

RESEARCH

Open Access



# A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations

Andrea J. De Micheli<sup>1,2</sup>, Jason A. Spector<sup>1,3</sup>, Olivier Elemento<sup>2</sup> and Benjamin D. Cosgrove<sup>1\*</sup> 

## Abstract

Single-cell RNA-sequencing (scRNA-seq) facilitates the unbiased reconstruction of multicellular tissue systems in health and disease. Here, we present a curated scRNA-seq dataset of human muscle samples from 10 adult donors with diverse anatomical locations. We integrated ~ 22,000 single-cell transcriptomes using Scanorama to account for technical and biological variation and resolved 16 distinct populations of muscle-resident cells using unsupervised clustering of the data compendium. These cell populations included muscle stem/progenitor cells (MuSCs), which bifurcated into discrete “quiescent” and “early-activated” MuSC subpopulations. Differential expression analysis identified transcriptional profiles altered in the activated MuSCs including genes associated with aging, obesity, diabetes, and impaired muscle regeneration, as well as long non-coding RNAs previously undescribed in human myogenic cells. Further, we modeled ligand-receptor cell-communication interactions and observed enrichment of the TWEAK-FN14 pathway in activated MuSCs, a characteristic signature of muscle wasting diseases. In contrast, the quiescent MuSCs have enhanced expression of the *EGFR* receptor, a recognized human MuSC marker. This work provides a new benchmark reference resource to examine human muscle tissue heterogeneity and identify potential targets in MuSC diversity and dysregulation in disease contexts.

## Introduction

Skeletal muscles are essential to daily functions such as locomotion, respiration, and metabolism. Upon damage, resident muscle stem cells (MuSCs) repair the tissue in coordination with supporting non-myogenic cell types such as immune cells, fibroblasts, and endothelial cells [1]. However, with age and disease, the repair capacity of MuSCs declines, leading to complications such as fibrotic scarring, reduced muscle mass and strength [2, 3], fat accumulation, and decreased insulin sensitivity [4], all of which severely affect mobility and quality of life [5].

Human MuSCs are defined by the expression of the paired box family transcription factor PAX7 and can be isolated using various surface marker proteins including

$\beta$ 1-integrin (CD29), NCAM (CD56), EGFR, and CD82 to varying purities [6–10]. With aging, human MuSCs exhibit a heterogeneous expression of the senescence marker p16<sup>Ink4a</sup> and accumulate other cell-intrinsic alterations in myogenic gene expression programs, cell cycle control, and metabolic regulation [2, 11]. However, given their varied molecular and functional states, our understanding of MuSCs in adult human muscle tissue remains incompletely defined. In addition, cellular coordination in the regulation of human muscle homeostasis and regeneration remains poorly understood due to the lack of experimentally tractable models with multiple human muscle cell types. Given these challenges, we posited that an unbiased single-cell reference atlas of skeletal muscle could provide a useful framework to explore MuSC variability and communication in adult humans.

Here, we deeply profiled the transcriptome of thousands of individual MuSCs and muscle-resident cells

\* Correspondence: [bdc68@cornell.edu](mailto:bdc68@cornell.edu)

<sup>1</sup>Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

from diverse adult human muscle samples using single-cell RNA-sequencing (scRNA-seq). After integrating these donor datasets to conserve biological information and overcome technical variation, we resolved two subpopulations of MuSCs with distinct gene expression signatures. Using differential gene expression analysis and ligand-receptor interaction modeling, we extend the known repertoire of human MuSC gene expression programs, suggesting new regulatory programs that may be associated with human MuSC activation, as well as features of human muscle aging and disease.

## Results

### Collection and integration of a diverse human scRNA-seq dataset

We used scRNA-seq to collect and annotate a single-cell transcriptomic dataset of diverse adult human muscle samples under homeostatic conditions. The muscle samples were from surgically discarded tissue from  $n = 10$  donors (range 41 to 81 years old) undergoing reconstructive procedures and originating from a wide variety of anatomical sites in otherwise healthy patients (Fig. 1a). Each sample was  $\sim 50$  mg after removal of extraneous fat and connective tissue. Muscle samples were enzymatically digested into single-cell suspensions and independently loaded into the 10X Chromium system. All together, we collected over 22,000 human muscle single-cell transcriptomes ( $2206 \pm 1961$  cells per dataset) into a single data compendium. Using unsupervised clustering, we resolved 16 types of cells of immune, vascular, and stromal origin, as well as two distinct subpopulations of MuSCs and some myofiber myonuclei (Fig. 1b).

Given important differences in anatomical site, donor health history, age, sex, and surgical procedures, the muscle samples were highly heterogeneous in terms of cell-type diversity and underlying gene expression profiles. Comparing the resulting scRNA-seq datasets is therefore a challenge that we addressed using recently developed bioinformatic integration methods [12–14]. Our goal was to assemble a unified dataset of human muscle tissue that faithfully conserved sources of biological variability such as donor, anatomical location, and cell composition heterogeneity, while accounting for technical biases. We tested four different scRNA-seq data integration methods (Fig. S1 and S3) and found that Scanorama [13] followed by scaling the output by regressing against the library chemistry technical variable (“10X chemistry”) and the number of genes detected per single-cell best satisfied this goal. Detailed information on our methodology is provided in Fig. S1. After integrating the 10 datasets, we noted remarkable consistency amid cell types across donors (Fig. 1c, e), owing to the robustness of scRNA-seq technology, the bioinformatic method chosen, and our sample preparation protocol. Differential gene

expression analysis between the 16 distinct subpopulations identified an extensive set of unique markers that we grouped into 4 categories (Fig. 1d).

### scRNA-seq resolves the cellular diversity of human muscle and novel markers

We annotated and interpreted the consensus cell atlas (Fig. 1b, d) into cell type subpopulations as follows. We identify four types of stromal cells starting with adipocytes found to be expressing apolipoprotein D (*APOD*) [15]), the brown fat tissue adipokine *CXCL14* [16], *GPX3*, and *GLUL*. Among the 3 other subpopulations of fibroblast-like cells, Fibroblast 1 expresses high levels of collagen 1 (*COL1A1*), *SFRP4*, *SERPINE1*, and *CCL2*; Fibroblast 2 expresses fibronectin (*FN1*), the microfibril-associated glycoprotein *MFAP5*, and *CD55* known to be expressed by synovocytes [17]; and Fibroblast 3 is mainly characterized by *SMOC2* identified in tendon fibroblasts [18]. The Fibroblast 3 cluster is similar to the adipocytes cluster though exhibits lower expression levels and frequencies of the marker genes *APOD*, *CXCL12*, and *GLUL*, and contain pre-adipocytes.

We also identify 5 types of vascular cells, including 3 endothelial subpopulations, and a subpopulation of pericytes and smooth muscle cells (SMCs). Pericytes and SMCs express the canonical markers *RGS5* and *MYH11*. Endothelial 1 express E-selectin (*SELE*), *IL6*, *ICAM1*, and *VCAM1*. These genes are upregulated at sites of inflammation to facilitate immune cell recruitment, suggesting this Endothelial 1 cell population may be involved in homeostatic muscle tissue remodeling [19, 20]. Endothelial 2 cells are distinguished by expressing high levels of claudin-5 (*CLDN5*), *ICAM2*, and the chemokine *CXCL2*. Endothelial 3 expresses high levels of the platelet-recruiting Von Willibrand Factor (*VWF*) and caveolin-1 (*CAV1*), a protein known to regulate cholesterol metabolism, atherosclerosis progression, and MuSC activation [21, 22]. Endothelial 3 cells are enriched for expression of *BTLN9*, suggesting they might represent a lymphatic endothelial phenotype [23].

We also noted two types of myeloid immune cells: first, tissue-resident and anti-inflammatory macrophages which express *CD74* and histocompatibility complex *HLA* proteins; second, activated macrophages and monocytes that express inflammatory markers such as *S100A9* (calgranulin) and *LYZ* (lysozyme). Moreover, *S100A9* transcript abundance levels have been shown to be a feature in aging and chronic inflammation [24]. We also identified a pool of T/B lymphocytes and NK cells characterized by *IL7R* and *NKG7*, respectively, as well as a small subset of *HBA1*<sup>+</sup> erythroblasts.

