrix of UMI counts. We used Seurat workflow for further quality control and downstream analysis of the snRNAseq gene expression data. The initial data set contained 2935 nuclei and 18104 expressed genes. Nuclei with a high fraction of their counts coming from mitochondrial and ribosomal genes were removed. Next, genes with at least one UMI count present in at least one nucleus were selected. After these steps, 2402 nuclei and 15416 genes were included in the downstream analysis. We normalized the gene expression data with the LogNormalize method of Seurat and standardized the count values prior to performing the Principal Component Analysis (Supplementary figure 1). JackStraw procedure<sup>38</sup> was applied as a denoising step in order to select an optimal number of Principal Components (PCs), indicating that 3 PCs to keep for further downstream analysis, which can be viewed as identifying the true intrinsic dimensionality of the snRNAseq data. Further, the number of cells predicted to be proliferative was investigated using a list of the genes annotated as functioning in the cell cycle according to<sup>39</sup>. The vast majority of cells were not detected to be proliferating and the cycling cells did not form any separate cluster (Supplementary figure 2a-b). Graph-based clustering based on Louvain modularity optimization<sup>23</sup> with resolution parameter equal to 0.1 was used for detecting boundaries between different populations of nuclei, and tSNE non-linear dimension reduction with perplexity 50, that was chosen as a square root of the number of nuclei according to the k-nearest neighbors rule of thumb, was applied for visualization of the nuclei populations (Fig 1a). The snRNAseq data was uploaded to Gene Expression Omnibus (GEO) and is available under accession number GSE190489.

### Deconvolution analysis

Deconvolution analysis<sup>24</sup> was performed using the snRNAseq data as a reference. For this purpose, the snRNAseq data was used to produce cluster signatures, which are marker gene expressions averaged across cells from each cluster. After that, the gene expression of each type of skeletal muscle fibers in a bulk RNAseq sample was inferred via linear matrix decomposition<sup>40</sup>. The deconvolution analysis was performed using DeconRNASeq R / Bioconductor package, and the results are presented in figure 2b-c.

### Study subjects

The MSAT cohort: 39 Swedish male subjects (Table 1) were enrolled in the study by advertising on social media and through local cycling clubs. Inclusion criteria were as follows: 1) Male, 2) Healthy, not on any medications, 3) Age range between 20 and 55. Subjects were given both oral and written information about the experimental procedures before giving their written informed consent. Each participant went

through three visits at different time points. All subjects completed all three visits. The first visit involved a regular doctor's examination with blood samples and measuring anthropometric characteristics. The second visit consisted, after an overnight fast, of a Wingate test followed by muscle biopsy and  $VO_{2max}$  was measured during the third and last visit. The study was approved by the local Ethics committee, Lund University (Dnr 2015/593). For determination of peak anaerobic power (Wingate) and  $VO_{2max}$ , subjects were instructed to perform only easy training 48 hours prior to each test. To determine peak anaerobic power, a 30-s all-out Wingate test<sup>33</sup> was conducted on a cycle ergometer (Monark Peak power). Before the test, a 5-minute low intensity  $\sim 150$ W warm-up, with instructions to perform a 5s high cadence drill each minute was performed. The test started with the subject pedaling as fast as possible. When a cadence of 120 rpm was reached, a braking resistance equivalent to  $0.7 N \times m \times kg^{-1}$  was applied to the freewheel and remained constant during the 30 seconds. Subjects were instructed to sit down throughout the test. Strong verbal encouragement was given throughout to ensure a maximal effort was provided. An incremental test to exhaustion was performed to determine  $VO_{2max}$ . The test started with 3 min of cycling at  $3 W \times kg^{-1}$  (rounded down to nearest 10 W) and then increased by 35 W every 2 minutes until voluntary exhaustion or failure to maintain  $\geq 60$  rpm. Strong verbal encouragement was given throughout.  $VO_2$  was measured using an Oxycon Pro (Jaeger GmbH, Germany) with a mixing chamber and a 30s sampling time. Gas sensors were calibrated according to instructions by the vendor before every test. Maximal oxygen uptake was determined as the mean of the two highest values attained during exercise from any 30-s period.

The Juntendo cohort: 23 Japanese subjects (Table 2) were recruited to examine the associations between RNA expression profiles and muscle fiber composition in the Japanese population. All subjects gave their signed informed consent before inclusion in the study. The study protocols were approved by the Ethics Committees of Juntendo University and were performed in accordance with the Declaration of Helsinki. No additional tests were performed in this cohort.

