Differentiation of the human PAX7-positive myogenic precursors/satellite cell lineage in vitro

Ziad Al Tanoury1,2,3, Jyoti Rao2,3, Olivier Tassy1, Bénédicte Gobert1,4, Svetlana Gapon2, Jean-Marie Garnier1, Erica Wagner2, Aurore Hick4, Arielle Hall5, Emanuela Gussoni5 and Olivier Pourquie1,2,3,6,*

ABSTRACT

Satellite cells (SC) are muscle stem cells that can regenerate adult muscles upon injury. Most SC originate from PAX7+ myogenic precursors set aside during development. Although myogenesis has been studied in mouse and chicken embryos, little is known about human muscle development. Here, we report the generation of human induced pluripotent stem cell (iPSC) reporter lines in which fluorescent proteins have been introduced into the PAX7 and MYOG loci. We use single cell RNA sequencing to analyze the developmental trajectory of the iPSC-derived PAX7+ myogenic precursors. We show that the PAX7+ cells generated in culture can produce myofibers and self-renew in vitro and in vivo. Together, we demonstrate that cells exhibiting characteristics of human fetal satellite cells can be produced in vitro from iPSC, opening interesting avenues for muscular dystrophy cell therapy. This work provides significant insights into the development of the human myogenic lineage.

KEY WORDS: PAX7, Skeletal muscle, Human development, Pluripotent stem cell, Satellite cell

INTRODUCTION

Adult skeletal muscles are endowed with significant regenerative capacity upon injury. This is made possible thanks to a small subpopulation of quiescent adult stem cells called satellite cells (SC) (Brack and Rando, 2012; Dumont et al., 2015). Most SC express the transcription factor paired box 7 (Pax7) and lie under the basal lamina of muscle fibers. Upon muscle damage, the SC become activated, proliferate and generate new fibers to reconstitute a healthy muscle. SC are also able to self-renew to recreate a quiescent muscle stem cell compartment. The mouse Pax7+ population is heterogeneous, with different populations of cells exhibiting different regenerative capacities (Cornelison, 2018; Tierney and Sacco, 2016). Whether this heterogeneity reflects distinct subpopulations or different maturation stages of the same lineage is presently unknown.

SC derive from the paraxial mesoderm, the embryonic tissue that forms the vertebral column and skeletal muscles (Chal and Pourquié, 2017; Gros et al., 2005; Hutcheson et al., 2009; Kassar-Duchossoy et al., 2005; Lepper and Fan, 2010; Relaix et al., 2005; Schienda et al., 2006). Pax3+ myogenic precursors arise from the dorsal epithelial compartment of the somite called the dermomyotome. Pax3+ cells of the dermomyotome lips activate Myf5, then downregulate Pax3 and delaminate and differentiate into elongated post-mitotic myocytes expressing myogenin (Myog) to form the first embryonic muscles called myotomes (Chal and Pourquié, 2017). At the same time, Pax3+ myogenic precursors delaminate from the lateral edge of the dermomyotome and migrate to the lateral plate to form the rudiments of the limbs, girdles, tongue and diaphragm muscles. This early phase of myogenesis, called embryonic or primary myogenesis, generates a first set of muscle fibers that serves as a scaffold for the secondary or fetal myogenesis (Biressi et al., 2007a; Chal and Pourquié, 2017). This next phase is initiated in the mouse embryo by a population of Pax7+expressing myogenic precursors originating from the dermomyotomal Pax3+ precursors (Hutcheson et al., 2009). These highly proliferative Pax7+ myogenic precursors are first found intermingled with primary muscle fibers (Gros et al., 2005; Hutcheson et al., 2009). Some of the Pax7+ cells activate Myf5 to become proliferative myoblasts, which eventually express Myog and exit the cell cycle to form myocytes, which fuse to form the fetal muscle fibers (Chal and Pourquié, 2017; Kassar-Duchossoy et al., 2005; Relaix et al., 2005). The Pax7+ cells become enclosed in the basal lamina of myofibers around embryonic day (E)16.5 to E18 in mouse (Ontell and Kozeka, 1984; Relaix et al., 2005). They then continue to proliferate, and many of the Pax7+ cells fuse with the developing fibers, increasing their number of nuclei (White et al., 2010). The myonuclear accretion of SC considerably slows down after birth and the SC that are left remain associated with the myofibers, which continue to grow by hypertrophy. Lineage tracing experiments show that incorporation of Pax7+ cells into mature muscle fibers proceeds throughout adult life, even in the absence of regeneration (Keefe et al., 2015; Pawlikowski et al., 2015).