Finally, we identified two subpopulations of MuSCs (henceforth called “MuSC1” and “MuSC2”). MuSC1 highly expressed the canonical myogenic transcription factor *PAX7* [25], as well as chordin-like protein 2 (*CHRD2*) and Delta-like non-canonical Notch ligand 1 (*DLK1*). *CHRD2* has been shown previously to be expressed in freshly isolated

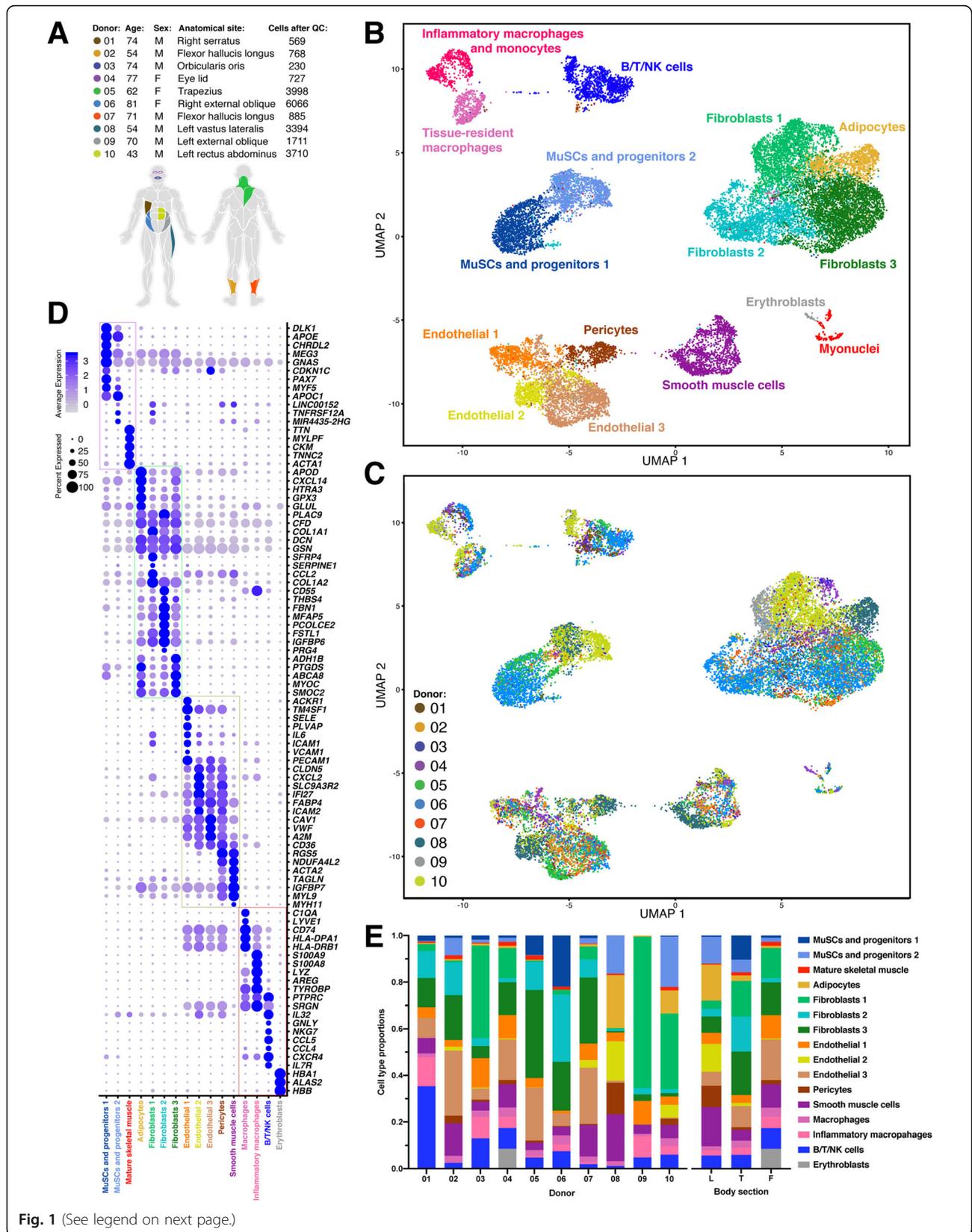


Fig. 1 (See legend on next page.)

(See figure on previous page.)

**Fig. 1** Single-cell transcriptomic map of human muscle tissue biopsies. **a** Metadata (sex, age, anatomical site, and the number of single-cell transcriptomes after quality control (QC) filtering) from  $n = 10$  donors. Colors indicate sample anatomical sites. **b** Scanorama-integrated and batch-effect corrected transcriptomic atlas revealing a consensus description of 16 distinct muscle-tissue cell populations. **c** Transcriptomic atlas colored by donor and anatomical location. **d** Dot-plot showing differentially expressed genes that distinguish the cell populations. Grouped in four compartments: muscle, endothelial/vascular, stromal, and immune. **e** Cell type proportions as annotated in **(b)** across the 10 donors and grouped by body sections. L, leg (donors 02, 07, 08); T, trunk (donors 01, 05, 06, 09, 10); F, face (donors 03, 04)

quiescent human MuSCs [7], though its function is still to be understood. *DLK1* is an inhibitor of adipogenesis whose role in muscle has mainly been recognized in the embryo but remains controversial in adult muscle regeneration [26–28]. In contrast to MuSC1, MuSC2 expressed lower levels of *PAX7* but maintain expression of *MYF5* (a marker of activated MuSCs) and *APOC1* (Fig. 2b). Interestingly, the MuSC2 population also had elevated expression of two long non-coding RNAs (lncRNAs), *LINC00152*, and *MIR4435-2HG*. lncRNAs are involved in regulating myogenesis [29]. Surprisingly, we detected low expression of the myogenic commitment factors *MYOD1* and *MYOG* (Fig. 2b), in contrast to scRNA-seq analyses of adult mouse muscle [30, 31]. These observations suggest that the MuSC1 and MuSC2 populations are both comprised largely of muscle stem cells, not committed myogenic progenitors. In addition, we noted that “Myonuclei” population (Fig. 1b) was enriched for myosin light chain (*MYLFP*), skeletal alpha-actin (*ATCA1*), and troponin C (*TNNC2*), proteins involved in muscle contraction. This multiple-donor scRNA-seq atlas highlights the cellular diversity of human muscle tissue and revealed two distinct MuSC subpopulations along with specific myogenic expression programs.

#### Homeostatic human muscle contains two distinct MuSC subpopulations

We examined genes that were differentially expressed between the MuSC1 and MuSC2 subpopulations and the biological processes that characterize them (Fig. 2a, b). The MuSC1 subpopulation was enriched for *PAX7*, *DLK1*, and *CHRDL2*, as well as for the cyclin-dependent kinase inhibitor *CKDNIC* (encoding P57<sup>KIP2</sup>), suggesting that these cells are quiescent and not cycling. In addition, this subpopulation expresses the transcription factor *BTG2*, which was identified in mouse to be enriched in quiescent MuSCs [30]. We also note that the MuSC1 subpopulation expressed elevated levels of mitochondrial genes as well as *FOS*, *JUN*, and *ERGL1*. Upregulation of these genes has been shown to be potential artefacts of the enzymatic digestion during the sample preparation [32–34].