### Muscle biopsies and histology

In the MSAT cohort (Table 1), muscle biopsies were taken from the *vastus lateralis* muscle under sterile conditions and local anesthesia (1% lidocaine) by using a 5 mm Bergström needle and frozen in liquid nitrogen. The biopsies was taken within 5 minutes after the Wingate test. Serial sections (10  $\mu$ m) were cut using a cryostat at  $-20$  °C. Myofibrillar ATPase histochemistry was performed by preincubation at pH 4.4, 4.6, and 10.3 to identify fiber types<sup>41</sup>; the proportion of fiber types (*i.e.*, Type I, IIa or IIx) were calculated as the number of each fiber type, divided by the total number of fibers in the section. Computer image analysis was performed using image analysis equipment (BioPix IQ 2.0.16 software, BioPix AB, Sweden).

In the Juntendo cohort (Table 2), muscle biopsies were taken from the *vastus lateralis* muscle under sterile conditions and local anesthesia (1% lidocaine) by using a disposal needle biopsy instrument (Max Core; C. R. Bard, Covington, GA). The biopsies were collected from approximately 15 cm above the patella in both legs of each subject under ultrasound imaging (Noblus; Aloka, Tokyo, Japan) and avoided the inclusion of subcutaneous fat and the subfascial and myotendinous parts as far as possible. In addition,

any visible non-muscle tissues (*e.g.*, adipose tissue) were removed from the biopsy samples. Samples were frozen immediately in liquid nitrogen and stored at -80°C until further analysis. Myosin heavy chain (MyHC) protein isoforms were assessed as markers of muscle fiber composition. Frozen muscle samples were homogenized in ice-cold lysis buffer [50 mM HEPES (pH 7.4), 10 mM EDTA, 4 mM EGTA, 50 mM  $\beta$ -glycerophosphate, 25 mM NaF, 5 mM  $\text{Na}_3\text{VO}_4$ ], containing a phosphatase inhibitor (PhosSTOP tablet; Roche Diagnostics, Indianapolis, IN) and a protease inhibitor (Complete tablet; Roche Diagnostics). The lysates obtained were centrifuged at 10,000 g for 10 min at 4°C. An insoluble pellet, obtained after centrifugation, was suspended in a sufficient volume of SDS sample buffer [30% glycerol, 5%  $\beta$ -mercaptoethanol, 2.3% SDS, 0.05% bromophenol blue, and 62.5 mM Tris-HCl (pH 6.8)] and boiled at 95°C for 5 min. MyHC composition was determined by glycerol SDS-PAGE, according to Kakigi *et al*<sup>42</sup>. Briefly, protein samples were resolved by performing glycerol SDS-PAGE [stacking gel: 4% acrylamide, 34.7% glycerol, and 125 mM Tris-HCl (pH 6.8); separating gel: 8% acrylamide, 33.3% glycerol, and 375 mM Tris-HCl (pH 8.3)]. Electrophoresis was started at 60 V with stacking gel at 8°C. The voltage was set to 150 V and run for 18 h at 8°C when the tracking dye had entered the separating gel completely. After separation, the gels were stained with Coomassie brilliant blue (Biosafe G250; Bio-Rad Laboratories, Hercules, CA) and rinsed repeatedly with water. Each gel was scanned using a calibrated densitometer (ChemiDoc Touch Imaging System; Bio-Rad Laboratories), and the relative proportion of MyHC-I, MyHC-IIa, and MyHC-IIx were determined using the calibrated densitometer (ChemiDoc Touch Imaging System) and analytical software (Image Laboratory software version 5.2.1; Bio-Rad Laboratories).

### RNA extraction and totRNAseq analysis

In the MSAT cohort RNA was extracted from 25-30 mg of muscle biopsies using a TissueLyser II (Qiagen) and the miRneasy Mini Kit (Qiagen). RNA concentration was determined using a NanoDrop ND-1000 spectrophotometer ( $A_{260} / A_{280} > 1.8$  and  $A_{260} / A_{230} > 1.0$ ) (NanoDrop Technologies, Wilmington, DE, USA). RNA integrity was verified using the 2200 TapeStation instrument (Agilent Technologies, CA, US), where all samples had an average RNA integrity number (RIN) above 8. In the Juntendo cohort, frozen muscle samples were crushed with 5.0 mm zirconia beads using a Micro Smash MS-100R (Tomy Seiko, Japan) at 3000 rpm twice for 15 s at 2 °C. Total RNA was extracted from muscle samples using TRIzol® Reagent (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. RNA concentration and purity were checked using a NanoDrop 8000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). RNA integrity was verified using the 2200 TapeStation instrument (Agilent Technologies, CA, US), where all samples had an average RNA integrity number (RIN) above 8.