We and others have established differentiation protocols recapitulating myogenesis in vitro from mouse pluripotent stem cells (PSC) such as embryonic stem cells (ESC) or reprogrammed stem cells (induced pluripotent stem cells, iPSC) (Chal et al., 2018; Magli and Perlingleo, 2017). In mouse, epiblast-like cells can be efficiently induced to a presomatic mesoderm fate characterized by Msgn1 and Pax3 expression by Wnt activation, in combination with BMP inhibition (Chal et al., 2015; Diaz-Cuadros et al., 2020). When exposed to myogenic growth factors, these cells downregulate Pax3 and activate Pax7 as observed in mouse myogenic precursors in vivo (Chal et al., 2018, 2015; Hutcheson et al., 2009; Relaix et al., 2005). Long stratified and multinucleated myofibers exhibiting characteristics of perinatal myofibers form after 3-4 weeks in vitro (Chal et al., 2015). Such cultures recapitulate mouse embryonic and fetal skeletal
myogenesis in vitro (Chal et al., 2015). These cultures also generate Pax7+ myogenic precursors/fetal SC (Chal et al., 2018, 2015). In cultures of a mouse Pax7GFP ESC reporter line, GFP expression appears around day 9 and peaks around 20% at 2-3 weeks (Chal et al., 2018, 2015). After 3 weeks in vitro, fast myosin heavy chain (MyHC)-positive (MyHC+) fibers are interspersed with a population of small cells expressing Pax7. These cells are sometimes found enclosed in the basal lamina, resembling SC positioning in vivo (Chal et al., 2015; Relaix et al., 2005; Sambasivan and Tajbakhsh, 2007). These dynamics of Pax7 cell production in vitro is strikingly similar to that reported in the developing mouse (Hutcheson et al., 2009; Relaix et al., 2005; Sambasivan and Tajbakhsh, 2007). The Pax7+ cells produced in vitro can engraft in adult mouse muscles in vivo and contribute Pax7+ satellite-like cells located under the basal lamina of myofibers (Chal et al., 2018, 2015). Thus, the mouse myogenic cultures spontaneously recreate a niche, allowing the differentiation of Pax7+ cells exhibiting a similar regenerative potential to endogenous SC.

SC showing similar features to those of mice have been identified in humans (Barruet et al., 2020; Charville et al., 2015; Xu et al., 2015). However, their origin and embryonic development is largely unknown. Owing to the difficulty in accessing human embryonic tissue, very few studies have analyzed human myogenesis at the molecular and cellular level (Belle et al., 2017; Hicks et al., 2018; Xi et al., 2017). Early electron microscopy studies of the developing gastrocnemius of human embryos have shown that a transition resembling the switch from embryonic to fetal muscles of rodents begins around gestational week 10 (Ishikawa, 1966). Recently, studies in cleared human embryos ranging from 8 to 14 weeks of gestation demonstrated abundant expression of PAX7, MYOG and MyHC in developing muscles (Belle et al., 2017). MyHC was detected at week 8 in muscles that were colonized with motoneuron axons (Belle et al., 2017). Thus, although the data suggests that aspects of the prenatal stages of muscle development in humans resemble those of mice, our understanding of human myogenesis remains extremely limited.

Transposition of the protocols developed for mouse to human ESC/iPSC has demonstrated that long striated myofibers and Pax7+ cells can also be produced in vitro from PSC (Chal et al., 2015; Hicks et al., 2018). Such cultures offer an excellent proxy to study the development of the human myogenic lineage. Here, we report a detailed characterization of the development of human skeletal muscle in vitro using two human iPSC reporter lines (PAX7Venus+ and MYOGVenus+), which identify subpopulations of the differentiating myogenic lineage. Using single cell RNA sequencing (scRNA-seq) of fluorescence-activated cell sorting (FACS)-sorted PAX7Venus+ populations, we characterize the developmental stages of the Pax7+ precursors produced in vitro. We also demonstrate the myogenic potential of the human Pax7+ cells in vivo and in vitro. Our analysis demonstrates that these cells exhibit characteristics of the fetal SC lineage, which could be used for the development of cell therapy approaches for muscle degenerative diseases such as Duchenne Muscular Dystrophy.