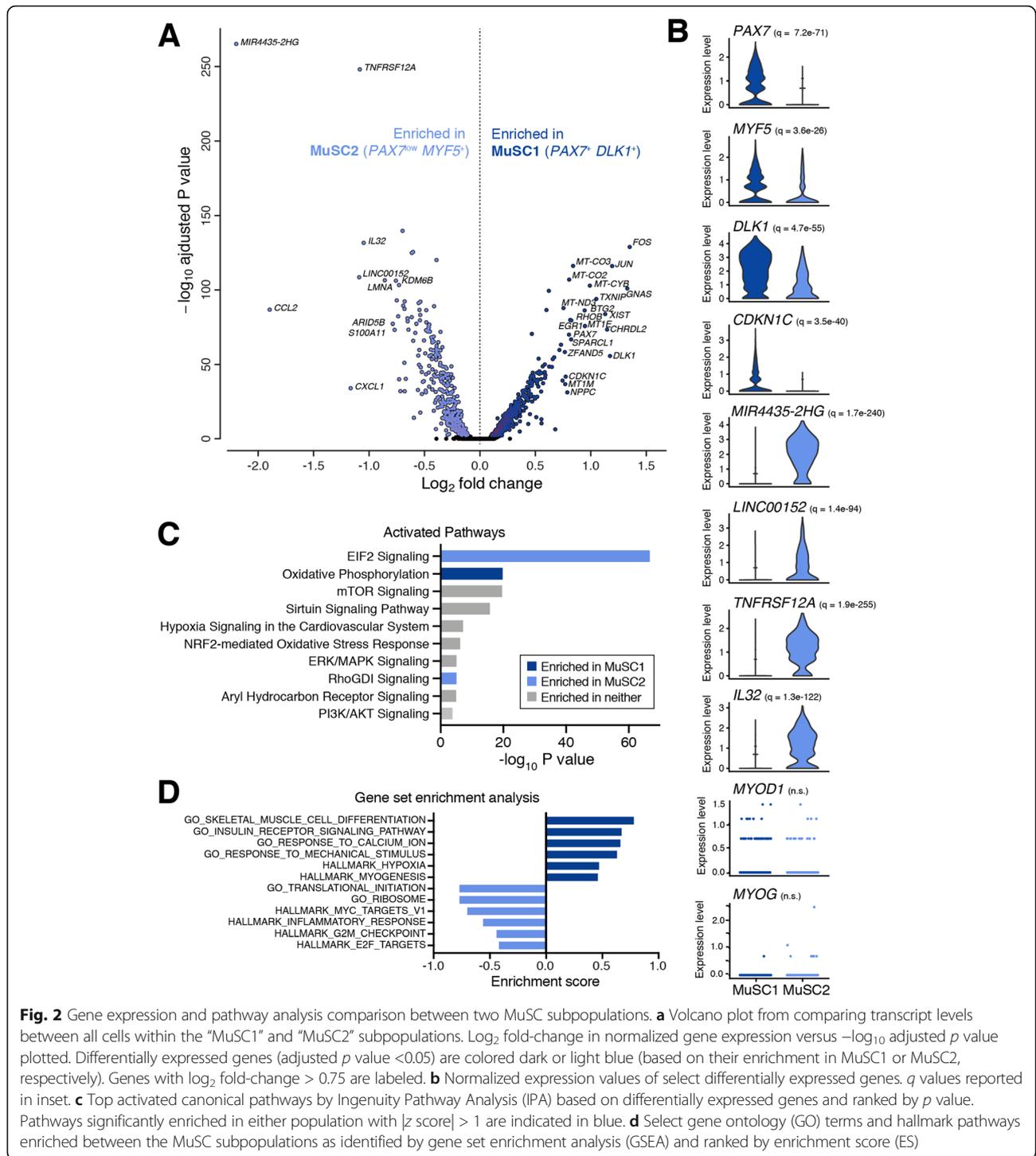
The MuSC2 subpopulation was enriched for multiple markers of inflammation including *CCL2*, *CXCL1*, *IL32*, and surface receptor *TNFRSF12/FN14*. In particular, *CCL2* and *CXCL1* are inflammatory cytokines known to be upregulated in muscle repair, exercise, and fat metabolism [35, 36].

In addition, *IL32* has been shown to have inflammatory properties in human obesity [37] and have a negative impact on insulin sensitivity and myogenesis [38], while *TNFRSF12/FN14* has been implicated in various muscle wasting diseases [39, 40] and metabolic dysfunction [41]. Furthermore, the MuSC2 population is enriched for ribosomal gene expression (e.g. *RPLP1* and *RPS6*; data not shown), indicating that these cells may have elevated translational mechanisms. Lastly, the MuSC1 population has enriched expression of the myogenic gene *PAX7* and, to a lesser extent, *MYF5*, compared the MuSC2 population. These observations suggest that MuSC1 is comprised of quiescent MuSCs, and MuSC2 is comprised of an early-activated MuSCs.

We performed Ingenuity Pathway Analysis (IPA) to compare biological processes differentially activated between the MuSC1 and MuSC2 populations. The IPA gene group “Oxidative Phosphorylation” is enriched in MuSC1 [42], while “EIF2 Signaling,” associated with protein translation processes, is enriched in MuSC2 (Fig. 2c). Furthermore, Gene Set Enrichment Analysis (GSEA) also found MuSC1 to be enriched for “myogenesis,” “muscle cell differentiation,” “hypoxia,” and “response to mechanical stimulus” gene sets, supporting the observation that these cells are both less differentiated and may exhibit enhanced transcriptional responses to mechanical disruption due to tissue dissociation [32–34] (Fig. 2d). MuSC2 cells are enriched for “ribosome and translational initiation,” “MYC targets,” and “E2F (cell proliferation),” “G2M checkpoint (cell division),” and “inflammation” gene sets, further supporting the interpretation that these cells may be in an early activated or partially differentiated state within an inflammatory environment (Fig. 2d). Taken together, these observations suggest that the MuSC1 population is comprised of quiescent MuSCs, while the MuSC2 population is comprised of active, proliferating, and/or dysregulated MuSCs, with expression alterations associated with inflammation, aging, and muscle wasting. Differentially expressed genes such as *IL32*, *CXCL1*, *CCL2*, and *TNFRSF12/FN14* may constitute a marker set for MuSC variation in chronic muscle inflammation in various pathologies.

#### Ligand-receptor interaction model identifies potential surface markers and cell-communication channels in human skeletal muscle homeostasis

We used a ligand-receptor (LR) interaction model and a database of LR pairs [43] to map cell signaling



communication channels in human muscle and uncover differences between MuSC1 and MuSC2 subpopulations (Fig. 3). The model also identifies interacting ligand(s) and is restricted to receptor genes differentially expressed by a specific cell type within the consensus human muscle cell atlas (Fig. 1b). For each LR pair, the model calculates an interaction score from differentially

expressed receptors on a given cell population (e.g., “MuSC1”) relative to all other population and ligands expressed by other cell types. The MuSC1 and MuSC2 subpopulations are involved in numerous LR interactions, as both ligand- and receptor-expressing cells (Fig. 3a), though a majority of all LR interaction pairs instead involve other cell types. This suggests that only a small

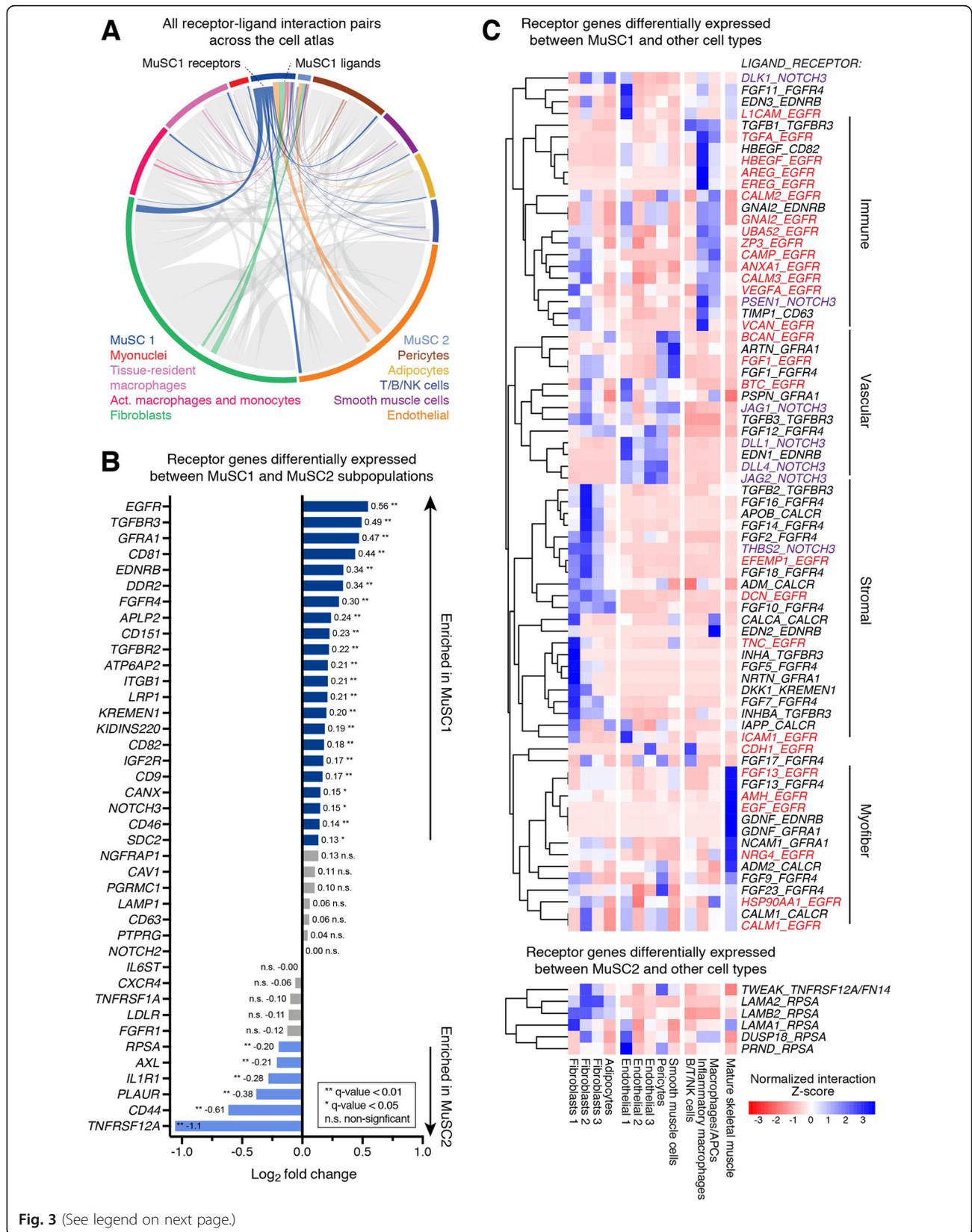


Fig. 3 (See legend on next page.)