All samples from both the MSAT and Juntendo cohorts were sequenced at Lund University using 800 ng input RNA. Library preparation was made using the TruSeq Stranded Total RNA Library Prep Kit with RiboZero Human/Mouse/Rat Set A (Illumina) and the 75bp paired-end sequencing was performed on a NextSeq instrument using the NextSeq® 500/550 High Output Kit v2 (150 cycles) (Illumina). The sequencing quality was checked with fastQC v0.11.9

(<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and multiQC<sup>43</sup> v1.9. Gene expression was assessed using Salmon<sup>44</sup> v1.2.1. Exon expression was obtained by mapping reads with STAR<sup>45</sup> v2.7.6

and counting with featureCounts<sup>46</sup> v2.0.1 (options: -p -f -C -O). We used the GRCh38 Ensembl v77<sup>47</sup> as a reference genome. We used the Seqtk tool (<https://github.com/lh3/seqtk>) for gradual random downsampling of the MSAT data, and applied Salmon for quantifying gene expression of the downsampled RNAseq data. The quantified gene expression for each downsampling iteration was normalized with TMM<sup>48</sup> and deconvolution analysis was performed using the gene markers identified for slow- and fast-twitch human clusters in the snRNAseq data as described above. We computed the Spearman correlation coefficient and mean square deviation between predicted and true fiber type composition for each downsampling iteration.

## Abbreviations

muscle tissue RNA sequencing data (totRNAseq)

single-cell RNA sequencing (scRNAseq)

single-nuclei RNA sequencing (snRNAseq)

slow-twitch fibers (Type I)

fast-twitch fibers (Type II)

myosin heavy chain (MyHC)

the Muscle SATellite cell study (MSAT)

Juntendo Muscle Study (JMS)

t-distributed stochastic neighbor embedding (tSNE)

ATPase sarcoplasmic/endoplasmic reticulum Ca<sup>2+</sup> transporting 2 (*ATP2A2*)

Tropomyosin 3 (*TPM3*)

Myosin heavy chain 7B (*MYH7B*)

ATPase sarcoplasmic/endoplasmic reticulum Ca<sup>2+</sup> transporting 1 (*ATP2A1*)

Troponin T3, fast skeletal type (*TNNT3*)

Myosin heavy chain 2 (*MYH2*)

Titin (*TTN*)

Maternally expressed 3 (*MEG3*)

paired-end (PE)

Unique Molecular Identifier (UMI)

Principal Components (PCs)

Principal Components Analysis (PCA)

Gene Expression Omnibus (GEO)

RNA integrity number (RIN)

European Genome-phenome Archive (EGA)

## Declarations

Ethics approval and consent to participate: The MSAT study was approved by the local Ethics committee at Lund University (Dnr 2015/593) and the JMS study was approved by the local Ethics committee at Juntendo University. All participants gave informed consent prior to participating in the studies.

Consent for publication: Not applicable

Availability of data and material: All scripts needed to utilize this new methodology is openly available at <https://github.com/OlaHanssonLab/PredictFiberType>. The datasets supporting the conclusions of this article are available in the the European Genome-phenome Archive and the Gene Expression Omnibus repositories, GSE190489 at <https://www.ncbi.nlm.nih.gov/geo>. The totRNAseq data has been submitted to the European Genome-phenome Archive (EGA). The reference numbers will be provided after the submission to EGA is approved by the EGA.

Funding: This work was financially supported by the following: the Knut and Alice Wallenberg Foundation for equipment, Swedish Research Council project grant 2018-02635, Crafoord Foundation, Novo Nordisk Foundation, Pålsson Foundation, Diabetes Wellness, the Swedish Diabetes foundation, the Hjelt Foundation, JSPS KAKENHI, Dnr 16KK0188 and by Institute of Health and Sports Science & Medicine, Juntendo University. LUDC-IRC: Swedish Foundation for Strategic Research, Dnr IRC15-0067, EXODIAB: Swedish Research Council, Strategic Research Area, Dnr 2009-1039. NO and BN are financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab. The funding bodies had no influence or were involved in any other way in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Competing interests: The authors declare that they have no competing interests.