RESULTS AND DISCUSSION
To study the differentiation of human iPSC towards the SC lineage, we targeted a Venus (a YFP derivative; Nagai et al., 2002) reporter cassette into exon 1 of the human PAX7 gene in the human NCRM1 iPSC line. The fluorescent protein was introduced by homologous recombination using CRISPR-Cas9 to produce a knock-in allele expressing a 2A peptide fused to a nuclear NLS-Venus fluorescent protein in frame with exon 1. This allows for a fusion transcript but distinct protein products (Fig. S1A). We next differentiated the targeted cells according to an established myogenic differentiation protocol (Chal et al., 2016) (Fig. 1A). The Venus nuclear signal first appeared after 2 weeks of differentiation in a subpopulation of mononuclear cells (Fig. 1B). After 21 days of primary differentiation and after replating and secondary differentiation in vitro, the reporter cells generated long striated myofibers (>1 mm) intermingled with Pax7+Venus+ mononucleated cells (Fig. 1A,C,D). Using anti-GFP and anti-PAX7 antibodies, we showed that Venus expression largely overlaps with that of PAX7 (Fig. 1E-G). From 20 to 30 days, Venus+ cells constitute ∼20-25% of the mononucleated cell population in the culture, as shown by FACS analysis (Fig. 1B,H). Moreover, when the Venus fraction was FACS-sorted from 21-day cultures, the PAX7 transcript was exclusively detected in the Venus fraction, validating the reporter specificity (Fig. 1I). The percentage of Venus+ cells decreased after 3 weeks, consistent with the dynamics of Pax7+ cells in vivo (Fig. 1B). When 3-4 week cultures were dissociated and replated in proliferation medium (SKGM) for 1-2 days before transferring into differentiation medium for a week (secondary differentiation, Fig. 1A), 20-25% of the mononucleated cell fraction was Venus+ (Fig. 1B). These results contrast with the recently reported differentiation of a human PAX7/MYF5 ESC reporter line in a different medium in which the peak of PAX7 expression was observed at day 4 in vitro (Wu et al., 2018). Our detailed scRNA-seq analysis of early myogenic cultures shows that at day 4, cells of the myogenic lineage are still at the presomitic mesoderm stage, significantly earlier than the first activation of Pax7 in mouse myogenic precursors (Chal et al., 2018; Diaz-Cuadros et al., 2020; Hutcheson et al., 2009). This suggests that the Pax7+ cells differentiating early in this study (Wu et al., 2018) more likely correspond to neural crest cells rather than to myogenic precursors.

We next compared the transcriptome of the Pax7Venus+ cells generated after 3 weeks (day 21) of myogenic differentiation in vitro with that of undifferentiated reporter iPSC. Pax7Venus+ cells upregulated several known markers of quiescent SCs, such as the vascular cell adhesion molecule 1 (VCAM1), M-cadherin (CDH15), MYF5, Nfia, PAX7 and NOTCH3 (Charville et al., 2015; Fukada et al., 2007; Liu et al., 2013; Pietrosomoli et al., 2017) (Table S1). Some genes coding for sarcomeric proteins, such as myosin light chain 1 (MYL1), which are normally associated with later stages of myogenic differentiation, were also upregulated, suggesting that more differentiated cells could be included in the Pax7Venus+ population due to the stability of the Venus protein. We also compared the transcriptome of human Pax7Venus+ myogenic progenitors from 3-week-old cultures with mouse Pax7GFP+ cells from FACS-sorted age-matched in vitro cultures (Chal et al., 2015). As observed for human cells, mouse Pax7+ cells upregulated well-known satellite cell markers, including Myf5, Cdh15, Nfia, Igf7 and Cds5 (Ncam1) (Tables S1 and S2). Thus, this analysis shows that the human Pax7Venus+ cells differentiated in vitro share a common transcriptional signature with the mouse ESC-derived Pax7GFP+ myogenic precursor cells. Together, our data indicate that expression of the Pax7Venus reporter allows tracking of the endogenous expression of Pax7 protein in human myogenic precursors/fetal SC differentiating in vitro.