(See figure on previous page.)

**Fig. 3** Differentially expressed receptors and ligand-receptor interaction between cell populations. **a** Chord plot of all ligand-receptor (LR) interactions across cell populations/types within the consensus atlas based on co-expression. Each cell type is color-coded with its receptor genes more displaced from the perimeter than its ligand genes. All interactions not involving MuSC1 or MuSC2 are presented in gray. **b** List of differentially expressed genes between the MuSC1 and MuSC2 subpopulation ranked by  $\log_2$  fold-change in expression. Positive average values correspond to genes that are upregulated in MuSC1, whereas negative values are upregulated in MuSC2. Receptors that are statistically significant (FDR-corrected  $q$  value < 0.05) are colored in blue. Receptors that are not statistically significant are in gray. **c** Heatmap representing row-normalized (Z-score) LR interaction scores. Rows represent ligand-receptor interaction pairs in the format *LIGAND\_RECEPTOR*, where the receptor is either differentially expressed in the MuSC1 or MuSC2 populations compared to all the other cell types. Columns identify cell types expressing the ligand. Asterisks after the pair name also indicate that the ligand is differentially expressed by the other cell type and that interaction is likely cell-type specific. Red pairs involve the *EGFR* receptor, purple pairs the *NOTCH3* receptor. A positive value indicates that the interaction has a high score for a particular ligand and cell type compared to other cell types

subset of potential paracrine interactions in human muscle may include MuSCs.

Given the distinct expression profiles between the MuSC1 and MuSC2 populations, we sought to identify genes that could facilitate surface antigen-based separation of these two human MuSC populations for prospective isolation strategies. We identified surface receptor genes that were differentially expressed between the MuSC1 and MuSC2 populations, using a database of 542 human surface “receptor” genes [43] (Fig. 3). MuSC1 exhibit elevated expression of *EGFR*, *ITGB1*, *FGFR4*, *SDC2*, as well as the three tetraspanins *CD81*, *CD82*, and *CD151* (Fig. 3b). *EGFR* is a recently established human MuSC marker and is required for basal-apical asymmetric cell division [7, 10]. The tetraspanin CD82 is also a recently recognized human MuSC maker [6], while CD9 and CD81 have been identified to control muscle myoblast fusion [44]. Furthermore, Syndecans (SDCs) have been identified in mouse to be heterogeneously expressed on MuSCs and myoblasts during muscle repair [30] and have been shown to form co-receptor complexes with integrin  $\beta 1$  (*ITGB1*) and *FGFR4* upstream of signaling pathways regulating myogenesis [45]. Only *SDC4* and *SDC3* have yet been identified to mark adult mouse MuSCs [46]. In comparison, the MuSC2 subpopulation has elevated expression of *CD44* and *TNFRSF12/FN14* as previously noted (Fig. 3b). The *CD44* receptor has been shown to regulate myoblast migration and fusion in mouse, but also mark MuSCs in osteoarthritis patients [47, 48].

Next, we focused the LR analysis on the MuSC1 and MuSC2 populations. We identified 73 and 6 significant LR interactions for the MuSC1 and MuSC2 populations, respectively (Fig. 3c). Over one third of all interactions in the MuSC1 subpopulation involve the *EGFR* receptor, which has recently been shown to play a critical role in directing MuSC asymmetric division in regenerating muscle [10]. A limited number of *EGFR* ligands have been identified in muscle repair, for example, amphiregulin. (*AREG*) secreted by  $T_{reg}$  cells [49]. According to our model findings, *EGFR* may also interact with ligands expressed by immune cells, such as with TGF- $\alpha$  (*TGFA*), heparin-binding EGF (*HBEGF*), amphiregulin (*AREG*), and epiregulin

(*EREG*). Other *EGFR* ligands include brevicin (*BCAN*), and betacellulin (*BTC*) produced by endothelial cells; ECM proteins fibulin 3 (*EFEMP1*), decorin (*DCN*), and tenascin C (*TNC*) expressed by fibroblasts; and FGF13, AHM, NRG4, and EGF, expressed by mature skeletal myofibers. We also detect seven interactions involving *NOTCH3* with a variety of ligands. Notch3 signaling is involved in maintaining MuSC quiescence, in particular through interaction with *DLL4* [50], which we found differently expressed by endothelial cells along with *JAG2*. In addition, *NOTCH3* also interacts with the ECM protein thrombospondin-2 (*THBS2*).

Only two receptors, *TNFRSF12/FN14* and *RPSA*, were found differentially expressed in MuSC2 compared to other cell types. The first, *TNFRSF12/FN14*, interacts with the *TWEAK* cytokine ligand. While typically recognized to be expressed by macrophages and other immune cells [51], our model suggests that *TWEAK* is also expressed by the Fibroblast 2 and pericyte cell populations, though not in a statistically significant manner. The second, *RPSA*, is surface ribosomal protein that interacts with laminins (*LAM*), a dual-specificity phosphatase 18 (*DUSP18*), and prion protein 2 (*PRND*), which taken together may suggest various pathological processes such as prion diseases and cancer [52, 53]. Together, this ligand-receptor analysis identifies a broad set of surface markers that could refine the molecular definition of human MuSCs and their subpopulations, as well as candidate cell-communication channels differentially involved in healthy and diseased muscle tissues.

Lastly, we performed a comparative analysis of receptor gene expression between mouse and human MuSCs. We integrated the human scRNA-seq datasets described in Fig. 1 and an adult mouse muscle injury-response scRNA-seq time-course previously reported [30] by converting mouse genes to their corresponding human ortholog. The multi-species scRNA-seq atlas was integrated with Scanorama and corrected with Harmony (Fig. S2A-B) [54]. From this integrated atlas, we annotated all clusters as in Fig. 1. We identified two MuSC clusters which both contained cells from both mouse and human samples. We then performed differential expression analysis between species comparing aggregated human MuSC1 and MuSC2

cells to mouse MuSCs from the uninjured timepoint (Fig. S2C). We found that *EGFR* and *CD99* were most differentially expressed by human MuSCs and, conversely, *CRLF1* and *SDC4* were most enriched in mouse MuSCs. This findings suggest that mouse and human MuSC exhibit species-specific receptor expression signatures.

## Discussion

Here we present an annotated multi-donor single-cell RNA sequencing dataset consisting of 22,000 single-cell transcriptomes from 10 different donors and unique anatomical sites, some of which difficult to access outside of reconstructive surgeries. Our study complements other recent reports by Rubenstein et al. and Barruet et al., which collected dissociated whole vastus lateralis muscles and FACS-sorted MuSC samples mostly from vastus lateralis muscles, respectively, by providing more diversity in anatomical sites and donor demographics [55, 56]. As such, these scRNA-seq data exhibited notable biological and technical variation, and therefore, we applied the bioinformatic method Scanorama to assemble an integrated cellular atlas with minimal technical biases so that we could examine the cellular heterogeneity across diverse adult human muscle tissue samples. We observed that Scanorama performed more successfully than other data integration approaches, especially when including a scaling regression for sequencing chemistry (Fig. S1 and S3). Notably, even after performing Scanorama with scaling, we still observed that integrated atlas exhibited biological (donor) and technical (sequencing chemistry) biases, but retained some degree of donor-specific cell-type subpopulations.

