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

Andrea J. De Micheli, Jason A. Spector, Olivier Elemento and Benjamin D. Cosgrove

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 β1-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^ink4a 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...
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).

A single sample was ~ 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 ± 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 regression 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 fibrinectin (FN1), the microfibril-associated glycoprotein MFAP5, and CD55 known to be expressed by synoviocytes [17]; and Fibroblast 3 is mainly characterized by SMO2 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 RG55 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 CXC12. Endothelial 3 express high levels of the platelet-recruiting Von Willebrand Factor (VWF) and caveolin-1 (CAVI), 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' 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.)

Table of donors:

| Donor | Age | Sex | Anatomical site       | Cells after QC |
|-------|-----|-----|-----------------------|----------------|
| 01    | 74  | M   | Right serratus        | 569            |
| 02    | 54  | M   | Flexor hallucis longus| 768            |
| 03    | 74  | M   | Orbicularis oculi     | 290            |
| 04    | 77  | F   | Eyebrow               | 727            |
| 05    | 62  | F   | Trapezius             | 3998           |
| 06    | 81  | F   | Right external oblique| 6396           |
| 07    | 71  | M   | Flexor hallucis longus| 865            |
| 08    | 54  | M   | Left vastus lateralis | 3394           |
| 09    | 70  | M   | Left external oblique | 1711           |
| 10    | 43  | M   | Left rectus abdominis | 3710           |

Diagram A: Cellular distribution across donors.

Diagram B: UMAP plots of cell types including inflammatory macrophages and monocytes, B/T/NK cells, tissue-resident macrophages, MuSCs and progenitors 1 and 2, fibroblasts 1, 2, and 3, adipocytes, erythroblasts, and myonuclei.

Diagram C: Additional UMAP plots focusing on different cell types and markers.

Diagram D: Heatmap showing average expression levels across donors.

Diagram E: Bar chart displaying cell type proportions across donors and body sections.
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 CHRD1, as well as for the cyclin-dependent kinase inhibitor CKDN1C (encoding P57Kip2), 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 ERG1. 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, while 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.)
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 coreceptor complexes with integrin β1 (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 Treg cells [49]. According to our model findings, EGFR may also interact with ligands expressed by immune cells, such as with TGF-α (TGFα), heparin-binding EGF (HBEGF), amphiregulin (AREG), and epiregulin (EREG). Other EGFR ligands include brevican (BCAN), and betacellulin (BTC) produced by endothelial cells; ECM proteins fibulin 3 (EFEMP1), decorin (DCN), and tenasin 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 a 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. 5A-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 EGF 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/ACKRI, 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/ACKRI+ 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, owning 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 IncRNAs 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 sub-populations (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 validated method [67, 68]. The single-cell suspensions were then sequentially digested with Dispase II and Collagenase D as described previously [67, 68]. The digested cells were sequentially passed through 70% Percoll, 40% Percoll, and 10% Percoll, and finally washed with PBS and treated with PAA to inactivate the protease. The resultant single-cell suspension was then stained with primary antibodies against CD133, CD271, and SSEA1, and subsequently with secondary antibodies against Alexa Fluor 555 and Alexa Fluor 488.

The stained single-cell suspensions were then loaded onto the Chromium Single Cell 3’ Library & Gel Bead Kit or Chromium Single Cell 3’ Library & Gel Bead Kit v3 for single-cell sequencing library preparation. The resulting libraries were sequenced on the Chromium platform using the Chromium Single Cell Gene Expression Kit. The resulting data were then analyzed using the Cell Ranger 3.1.0, Seurat 3.1.0, Scanorama (online version as of 2019-11-19), Harmony (online version as of 2019-11-19), biomaRt 2.43.1, Gene Set Enrichment Analysis 4.0.3, and Ingenuity Pathway Analysis (IPA, 2019-08-30) tools.