To analyze the commitment of differentiating human iPSC toward the skeletal muscle fate in vitro, we also generated a MYOG-Venus reporter line by targeting a 2A peptide-NLS-Venus-reporter cassette into the stop codon located in exon 3 of MYOG (Fig. S1B). When these reporter cells were differentiated in myogenic conditions (Chal et al., 2016), we detected Venus+ myocytes as early as 10 days in culture (Fig. 1B), i.e. before Pax7 but after Pax3 expression (Chal et al., 2016). These cells likely correspond...
to myogenic cells generated during the equivalent of the primary or embryonic phase from PAX3+ precursors (Chal et al., 2016; Hutcheson et al., 2009). The MYOGVenus+ cells progressively fused to produce long myotubes expressing sarcomeric proteins such as α-actinin (ACTA2) (Fig. 1J,K). Venus protein expression overlapped extensively with the MYOG protein (Fig. 1L-N). After 3 weeks of differentiation in vitro, Venus+ cells peaked to 12-15% of the mononucleated cell fraction (Fig. 1B,O). The MYOG transcript was only detected in the Venus+ population in FACS-sorted cells, thus supporting the specificity of the reporter (Fig. 1P).
Thus at 3-4 weeks of differentiation, the proportion of myogenic cells (PAX7+ and MYOG+) in the mononucleated fraction of the culture is ~30%. The remaining fraction appears to be mostly composed of fibroblastic populations.

We next compared the transcriptome of the MYOGVenus+ cells with that of undifferentiated reporter iPSC. In contrast to the PAX7Venus+ cells, the FACs-sorted MYOGVenus+ progenitors from 27-day cultures showed a clear myogenic identity. They expressed muscle-specific transcription factors such as MYOG, MYOD1 and MEF2C, as well as sarcomeric proteins, including the myosin light (MYL) and heavy (MYH) chain isoforms MYL1, MYL4, MYH3 (embryonic), MYH8 (perinatal), troponin C type 2 (TNVC2) and titin (TTN), and many muscle-specific proteins such as muscle creatine kinase (CKM) (Table S3). Thus, the MYOGVenus reporter line allows tracking of the differentiation of human myocytes in vitro.

We next performed a Gene Ontology (GO) analysis of the genes differentially expressed between the PAX7Venus+ and the MYOGVenus+ cells. This identified distinct over-represented pathways between the two populations. The MYOGVenus+ cells showed enrichment in GO categories such as muscle protein, muscle filament sliding, Z disc or muscle organ development, whereas extracellular matrix (ECM) organization was in the top GO categories for the PAX7Venus+ cells (Tables S1 and S3). Several genes coding for ECM proteins such as lumican (LUM), decorin (DCN), and COL3A1 were among the most strongly upregulated genes in PAX7Venus+ compared with undifferentiated iPSC (Tables S1 and S3). Tenascin-C (TNC), previously described as one of the most highly upregulated genes in mouse fetal SC when compared with adult SC (Tierney et al., 2016), was also highly upregulated in PAX7Venus+ cells. Together, these data show that the MYOGVenus+ cells represent a more differentiated population of myogenic cells compared with the PAX7Venus+ cells.

To analyze the heterogeneity of the differentiating human PAX7 cell population in vitro, we performed scRNA-seq of the PAX7Venus+ cell population from 30-day-old myogenic cultures. We analyzed 1427 FACs-sorted PAX7Venus+ cells from two different myogenic cultures using the inDrops platform (Klein et al., 2015). Both myogenic cultures revealed similar patterns of heterogeneity and clustering (Fig. S2A). The Louvain-based unsupervised clustering analysis identified four different clusters (Fig. 2A, Table S4). The clusters are organized as a continuum, arguing for limited heterogeneity among the different populations composing the PAX7Venus+ fraction.

Two of the clusters are characterized by expression of PAX7 transcripts and lack of MYOG expression (Fig. 2B). A first cluster (cluster 1) representing 18% of the PAX7Venus+ cells formed a distinct cluster of cycling cells that expressed genes involved in cell division such as MKI67 or PCNA (green; Fig. 2A,B, Fig. S3). A second cluster (cluster 2) contains the largest number of PAX7+ cells (60%) (blue; Fig. 2A,B, Fig. S2). We could not identify markers strictly specific for this cluster. The most specific markers such as MYF3 or CXC4 were also shared with the cluster of cycling cells (Fig. 2B, Fig. S3). Gene signatures specific for mouse and human SC, including YCAM1, MYF5, NFIA, ITGA7 and CAV1 were enriched in these two clusters (Barruet et al., 2020; Charvile et al., 2015; Fukada et al., 2007; Liu et al., 2013) (Fig. 2B, Fig. S3). These cells were also enriched in genes coding for ECM proteins such as COL6A1, COL3A1, TNC, DCN and GPC3, and for effectors of cell signaling (Figs S2B and S3). Members of the TGFβ pathway (such as TGFB1 and TGFBRII), which is involved in the control of myogenesis, were upregulated in these two clusters (Biressi et al., 2007b; Hicks et al., 2018) (Fig. S3) along with targets of the FGF (SPRY1) and YAP pathway (CYR61; also known as CCN1) (Fig. S3) (Chakkalakal et al., 2012; Esteves de Lima et al., 2016; Flanagan-Stee et al., 2000).