We describe the muscle tissue cellular heterogeneity and provide a comprehensive analysis of differentially expressed genes for 16 resolved cell subpopulations (Fig. 1), adding to a growing literature documenting human muscle cell transcriptional diversity [55–57]. Compared to other studies, the broader variety of muscle tissue samples combined with the lack of FACS selection allowed us to identify candidate subpopulations of muscle fibroblasts and vascular endothelial cells that may provide unique perspective to human muscle physiology. In particular, we remark that Endothelial 1 expressed *DARC/ACKR1*, a gene identified in mouse and human [56, 58] to mark cells of post-capillary venular origin (Fig. 1d). Rubenstein et al. also found a *DARC/ACKR1*+ post-capillary venular endothelial cluster and a second *VWF*+ *FABP*+ cluster, which overlaps with the Endothelial 2 and 3 clusters reported here. We suggest that the Endothelial 2 cluster may contain both arterial and capillary endothelial cells, but could not further partition and classify this cluster. We suggest that the Endothelial 3 cluster may represent lymphatic endothelium

due to its differential expression of *BTLN9*, a marker of lymphatic endothelial cells [23].

Most notably, this analysis suggests that human muscle may contain two distinct MuSC subpopulations (Fig. 2). This finding contrasts with Rubenstein et al. which observed a single MuSC (“satellite cell”) population from dissociated whole muscle samples and Barruet et al. which observed ~ 12 clusters from human MuSCs prospectively enriched by *CXCR4*+/*CD29*+/*CD56*+ FACS. Since cluster distinction depends on both the cellular diversity and sample complexity, it is expected that variation in study design and methods will yield differing conclusions regarding sub-population resolution. In this work, we found a “MuSC1” subpopulation to be largely comprised of “quiescent” MuSCs, owing to high levels of *PAX7*, the mitotic inhibitor *CDKN1C*, and *DLK1*. Interestingly, *DLK1* may be an important regulator for human MuSC maintenance and a marker of healthy tissue given its role in inhibiting adipogenesis [26]. Conversely, we identified in the “MuSC2” population signatures of inflammation and increased fat metabolism (*CCL2* and *CXCL1*), reduced insulin sensitivity (*IL32*), cell cycle (EIF2 Signaling terms), and muscle wasting (*TNFRSF12/FN14*), thereby suggesting that these cells may constitute an “early-activated” and possibly dysfunctional MuSC pool. These markers are consistent with prior observations that excessive fat accumulation in muscle can be attributed to obesity, diabetes, and aging [4]. In addition, we identify two upregulated lncRNAs that warrant further investigation as candidate non-coding regulators of myogenesis [29]. Moreover, the finding of two human MuSC subpopulations mirrors similar observations made from mouse muscle scRNA-seq analyses [30, 31] and agrees with the general conceptual framework that MuSCs transition between quiescent, activated, and cycling states [1]. Future studies comparative analysis of these MuSC subpopulations across species may reveal human-specific aspects of myogenesis.

Ligand-receptor interaction models from scRNA-seq data can help formulate new hypotheses about cell-communication channels that regulate muscle function [30]. Identifying new MuSCs surface receptors will also help us refine MuSC purification protocols for prospective isolation studies used for in vitro and transplantation models. Our LR model revealed a set of 40 surface receptor genes that are distinctly expressed between MuSC1 and MuSC2, confirming some prior reports but also providing new candidate surface antigens for human MuSC subpopulation fractionation (Fig. 3). For example, we identify that *SDC2* may mark “quiescent” MuSCs while *CD44*, *TNFRSF12*, and *RPSA* “early-activated” MuSCs in aging and disease contexts. In addition, our model proposed 79 cell-communication signals that may act between MuSCs and other cell types, in particular

with fibroblasts, myofibers and immune cells through the EGFR receptor, and with vascular cells through the NOTCH3 receptor. These interactions may be critical regulators of muscle homeostasis and should be further investigated.

This study presents a new set of candidate receptor expression signatures that may define human MuSC subpopulations (Fig. 3b) and provide human-specific receptor patterns (Fig. S2C). This approach is complimentary to receptor screening approaches, which have previously been useful to identify EGFR and CD82 as human MuSC receptor markers for flow cytometry [6, 7, 9]. The subpopulation-specific receptor genes identified here may allow for further comparison of molecular and functional human MuSC diversity across muscle groups [59, 60].

Our study has some limitations. First, the sample size is small, and donors are very diverse, thus limiting our ability to control for age and sex. Therefore, we could not examine cell composition or gene expression trends based on muscle group, donor sex, or donor age. Even for samples from the same muscle (e.g., flexor hallucis longus [donors 2 and 7] or external oblique [donors 6 and 9]), we were unable to perform these comparisons with statistical power. Further, we performed differential expression and gene set enrichment analyses within the MuSC1 and MuSC2 populations between the four middle-age (43–69 years old) and six aged (70–81 years old) donors, but found few age-cohort specific differences (data not shown). Second, future studies should aim at collecting muscle specimens in a more controlled manner, for example using a Bergström needle [61, 62] from a unique anatomical site; though this

would not be possible for some muscles presented in this study. These biopsies would allow for aging and disease comparative analyses. Indeed, a recent report by Rubenstein et al. [56] performed scRNA-seq on four human vastus lateralis muscle biopsies found that myofiber type composition and gene expression alterations based on donor age.

Nevertheless, our dataset offers a new transcriptomic cell reference atlas and computational data integration approaches as a benchmark resource to examine human muscle cell diversity in health, aging, and disease.

## Methods

### Human participation for muscle sample collection

All procedures were approved by the Institutional Review Board at Weill Cornell Medical College (WCMC IRB Protocol # 1510016712) and were performed in accordance with relevant guidelines and regulations. All specimens were obtained at the New York-Presbyterian/Weill Cornell campus. All subjects provided written informed consent prior to participation. Samples were de-identified in accordance to IRB guidelines, and only details concerning age, sex, and anatomic origin were included. Sample anatomic locations and donor details are provided in Fig. 1a.

### Muscle digestion and single-cell sequencing library preparation

After collection from donors during surgery, the muscle samples were cleared from excessive fat and connective tissue and weighted. About 50–65 mg of tissue was then digested into a single-cell suspension following a previously

**Table 1** List of reagents and other resources used in this study

<b>Reagents</b>		
Dispase II (neutral protease, grade II)	Sigma-Aldrich	04942078001
Collagenase D, from <i>Clostridium histolyticum</i>	Sigma-Aldrich	11088866001
<b>Commercial kits</b>		
Chromium Single Cell 3' Library & Gel Bead Kit v2	10X Genomics	CG00052 (protocol)
Chromium Single Cell 3' Library & Gel Bead Kit v3	10X Genomics	CG000183 (protocol)
<b>Deposited data</b>		
Human ligand-receptor database	[43]	<a href="https://www.ncbi.nlm.nih.gov/pubmed/26198319">https://www.ncbi.nlm.nih.gov/pubmed/26198319</a>
Human scRNA-seq dataset	This paper	GSE143704
Mouse scRNA-seq dataset	[29]	GSE143437
<b>Software packages and algorithms</b>		
Cell Ranger 3.1.0 (July 24, 2019)	10X Genomics	<a href="https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest">https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest</a>
Seurat 3.1.0	[14]	<a href="https://github.com/satijalab/seurat">https://github.com/satijalab/seurat</a>
Scanorama (online version as of 2019-11-19)	[13]	<a href="https://github.com/brianhie/scanorama">https://github.com/brianhie/scanorama</a>
Harmony (online version as of 2019-11-19)	[54]	<a href="https://github.com/immunogenomics/harmony">https://github.com/immunogenomics/harmony</a>
biomaRt 2.43.1 (online version as of 2020-01-08)	[64]	<a href="https://bioconductor.org/packages/release/bioc/html/biomaRt.html">https://bioconductor.org/packages/release/bioc/html/biomaRt.html</a>
Gene Set Enrichment Analysis (4.0.3)	[65]	<a href="http://software.broadinstitute.org/gsea/index.jsp">http://software.broadinstitute.org/gsea/index.jsp</a>
Ingenuity Pathway Analysis (IPA, 2019-08-30)	QIAGEN	<a href="https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/">https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/</a>

reported protocol [63]. Briefly, the specimen was digested in 8 mg/mL Collagenase D (Roche) and 4.8 U/mL Dispase II (Roche) for 1 h followed by manual dissociation, filtration, and red blood cell lysis (Table 1). All single-cell suspensions were then frozen at  $-80^{\circ}\text{C}$  in 90% FBS, 10% DMSO and were re-filtered after thawing and prior to generating scRNA-seq libraries. The sequencing libraries were prepared using the Chromium Single Cell 3' reagent V2 or V3 kit (10X Genomics) in accordance with the manufacturer's protocol and diluted as to yield a recovery of  $\sim 6000$  single-cell transcriptomes with  $< 5\%$  doublet rate (Table 1). The libraries were sequenced in multiplex ( $n = 2$  per sequencing run) on the NextSeq 500 sequencer (Illumina) to produce between 200 and 250 million reads per library.