Authors' contributions: NO, SP, BN and OH wrote the main manuscript text and NO, GC, BT, DK and OH prepared the figures. OE, EMM, NF and OH made the tables. MS, KS, BM performed experiments. NO, GC, HP, DK, BT and OH made the bioinformatic analyses. OE, EMM, ML, HK, RK, HN, KFE, NF, SP and OH performed the human studies and/or collected samples. All authors reviewed and approved the manuscript.

Acknowledgements: The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, partially funded by the Swedish Research Council through grant agreement no. 2018-05973 for providing assistance in massive parallel sequencing and computational infrastructure.

## References

- 1 Schiaffino, S. & Reggiani, C. Myosin isoforms in mammalian skeletal muscle. *J Appl Physiol* (1985) 77, 493-501 (1994).
- 2 Schiaffino, S. & Reggiani, C. Fiber types in mammalian skeletal muscles. *Physiol Rev* 91, 1447-1531, doi:10.1152/physrev.00031.2010 (2011).
- 3 Saltin, B., Henriksson, J., Nygaard, E., Andersen, P. & Jansson, E. Fiber types and metabolic potentials of skeletal muscles in sedentary man and endurance runners. *Ann N Y Acad Sci* 301, 3-29 (1977).
- 4 Simoneau, J. A. & Bouchard, C. Human variation in skeletal muscle fiber-type proportion and enzyme activities. *The American journal of physiology* 257, E567-572 (1989).
- 5 Snijders, T., Verdijk, L. B. & van Loon, L. J. The impact of sarcopenia and exercise training on skeletal muscle satellite cells. *Ageing Res Rev* 8, 328-338, doi:10.1016/j.arr.2009.05.003 (2009).
- 6 Wilson, J. M. et al. The effects of endurance, strength, and power training on muscle fiber type shifting. *J Strength Cond Res* 26, 1724-1729, doi:10.1519/JSC.0b013e318234eb6f (2012).
- 7 Harridge, S. D. et al. Whole-muscle and single-fibre contractile properties and myosin heavy chain isoforms in humans. *Pflugers Arch* 432, 913-920 (1996).
- 8 Tanner, C. J. et al. Muscle fiber type is associated with obesity and weight loss. *Am J Physiol Endocrinol Metab* 282, E1191-1196, doi:10.1152/ajpendo.00416.2001 (2002).
- 9 Mogensen, M. et al. Mitochondrial respiration is decreased in skeletal muscle of patients with type 2 diabetes. *Diabetes* 56, 1592-1599, doi:10.2337/db06-0981 (2007).
- 10 Simoneau, J. A., Colberg, S. R., Thaete, F. L. & Kelley, D. E. Skeletal muscle glycolytic and oxidative enzyme capacities are determinants of insulin sensitivity and muscle composition in obese women. *FASEB J* 9, 273-278 (1995).
- 11 Komi, P. V. et al. Skeletal muscle fibres and muscle enzyme activities in monozygous and dizygous twins of both sexes. *Acta Physiol Scand* 100, 385-392 (1977).
- 12 Simoneau, J. A. & Bouchard, C. Genetic determinism of fiber type proportion in human skeletal muscle. *FASEB J* 9, 1091-1095 (1995).

- 13 Schiaffino, S., Reggiani, C. & Murgia, M. Fiber type diversity in skeletal muscle explored by mass spectrometry-based single fiber proteomics. *Histol Histopathol* 35, 239-246, doi:10.14670/HH-18-170 (2020).
- 14 Hedlund, E. & Deng, Q. Single-cell RNA sequencing: Technical advancements and biological applications. *Mol Aspects Med* 59, 36-46, doi:10.1016/j.mam.2017.07.003 (2018).
- 15 De Micheli, A. J., Spector, J. A., Elemento, O. & Cosgrove, B. D. A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations. *Skelet Muscle* 10, 19, doi:10.1186/s13395-020-00236-3 (2020).
- 16 Jensen, J. B. et al. Isolation and characterization of muscle stem cells, fibro-adipogenic progenitors and macrophages from human skeletal muscle biopsies. *Am J Physiol Cell Physiol*, doi:10.1152/ajpcell.00127.2021 (2021).
- 17 van den Heuvel, A. et al. Single-cell RNA sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. *Hum Mol Genet* 28, 1064-1075, doi:10.1093/hmg/ddy400 (2019).
- 18 Barriet, E. et al. Functionally heterogeneous human satellite cells identified by single cell RNA sequencing. *Elife* 9, doi:10.7554/eLife.51576 (2020).
- 19 Xi, H. et al. A Human Skeletal Muscle Atlas Identifies the Trajectories of Stem and Progenitor Cells across Development and from Human Pluripotent Stem Cells. *Cell Stem Cell* 27, 158-176.e110, doi:10.1016/j.stem.2020.04.017 (2020).
- 20 Camps, J. et al. Interstitial Cell Remodeling Promotes Aberrant Adipogenesis in Dystrophic Muscles. *Cell Rep* 31, 107597, doi:10.1016/j.celrep.2020.107597 (2020).
- 21 Wu, H., Kirita, Y., Donnelly, E. L. & Humphreys, B. D. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J Am Soc Nephrol* 30, 23-32, doi:10.1681/ASN.2018090912 (2019).
- 22 Dos Santos, M. et al. Single-nucleus RNA-seq and FISH identify coordinated transcriptional activity in mammalian myofibers. *Nat Commun* 11, 5102, doi:10.1038/s41467-020-18789-8 (2020).
- 23 Blondel, V. D., Guillaume, J. L., Hendrickx, J. M., de Kerchove, C. & Lambiotte, R. Local leaders in random networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 77, 036114, doi:10.1103/PhysRevE.77.036114 (2008).
- 24 Gong, T. & Szustakowski, J. D. DeconRNASeq: a statistical framework for deconvolution of heterogeneous tissue samples based on mRNA-Seq data. *Bioinformatics* 29, 1083-1085, doi:10.1093/bioinformatics/btt090 (2013).