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

| Reagents | Sigma-Aldrich | 10X Genomics | biomaRt 2.43.1 | Gene Set Enrichment Analysis 4.0.3 | Ingenuity Pathway Analysis (IPA, 2019-08-30) |
|----------|---------------|--------------|----------------|------------------------------|-----------------------------------------------|
| Dispase II (neutral protease, grade II) | Sigma-Aldrich | 04942078001 | | | |
| Collagenase D, from Clostridium histolyticum | Sigma-Aldrich | 11088866001 | | | |
| Commercial kits | | | | | |
| 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) | | | |
| Deposited data | | | | | |
| Human ligand-receptor database | [43] | https://www.ncbi.nlm.nih.gov/pubmed/26198319 | | | |
| Human scRNA-seq dataset | This paper | GSE143704 | | | |
| Mouse scRNA-seq dataset | [29] | GSE143437 | | | |
| Software packages and algorithms | | | | | |
| Cell Ranger 3.1.0 (July 24, 2019) | 10X Genomics | | | | https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest|
| Seurat 3.1.0 | [14] | | | | https://github.com/satijalab/seurat|
| Scanorama (online version as of 2019-11-19) | [13] | | | | https://github.com/brianhie/scanorama|
| Harmony (online version as of 2019-11-19) | [54] | | | | https://github.com/immunogenomics/harmony|
| biomaRt 2.43.1 (online version as of 2020-01-08) | [64] | | | | https://bioconductor.org/pkg/packages/release/bioc/html/biomaRt.html|
| Gene Set Enrichment Analysis 4.0.3 | [65] | | | | http://software.broadinstitute.org/gsea/index.jsp|
| Ingenuity Pathway Analysis (IPA, 2019-08-30) | QIAGEN | | | | https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/|
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 °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 ~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 >log₂(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 function from Scanorama. The integrated values 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 log₂ 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 types) (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 unjinjured 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. Transcriptomic 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 R01AG066360 (to B.D.C.), a National Institutes of Health 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 # 1510017612) 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

1Meing School of Biomedical Engineering, Cornell University, Ithaca, NY 14853, USA. 2Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY 10021, USA. 3Division 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

1. Bentzinger CF, Wang YX, Dumont NA, Rudnicki MA. Cellular dynamics in the muscle satellite cell niche. EMBO Rep. 2013;14:1062–72.
2. Blau HM, Cosgrove BD, Ho ATV. The central role of muscle stem cells in regenerative failure with aging. Nat Med. 2015;21:854.
3. Järvinen TA, Järvinen M, Kalimo H. Regeneration of injured skeletal muscle after the injury. Muscles Ligaments Tendons J. 2014;3:37–45.
4. Addison O, Marcus RL, LaStayo PC, Ryan AS. Interspersed fat: a review of adipogenic progenitors within human skeletal muscle. Stem Cells. 2010;28:427–511.
5. Alexander MS, Rozkalne A, Colletta A, Spinazzola JM, Johnson S, Rahimov F, Meng H, Ravlov 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.
6. Charville GW, Cheung TH, Yoo B, Santos PJ, Lee GK, Shager JB, Rand R. Ex vivo expansion and in vivo self-renewal of human muscle stem cells. Stem Cell Reports. 2015;5:562–32.
7. Pisani DF, Clement N, Loubat M, Saconi 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.
8. Uezumi A, Nakatani M, Ikemoto-Uezumi M, Yamamoto N, Morita M, Yamaguchi A, Yamada H, Kusai T, Masuda S, Narta A, et al. Cell-surface protein profiling identifies distinctive markers of progenitor cells in human skeletal muscle. Stem Cell Reports. 2016;7:263–78.
9. Wang YX, Feige P, Brun CE, Hekmatnejad B, Dumont NA, Renaud J-M, Faulkes S, Guirond 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.
10. Sousa-Victor P, Gunna S, Guirond 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.
11. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet. 2019;20:257–72.
12. Hie B, Lysyov B, Berger B. Efficient integration of heterogeneous single-cell data across datasets using Scanorama. Nat Biotechnol. 2019;37:685–91.
13. Hie B, 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.
14. Mufti J, Walker DW, Aliporeprotein D: an overview of its role in aging and age-related diseases. Cell Cycle. 2010;9:269–73.
16. Cereijo R, Gavaldá-Navarro A, Cairó M, Quesada-López T, Villarroya J, Morón-Ros S, Sánchez-Infantes D, Peyrou M, Iglesias R, Rampet 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. Karups ON, Kiener HP, Niederreiter B, Vilmaz-Elis AS, van der Kaa J, Ramaglia–CS, Robin C, van Oudenaarden A. Single-cell sequencing reveals heterogeneity in the cellular composition of mouse Achilles tendons. BioRev. 2020b;8(126).