### Single-cell data analysis

Sequencing reads were processed with the Cell Ranger version 3.1 (10X Genomics) using the human reference transcriptome GRCh38. The downstream analysis was carried out with R 3.6.1 (2019-07-05). Quality control filtering, data clustering, visualization, and differential gene expression analysis was carried out using Seurat 3.1.0 R package [14]. Each of the 10 datasets was first analyzed and annotated independently before integration with Scanorama [13] (Table 1). Filtering retained cells with  $> 1000$  unique molecular identifiers (UMIs),  $< 20\%$  UMIs mapped to mitochondrial genes, and genes expressed in at least 3 cells (Fig. S4). Unsupervised shared nearest neighbor (SNN) clustering was performed with a resolution of 0.4 following which clusters were annotated with a common nomenclature of 12 cell type terms (Fig. S1). Differential expression analysis was achieved using either Seurat's "FindAllMarkers" (Fig. 1d) or "FindMarkers" (Fig. 2a) function using a Wilcoxon Rank Sum test and only considering genes with  $> \log_2(0.25)$  fold-change and expressed in at least 25% of cells in the cluster.  $P$  values were corrected for false-discovery (FDR) and then reported as  $q$  values. Integration of raw counts was achieved using the "scanorama.correct" function from Scanorama. The integrated values were finally scaled in Seurat regressing out the 10X chemistry type and the number of genes per cell. Visualization was done using uniform manifold approximation and projection (UMAP) [66]. In Fig. S2, we integrated these human scRNA-seq datasets with a cohort of adult mouse muscle scRNA-seq datasets collected 0–7 days post-notexin injury [30]. For multi-species integration, scRNA-seq datasets were integrated using first Scanorama and then Harmony [54] to align related cell populations across species. Mouse genes were converted to human orthologs using biomaRt Bioconductor R package [64] (Table 1). For differential expression analysis between human and mouse samples, we compared human MuSCs (combining MuSC1 + 2 clusters)

and the uninjured mouse MuSCs to focus on cells from the homeostatic conditions.

### Pathway and gene set enrichment analysis

The list of differentially expressed genes between MuSC1 and MuSC2 (Fig. 2a) was used in Ingenuity Pathway Analysis (IPA) (QIAGEN, 2019-08-30). Activated (canonical) pathways were calculated by "Core Analysis" setting a  $q$  value cutoff of 0.05, which yielded 964 genes (366 down, 598 up). Top canonical pathways were chosen based of  $-\log(p)$  value and  $z$  score values. Gene set enrichment analysis (GSEA, v.4.0.3) [65] was ran on the same gene list as IPA ranked by  $\log_2$  fold-change and with default program settings (Table 1). Gene sets database used the following: h.all.v7.0.symbols.gmt, c2.all.v7.0.symbols.gmt, c5.all.v7.0.symbols.gmt (Broad Institute). Gene sets enriched in phenotype were selected based on  $q$  value and enrichment score (ES).

### Ligand-receptor cell communication model

The model aims at scoring potential ligand-receptor interactions between MuSCs (receptor) and other cell types (ligand). We used the ligand-receptor interaction database from Ramilowski et al. [43] (Table 1). From the database, we considered 1915 ligand-receptor pairs (from 542 receptors and 518 ligands) to test for differential expression in our scRNA-seq dataset. To calculate the score for a given ligand-receptor pair, we multiply the average receptor expression in MuSCs by the average ligand expression per other cell type. We only considered receptors that are differentially expressed in either the MuSC1 or MuSC2 subpopulation when compared individually to all other cell types.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13395-020-00236-3>.

#### **Additional file 1: Figure S1. Comparison of scRNA-seq integration and batch correction methods.** We compared four scRNA-seq data integration methods to evaluate which most faithfully conserves donor, anatomical, and biological information while minimizes technical biases.

**(A)** The  $n = 10$  donor datasets were first annotated independently using a nomenclature of 12 common cell type terms following unsupervised SNN clustering. Then we evaluated the integration method by UMAP and by coloring the data either by cell type, donor ID, or 10X library chemistry used. *First*, we integrated the data by merging the individually normalized gene expression matrices without any further correction. We saw strong technical biases that overwhelmed biological information as the different cell populations segregate by sample/donor and chemistry type. For instance, the two MuSC and progenitor subpopulations are grouped with fibroblasts and endothelial cells. *Second*, we tested the Seurat SCT integration method [14]. This method first calculates a cross-correlation subspace from genes that are shared between datasets. We noticed that this method better "aligns" donor and chemistry type but at the expense of masking biological variability. For instance, we observed that the two MuSC and four stromal subpopulations (Fibroblast 1,2,3 and Adipocytes)

were grouped together, hiding important biological heterogeneity. Although certainly useful to validate reproducibility in scRNA-seq experiments, the Seurat SCT integration approach overcorrected biological heterogeneity for heterogeneous samples. *Third*, we tested the Scanorama method [13], which relies on a computer vision algorithm that “stitches” datasets together even when the cell type composition between dataset is considerably different. We see that this method groups similar cell populations together while acknowledging donor differences. Yet, surprisingly, this method is also very sensitive at picking up differences in chemistry. To correct this chemistry effect, we scaled the Scanorama output by regressing out the chemistry and the number of genes detected per cell (significantly different between chemistry type) (B). Using this integration method, we observed clear separation of the independently annotated cell populations. We present the resulting Scanorama-integrated dataset as a “consensus atlas” (see Fig. 1b-c) of human muscle that describes donor-to-donor differences while grouping cells that are similar together and removing technical biases. **Figure S2. Integration of human and mouse scRNA-seq data sets allows comparison of MuSC receptor gene expression across species.** We generated an integrated scRNA-seq atlas including human sample datasets from Fig. 1 and an adult mouse muscle regeneration time-course from De Micheli et al. [29]. These datasets were integrated using first Scanorama and then Harmony for alignment across species. (A) Multi-species integrated atlas presented by UMAP plot a colored by sample type. (B) Multi-species integrated atlas presented by UMAP plot and annotated by cell-type clusters. (C) The human MuSC1 and MuSC2 clusters were grouped into a cumulative human MuSC cell population, which was compared to mouse MuSCs from the uninjured samples only. Receptor genes were analyzed between the mouse and human MuSC cells for differential expression. Differentially expressed genes with an FDR-corrected  $q$ -value < 0.05 are shown in (C). **Figure S3. Composition of single-cell reference atlas as a whole and in cell-type clusters by donor.** (A) Visualization of donor ( $n = 10$ ) contributions to the whole single-cell reference atlas. In each panel, the full atlas is presented as a UMAP plot, with the cells for an individual donor are colored and overlaid on cells from all other donors (in gray). Note the total number of cells assayed differs for each donor (see Fig. 1a). (B) Bar plot representing the relative contribution of cells with each cell type cluster from each donor. Note that the MuSC1 and MuSC2 clusters are also plotted as a combined cluster on the left side of the bar plot for reference. **Figure S4. Transcriptional detection variation within human muscle reference atlas.** UMAP plots featuring (left) the number of unique molecular identifiers (UMIs) and (right) number of genes detected per single cell. Note that QC filtering removed all cells with less than 1000 UMIs (see Methods).

#### Acknowledgements

The authors acknowledge helpful advice from colleagues in the Cosgrove and Elemento groups, as well as Christopher Mendias at the Hospital for Special Surgery and Peter Schweitzer of Genomics Facility at the Cornell University Biotechnology Resource Center. Lastly, the authors are grateful for the human tissue donors.

#### Authors' contributions

A.J.D. and B.D.C. designed the study and wrote the manuscript. J.A.S. obtained the human tissue samples. A.J.D. performed the tissue dissociations, scRNA-seq, and data analysis, with supervision and assistance from B.D.C. and O.E. All authors reviewed, read, and approved the final manuscript.

#### Funding

This work was financially supported by the National Institutes of Health under award R01AG058630 (to B.D.C.), a Glenn Medical Research Foundation and American Federation for Aging Research Grant for Junior Faculty (to B.D.C.), and a US Department of Education Graduate Assistantship in Areas of National Need under Award P200A150273 (to A.J.D.). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of these funding sources.