- 25 Costill, D. L., Fink, W. J. & Pollock, M. L. Muscle fiber composition and enzyme activities of elite distance runners. *Med Sci Sports* 8, 96-100 (1976).
- 26 Harber, M. & Trappe, S. Single muscle fiber contractile properties of young competitive distance runners. *J Appl Physiol* (1985) 105, 629-636, doi:10.1152/jappphysiol.00995.2007 (2008).
- 27 Widrick, J. J., Trappe, S. W., Costill, D. L. & Fitts, R. H. Force-velocity and force-power properties of single muscle fibers from elite master runners and sedentary men. *Am J Physiol* 271, C676-683, doi:10.1152/ajpcell.1996.271.2.C676 (1996).
- 28 Bellinger, P. et al. Determinants of last lap speed in paced and maximal 1500-m time trials. *Eur J Appl Physiol*, doi:10.1007/s00421-020-04543-x (2020).
- 29 Oberbach, A. et al. Altered fiber distribution and fiber-specific glycolytic and oxidative enzyme activity in skeletal muscle of patients with type 2 diabetes. *Diabetes Care* 29, 895-900 (2006).
- 30 Lillioja, S. et al. Skeletal muscle capillary density and fiber type are possible determinants of in vivo insulin resistance in man. *J Clin Invest* 80, 415-424, doi:10.1172/JCI113088 (1987).
- 31 Henriksen, E. J. et al. Glucose transporter protein content and glucose transport capacity in rat skeletal muscles. *The American journal of physiology* 259, E593-598, doi:10.1152/ajpendo.1990.259.4.E593 (1990).
- 32 Daugaard, J. R. et al. Fiber type-specific expression of GLUT4 in human skeletal muscle: influence of exercise training. *Diabetes* 49, 1092-1095, doi:10.2337/diabetes.49.7.1092 (2000).
- 33 Ciciliot, S., Rossi, A. C., Dyar, K. A., Blaauw, B. & Schiaffino, S. Muscle type and fiber type specificity in muscle wasting. *Int J Biochem Cell Biol* 45, 2191-2199, doi:10.1016/j.biocel.2013.05.016 (2013).
- 34 Deshmukh, A. S. et al. Deep muscle-proteomic analysis of freeze-dried human muscle biopsies reveals fiber type-specific adaptations to exercise training. *Nat Commun* 12, 304, doi:10.1038/s41467-020-20556-8 (2021).
- 35 Bellinger, P. et al. Muscle fiber typology is associated with the incidence of overreaching in response to overload training. *J Appl Physiol* (1985) 129, 823-836, doi:10.1152/jappphysiol.00314.2020 (2020).
- 36 Lievens, E., Klass, M., Bex, T. & Derave, W. Muscle fiber typology substantially influences time to recover from high-intensity exercise. *J Appl Physiol* (1985) 128, 648-659, doi:10.1152/jappphysiol.00636.2019 (2020).
- 37 Hayot, M. et al. Skeletal muscle microbiopsy: a validation study of a minimally invasive technique. *Eur Respir J* 25, 431-440, doi:10.1183/09031936.05.00053404 (2005).