18. 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;2017:979375.

19. 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;106(1):112–9.

20. Fernández-Hernando C, Yu J, Dívalos 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.

21. Volonte D, Liu Y, Galbiati F. The modulation of caveolin-1 expression controls satellite cell activation during muscle repair. FASEB J. 2004;18:237–9.

22. Fujimoto N, He Y, D’Adzio 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(10):e3000704.

23. Swindell WR, Johnston A, Xiong 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.

24. Kuang S, Chargé SB, Seale P, Huh M, Rudnicki MA. Distinct roles for Pax7 in lymph node myofibers. Cell Stem Cell. 2015;16:171–8.

25. Ramlowski JA, Goldberg T, Hashbarger J, Koppmann E, Listo M, Satapathy VP, 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;180(8):8833–49.

26. Sato S, Ogora Y, Kumar A. TWEAK-Fn14 signaling axis mediates skeletal muscle atrophy and metabolic dysfunction. Front Immunol. 2014;5:18.

27. Ryall JD, Dell’Oro S, Derfoul A, Anfi A, Hare J, Cheng X, Clermont D, Cotrina MD, Lecky G, Aslan E, Cao Y, et al. The NAD + dependent SIRT1 deacetylase translates a metabolic switch into regulatory epigenetics in skeletal muscle stem cells. Cell Stem Cell. 2015;16:171–8.

28. Aebischer P, van den Brink SC, Sage F, Vértesy Á, Spanjaard B, Peterson-Maduro J, Baron M, Hansen NS, Volkov P, Kjestedt R, Wojszawski JF, et al. Abnormal epigenetic changes during differentiation of human skeletal muscle stem cells from obese subjects. BMC Med. 2017;15:39.

29. Emwere BK, Lacasse EC, Adam NJ, Korneluk RG. Role of the TWEAK-Fn14–CD27–NF-kB signaling axis in the regulation of myogenesis and muscle homeostasis. Front Immunol. 2014;5:34.

30. 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;180(8):8833–49.

31. Dell’Oro S, Jiang Y, Leblanc L, Durrant N, Conforti M, Bootsma D, Korneluk RG. The TWEAK-Fn14–Sup35, the [PSI+] prion-forming protein from S. cerevisiae: a yeast model for studies of Lamk interactions with amyloidogenic proteins. PLoS One. 2013;8(1):e50555.

32. Zhang L, Uezumi A, Kaji T, Tsujikawa K, Andersen DC, Jensen CH, Fukuda S. Expression and Functional Analyses of Dlk1 in Muscle stem cells and mesenchymal progenitors during muscle regeneration. Int J Mol Sci. 2019;20(3):1269.

33. Catalán V, Gómez-Ambrosio J, Rodríguez A, Ramírez B, Ortega VA, Hernández-Lizain JA, Baquial J, Becerril S, Rotellar F, Valenti V, et al. IL-32-induced inflammation constitutes a link between obesity and colon cancer. Oncoimmunology. 2017;6:e1283388.

34. Davegård C, Boholm C, Perhyvel A, Henricksen T, García-Calbón S, Pejs J, Hansen NS, Volkov P, Kjestedt R, Wojszawski JF, et al. Abnormal epigenetic changes during differentiation of human skeletal muscle stem cells from obese subjects. BMC Med. 2017;15:39.

35. Emwere BK, Lacasse EC, Adam NJ, Korneluk RG. Role of the TWEAK-Fn14–CD27–NF-kB signaling axis in the regulation of myogenesis and muscle homeostasis. Front Immunol. 2014;5:34.

36. Fujimoto N, He Y, D’Adzio 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(10):e3000704.
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.