#### Availability of data and materials

The human muscle scRNA-seq datasets supporting the conclusions of this article are archived at the NIH GEO repository under accession number GSE143704.

#### Ethics approval and consent to participate

All procedures were approved by the Institutional Review Board at Weill Cornell Medical College (WCMC IRB Protocol # 1510016712) and were performed in accordance with relevant guidelines and regulations. All specimens were obtained at the New York-Presbyterian/Weill Cornell campus. All subjects provided written informed consent prior to participation. Samples were de-identified in accordance to IRB guidelines and only details concerning age, sex, and anatomic origin were included.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no conflicts of interest.

#### Author details

<sup>1</sup>Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA. <sup>2</sup>Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY 10021, USA. <sup>3</sup>Division of Plastic Surgery, Weill Cornell Medical College, New York, NY 10021, USA.

Received: 30 January 2020 Accepted: 10 June 2020

Published online: 06 July 2020

#### References

- Bentzinger CF, Wang YX, Dumont NA, Rudnicki MA. Cellular dynamics in the muscle satellite cell niche. *EMBO Rep.* 2013;14:1062–72.
- Blau HM, Cosgrove BD, Ho ATV. The central role of muscle stem cells in regenerative failure with aging. *Nat Med.* 2015;21:854.
- Järvinen TA, Järvinen M, Kalimo H. Regeneration of injured skeletal muscle after the injury. *Muscles Ligaments Tendons J.* 2014;3:337–45.
- Addison O, Marcus RL, LaStayo PC, Ryan AS. Intermuscular fat: a review of the consequences and causes. *Int J Endocrinol.* 2014;2014:1–11.
- Larsson L, Degens H, Li M, Salvati L, Lee YI, Thompson W, Kirkland JL, Sandri M. Sarcopenia: aging-related loss of muscle mass and function. *Physiol Rev.* 2018;99:427–511.
- Alexander MS, Rozkalne A, Colletta A, Spinazzola JM, Johnson S, Rahimov F, Meng H, Lawlor MW, Estrella E, Kunkel LM, et al. CD82 is a marker for prospective isolation of human muscle satellite cells and is linked to muscular dystrophies. *Cell Stem Cell.* 2016;19:800–7.
- Charville GW, Cheung TH, Yoo B, Santos PJ, Lee GK, Shrager JB, Rando TA. Ex vivo expansion and in vivo self-renewal of human muscle stem cells. *Stem Cell Reports.* 2015;5:621–32.
- Pisani DF, Clement N, Loubat A, Plaisant M, Sacconi S, Kurzenne J-Y, Desnuelle C, Dani C, Dechesne CA. Hierarchization of myogenic and adipogenic progenitors within human skeletal muscle. *Stem Cells.* 2010;28:2182–94.
- Uezumi A, Nakatani M, Ikemoto-Uezumi M, Yamamoto N, Morita M, Yamaguchi A, Yamada H, Kasai T, Masuda S, Narita A, et al. Cell-surface protein profiling identifies distinctive markers of progenitor cells in human skeletal muscle. *Stem Cell Reports.* 2016;7:263–78.
- Wang YX, Feige P, Brun CE, Hekmatnejad B, Dumont NA, Renaud J-M, Faulkes S, Guindon DE, Rudnicki MA. EGFR-Aurka signaling rescues polarity and regeneration defects in dystrophin-deficient muscle stem cells by increasing asymmetric divisions. *Cell Stem Cell.* 2019;24:419–432.e6.
- Sousa-Victor P, Gutarra S, García-Prat L, Rodríguez-Ubrea J, Ortet L, Ruiz-Bonilla V, Jardí M, Ballestar E, González S, Serrano AL, et al. Geriatric muscle stem cells switch reversible quiescence into senescence. *Nature.* 2014;506:316–21.
- Stuart T, Satija R. Integrative single-cell analysis. *Nat Rev Genet.* 2019;20:257–72.
- Hie B, Bryson B, Berger B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat Biotechnol.* 2019;37:685–91.
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive integration of single-cell data. *Cell.* 2019;177:1888–1902.e21.
- Muffat J, Walker DW. Apolipoprotein D: an overview of its role in aging and age-related diseases. *Cell Cycle.* 2010;9:269–73.