- 38 Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* (Oxford, England) 31, 545-554, doi:10.1093/bioinformatics/btu674 (2015).
- 39 Kowalczyk, M. S. et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res* 25, 1860-1872, doi:10.1101/gr.192237.115 (2015).
- 40 Sofeu Feugaing, D. D., Götte, M. & Viola, M. More than matrix: the multifaceted role of decorin in cancer. *Eur J Cell Biol* 92, 1-11, doi:10.1016/j.ejcb.2012.08.004 (2013).
- 41 Brooke, M. H. & Kaiser, K. K. Three "myosin adenosine triphosphatase" systems: the nature of their pH lability and sulfhydryl dependence. *J Histochem Cytochem* 18, 670-672 (1970).
- 42 Kakigi, R. et al. Heat stress enhances mTOR signaling after resistance exercise in human skeletal muscle. *J Physiol Sci* 61, 131-140, doi:10.1007/s12576-010-0130-y (2011).
- 43 Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047-3048, doi:10.1093/bioinformatics/btw354 (2016).
- 44 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14, 417-419, doi:10.1038/nmeth.4197 (2017).
- 45 Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 46 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
- 47 Yates, A. D. et al. Ensembl 2020. *Nucleic Acids Res* 48, D682-D688, doi:10.1093/nar/gkz966 (2020).
- 48 Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11, R25, doi:10.1186/gb-2010-11-3-r25 (2010).

## Tables

**Table 1.** Description of the Muscle SATellite cell study (MSAT)

	N	Minimum	Maximum	Mean	SD
Age (yrs)	39	21	54	36	8
Weight (kg <sup>BW</sup> )	39	70	97	79	6
BMI (kg <sup>BW</sup> /m <sup>2</sup> )	39	19	31	24	3
VO <sub>2peak</sub> (ml / kg <sup>BW</sup> per min)	39	31	70	52	8
Peak power (W/ kg <sup>BW</sup> )	39	11	16	13	1
Average power (W/ kg <sup>BW</sup> )	39	8	10	9	1
Type I (%)	35	40	81	62	10
Type II (%)	35	19	60	38	10

Note: Type II fiber types are given as Type II = Type IIa + Type IIx, all male subjects

**Table 2.** Description of the Juntendo Muscle Study (JMS)

	N	Minimum	Maximum	Mean	SD
Age (yrs)	23	20	32	23	3
Weight (kg <sup>BW</sup> )	23	50	72	61	6
BMI (kg <sup>BW</sup> /m <sup>2</sup> )	23	17	26	22	2
Type I (%)	23	27	60	40	9
Type II (%)	23	40	73	60	9

Note: Type II fiber types are given as Type II = Type IIa + Type IIx, 10 female and 13 male subjects

## Figures

**Figure 1**

single-nuclei RNAseq of human skeletal muscle. Slow- (Type I) and fast-twitch (Type II) fibers form species-specific distinct clusters of nuclei. **(a)** Three major clusters of nuclei were identified using graph-based clustering built on Louvain modularity optimization. For visualization of the nuclei populations, t-distributed stochastic neighbor embedding (tSNE) non-linear dimension reduction was applied. **(b)** Examples of marker genes for the three clusters. **(c)** Examples of nuclei expression patterns for genes separating the different clusters, *i.e.*, *EMC10* and *TTN* are muscle enriched (cluster A and B) and *MEG3* is

enriched in cluster C. Type II fiber markers, *i.e.*, *ATP2A1*, *TNNT3* and *MYH2* are enriched in cluster A, whereas Type I fiber markers, *i.e.*, *ATP2A2*, *TPM3* and *MYH7B* are enriched in cluster B.

## Figure 2

**(a)** Correlation between the estimated proportions of muscle fiber nuclei types from totRNAseq data and muscle fiber types from ATPase staining in the MSAT study,  $r = 0.65$  [0.46 – 0.84], [95% CI] at  $p_{\text{spearman}} = 7.96 \times 10^{-6}$ ,  $n = 39$ . **(b)** Correlation between the estimated proportions of muscle fiber nuclei types from totRNAseq data and muscle fiber types from myosin heavy chain distribution in the JMS study,  $r = 0.80$  [0.71 – 0.89], [95% CI] at  $p_{\text{spearman}} = 8.06 \times 10^{-6}$ ,  $n = 22$ . **(c and d)** Correlations between the estimated proportions of muscle fiber nuclei types from totRNAseq data and muscle fiber types from ATPase staining in the MSAT study at different sequencing depths.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfigures.docx](#)