16. Cereijo R, Gavalda-Navarro A, Cairó M, Quesada-López T, Villarroya J, Morón-Ros S, Sánchez-Infantes D, Peyrou M, Iglesias R, Mampel T, et al. CXCL14, a brown adipokine that mediates brown-fat-to-macrophage communication in thermogenic adaptation. *Cell Metab.* 2018; 28:750–763.e6.
17. Karpus ON, Kiener HP, Niederreiter B, Yilmaz-Elis AS, van der Kaa J, Ramaglia V, Arens R, Smolen JS, Botto M, Tak PP, et al. CD55 deposited on synovial collagen fibers protects from immune complex-mediated arthritis. *Arthritis Research & Therapy.* 2015;17:6.
18. De Micheli AJ, Swanson JB, Dissler NP, Martinez LM, Walker NR, Oliver DJ, Cosgrove BD, Mendias CL. Single-cell transcriptomics identify extensive heterogeneity in the cellular composition of mouse Achilles tendons. *BioRxiv.* 2020b;801266.
19. Goncharov NV, Nadeev AD, Jenkins RO, Avdonin PV. Markers and biomarkers of endothelium: when something is rotten in the state. *Oxidative Med Cell Longev.* 2017;2017:9759735.
20. Watson C, Whittaker S, Smith N, Vora AJ, Dumonde DC, Brown KA. IL-6 acts on endothelial cells to preferentially increase their adherence for lymphocytes. *Clin Exp Immunol.* 1996;105(1):112–9.
21. Fernández-Hernando C, Yu J, Dávalos A, Prendergast J, Sessa WC. Endothelial-specific overexpression of caveolin-1 accelerates atherosclerosis in apolipoprotein E-deficient mice. *Am J Pathol.* 2010;177:998–1003.
22. Volonte D, Liu Y, Galbiati F. The modulation of caveolin-1 expression controls satellite cell activation during muscle repair. *FASEB J.* 2004;19:237–9.
23. Fujimoto N, He Y, D'Addio M, Tacconi C, Detmar M, Dieterich LC. Single-cell mapping reveals new markers and functions of lymphatic endothelial cells in lymph nodes. *PLoS Biol.* 2020;18:e3000704.
24. Swindell WR, Johnston A, Xing X, Little A, Robichaud P, Voorhees JJ, Fisher G, Gudjonsson JE. Robust shifts in S100a9 expression with aging: a novel mechanism for chronic inflammation. *Sci Rep.* 2013;3:1215.
25. Kuang S, Chargé SB, Seale P, Huh M, Rudnicki MA. Distinct roles for Pax7 and Pax3 in adult regenerative myogenesis. *J Cell Biol.* 2006;172:103.
26. Andersen DC, Laborda J, Baladron V, Kassem M, Sheikh SP, Jensen CH. Dual role of delta-like 1 homolog (DLK1) in skeletal muscle development and adult muscle regeneration. *Development.* 2013;140:3743.
27. Waddell JN, Zhang P, Wen Y, Gupta SK, Yevtodiyenko A, Schmidt JV, Bidwell CA, Kumar A, Kuang S. Dlk1 is necessary for proper skeletal muscle development and regeneration. *PLoS One.* 2010;5:e15055.
28. Zhang L, Uezumi A, Kaji T, Tsujikawa K, Andersen DC, Jensen CH, Fukada S. Expression and Functional Analyses of Dlk1 in Muscle stem cells and mesenchymal progenitors during muscle regeneration. *Int J Mol Sci.* 2019; 20:3269.
29. Hagan M, Zhou M, Ashraf M, Kim I-M, Su H, Weintraub NL, Tang Y. Long noncoding RNAs and their roles in skeletal muscle fate determination. *Noncoding RNA Investig.* 2017;1:24.
30. De Micheli AJ, Laurillard EJ, Heinke CL, Ravichandran H, Fraczek P, Soueid-Baumgarten S, De Vlaminck I, Elemento O, Cosgrove BD. Single-cell analysis of the muscle stem cell hierarchy identifies heterotypic communication signals involved in skeletal muscle regeneration. *Cell Rep.* 2020;30:3583–3595.e5.
31. Dell'Orso S, Juan A-H, Ko K-D, Naz F, Gutierrez-Cruz G, Feng X, and Sartorelli V. (2019). Single-cell analysis of adult skeletal muscle stem cells in homeostatic and regenerative conditions. *Development dev.* 174177.
32. Machado L, Esteves de Lima J, Fabre O, Proux C, Legendre R, Szegedi A, Varet H, Ingerslev LR, Barrès R, Relaix F, et al. In situ fixation redefines quiescence and early activation of skeletal muscle stem cells. *Cell Rep.* 2017; 21:1982–93.
33. van den Brink SC, Sage F, Vértsey Á, Spanjaard B, Peterson-Maduro J, Baron CS, Robin C, van Oudenaarden A. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat Methods.* 2017;14:935–6.
34. van Velthoven CTJ, de Morree A, Egner IM, Brett JO, Rando TA. Transcriptional profiling of quiescent muscle stem cells in vivo. *Cell Rep.* 2017;21:P1994–2004.
35. Harmon B.T., Orkunoglu-Suer, E.F., Adham, K., Larkin, J.S., Gordish-Dressman, H., Clarkson, P.M., Thompson, P.D., Angelopoulos, T.J., Gordon, P.M., Moyna, N.M., et al. (2010). CCL2 and CCR2 variants are associated with skeletal muscle strength and change in strength with resistance training. *J Appl Physiol* (1985) 109, 1779–1785.
36. Pedersen L, Olsen CH, Pedersen BK, Hojman P. Muscle-derived expression of the chemokine CXCL1 attenuates diet-induced obesity and improves fatty acid oxidation in the muscle. *American Journal of Physiology-Endocrinology and Metabolism.* 2012;302:E831–40.
37. Catalán V, Gómez-Ambrosi J, Rodríguez A, Ramírez B, Ortega VA, Hernández-Lizoain JL, Baixauli J, Becerril S, Rotellar F, Valentí V, et al. IL-32 $\alpha$ -induced inflammation constitutes a link between obesity and colon cancer. *Oncoimmunology.* 2017;6:e1328338.
38. Davegårdh C, Broholm C, Perflyev A, Henriksen T, García-Calzón S, Peijs L, Hansen NS, Volkov P, Kjøbsted R, Wojtaszewski JFP, et al. Abnormal epigenetic changes during differentiation of human skeletal muscle stem cells from obese subjects. *BMC Med.* 2017;15:39.
39. Enwere EK, Lacasse EC, Adam NJ, Korneluk RG. Role of the TWEAK-Fn14-clAP1-NF- $\kappa$ B signaling axis in the regulation of myogenesis and muscle homeostasis. *Front Immunol.* 2014;5:34.
40. Mittal A, Kumar A, Lach-Trifilieff E, Wauters S, Li H, Makonchuk D, Glass D, Kumar A. The TWEAK-Fn14 system is a critical regulator of denervation-induced skeletal muscle atrophy in mice. *J Cell Biol.* 2010;188:833–49.
41. Sato S, Ogura Y, Kumar A. TWEAK/Fn14 signaling axis mediates skeletal muscle atrophy and metabolic dysfunction. *Front Immunol.* 2014;5:18.
42. Ryall JG, Dell'Orso S, Derfoul A, Juan A, Zare H, Feng X, Clermont D, Kounlis M, Gutierrez-Cruz G, Fulco M, et al. The NAD<sup>+</sup>-dependent SIRT1 deacetylase translates a metabolic switch into regulatory epigenetics in skeletal muscle stem cells. *Cell Stem Cell.* 2015;16:171–83.
43. Ramilowski JA, Goldberg T, Harshbarger J, Kloppmann E, Lizio M, Satagopam VP, Itoh M, Kawaji H, Carninci P, Rost B, et al. A draft network of ligand–receptor-mediated multicellular signalling in human. *Nat Commun.* 2015;6:7866.
44. Charrin S, Latil M, Soave S, Poleskaya A, Chrétien F, Boucheix C, Rubinstein E. Normal muscle regeneration requires tight control of muscle cell fusion by tetraspanins CD9 and CD81. *Nat Commun.* 2013;4:1674.
45. Pawlikowski B, Vogler TO, Gadek K, Olwin BB. Regulation of skeletal muscle stem cells by fibroblast growth factors. *Dev Dyn.* 2017;246:359–67.
46. Pisconti A, Bernet JD, Olwin BB. Syndecans in skeletal muscle development, regeneration and homeostasis. *Muscles Ligaments Tendons J.* 2012;2:1–9.
47. Mylona E, Jones KA, Mills ST, Pavlath GK. CD44 regulates myoblast migration and differentiation. *J Cell Physiol.* 2006;209:314–21.
48. Scimeca M, Bonanno E, Piccirilli E, Baldi J, Mauriello A, Orlandi A, Tancredi V, Gasbarra E, Tarantino U. Satellite cells CD44 positive drive muscle regeneration in osteoarthritis patients. *Stem Cells Int.* 2015;2015:469459.
49. Burzyn D, Kuswanto W, Kolodin D, Shadrach JL, Cerletti M, Jang Y, Sefik E, Tan TG, Wagers AJ, Benoist C, et al. A special population of regulatory T cells potentiates muscle repair. *Cell.* 2013;155:1282–95.
50. Low S, Barnes JL, Zammit PS, Beauchamp JR. Delta-like 4 activates Notch 3 to regulate self-renewal in skeletal muscle stem cells. *Stem Cells.* 2018;36: 458–66.
51. Tajrishi MM, Zheng TS, Burkly LC, Kumar A. The TWEAK-Fn14 pathway: a potent regulator of skeletal muscle biology in health and disease. *Cytokine Growth Factor Rev.* 2014;25:215–25.
52. Pampeno C, Derkatch IL, Meruelo D. Interaction of human laminin receptor with Sup35, the [PSI<sup>+</sup>] prion-forming protein from *S. cerevisiae*: a yeast model for studies of LamR interactions with amyloidogenic proteins. *PLoS One.* 2014;9:e86013.
53. Wu Y, Tan X, Liu P, Yang Y, Huang Y, Liu X, Meng X, Yu B, Wu M, Jin H. ITGA6 and RPSA synergistically promote pancreatic cancer invasion and metastasis via PI3K and MAPK signaling pathways. *Exp Cell Res.* 2019;379:30–47.
54. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods.* 2019;16:1289–96.
55. Barruet E, Garcia SM, Striedinger K, Wu J, Lee S, Byrnes L, Wong A, Xuefeng S, Tamaki S, Brack AS, Pomerantz JH. Functionally heterogeneous human satellite cells identified by single cell RNA sequencing. *eLife.* 2020;9:e51576.
56. Rubenstein AB, Smith GR, Raue U, Begue G, Minchev K, Ruf-Zamojski F, Nair VD, Wang X, Zhou L, Zaslavsky E, Trappe TA, Sealon SC. Single-cell transcriptional profiles of human skeletal muscle. *Sci Rep.* 2020;10:229.
57. Riddle ES, Bender EL, Thalacker-Mercer AE. Transcript profile distinguishes variability in human myogenic progenitor cell expansion capacity. *Physiol Genomics.* 2018;50:817–27.
58. Thiriout A, Perdomo C, Cheng G, Novitzky-Basso I, McArdle S, Kishimoto JK, Barreiro O, Mazo I, Triboulet R, Ley K, et al. Differential DARC/ACKR1 expression distinguishes venular from non-venular endothelial cells in murine tissues. *BMC Biol.* 2017;15:45.
59. Garcia SM, Tamaki S, Lee S, Wong A, Jose A, Dreux J, Kouklis G, Sbitany H, Seth R, Knott PD, et al. High-yield purification, preservation, and serial Transplantation of Human Satellite Cells. *Stem Cell Reports.* 2018;10:1160–74.

60. Xu X, Wilschut KJ, Kouklis G, Tian H, Hesse R, Garland C, Sbitany H, Hansen S, Seth R, Knott PD, Hoffman WY, Pomerantz JH. Human satellite cell transplantation and regeneration from diverse skeletal muscles. *Stem Cell Reports*. 2015;5:419–34.
61. Sarver DC, Sugg KB, Disser NP, Enselman ERS, Awan TM, Mendias CL. Local cryotherapy minimally impacts the metabolome and transcriptome of human skeletal muscle. *Sci Rep*. 2017;7.
62. Tarnopolsky MA, Pearce E, Smith K, Lach B. Suction-modified Bergström muscle biopsy technique: Experience with 13,500 procedures. *Muscle Nerve*. 2011;43:716–25.
63. Spinazzola JM, Gussoni E. Isolation of primary human skeletal muscle cells. *Bio-Protocol*. 2017;7:e2591.
64. Durinck S, Spellman P, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*. 2009;4:1184–91.
65. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545.
66. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2018.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