A third cluster (cluster 3) includes cells expressing both PAX7 and MYOG (13% of total cells; orange; Fig. 2A,C). Genes enriched in this cluster include several genes associated with muscle differentiation and fusion, such as the SC marker S0X8 (Schmidt et al., 2003), the transcription factor HES6 (Gao et al., 2001), creatine kinase B (CKB) (Simionescu-Bankston et al., 2015), CDH15 (Martí et al., 2013) and the small fusogenic peptide myomixer (MIMX), which in mouse peaks in skeletal muscles at E14.5 (Bi et al., 2017) (Figs S2 and S3). This cluster was located between the PAX7 MYOG− clusters and a fourth cluster (cluster 4) characterized by lack of PAX7 and expression of MYOG (8% total cells; red; Fig. 2A,C), suggesting that cluster 3 represents a transition state between PAX7 MYOG− and PAX7 MYOG+ cells. We also detected expression of the markers of human myogenic progenitors/fetal SC NGFR, ERBB3 and CD82 in virtually all PAX7+ cells (Alexander et al., 2016; Hicks et al., 2018) (Fig. 2B).

Cells of cluster 4 lack PAX7 and express myocyte markers such as MYOG or MEF2C as well as sarcomeric proteins usually expressed by post-mitotic myogenic cells such as ACTA1, TTN, TNCC2, MYL1 or MYH8, indicating the differentiated status of these cells (Fig. 2B, Fig. S3). The temporal stability of the Venus protein likely explains why cells that have downregulated PAX7 mRNA to start their myogenic differentiation are captured as part of the PAX7Venus+ population (Relaix et al., 2005). Cells of the two MYOG− clusters organized along a gradient of differentiation, with cells progressing from expressing the determination and signaling genes associated with muscle maturation and fusion described above to structural genes found in post-mitotic myocytes such as MYH3 (embryonic MHC) and MYH8 (fetal MHC) (Fig. 2B, Fig. S3).

This analysis suggests that the PAX7 MYOG+ cells gives rise to a population of PAX7 MYOG+ cells, which differentiate into PAX7 MYOG+ fetal myocytes. This developmental trajectory is supported by PAGA graph analysis (Wolf et al., 2019) (Fig. 2C,D). Interestingly, both PAX7 MYOG− clusters are directly linked by an edge to the PAX7 MYOG+ cluster, suggesting that they could both directly contribute to this fate. This trajectory is reminiscent of the differentiation sequence reported for chicken and mouse myogenesis (Esteves de Lima et al., 2014; Kassar-Duchossoy et al., 2005; Manceau et al., 2008; Murphy and Kardon, 2011; Relaix et al., 2005; Zale et al., 2014).

We next used scRNA-seq data to analyze the cell cycle status of the PAX7Venus+ cells (Kowalczyk et al., 2015). Based on this analysis, ~68% of the cells were in G0/G1 state and the remaining cells were in G2/M and S phase (Fig. 3A,B). We confirmed this data with an analysis of DNA content of FACs-sorted PAX7Venus+ cells by flow cytometry. This identified around 70% cells in the G0/G1 state (Fig. 3C,D). Approximately 6% of cells were in S phase and 14% had increased their DNA content, suggesting they were in G2/M phase. To further characterize the proliferating cells in the PAX7Venus+ fraction, we labeled 21-day primary PAX7Venus reporter cell cultures with EdU for 40 h and FACs-sorted the cells based on Venus expression. Immunohistochemistry with an anti-PAX7 antibody showed that ~98% of the sorted cells expressed the PAX7 protein (Fig. 3F). Among the fraction of cells expressing the PAX7 protein, ~80% of the cells also expressed EdU indicating that they are actively proliferating (Fig. 3F). We next performed cell cycle scoring and reexpressing the genes pertaining to S and G2/M phase from the cluster of PAX7 MYOG− cyclic cells. This resulted in cells of cluster...
becoming mostly incorporated into cluster 2 (Fig. 3G,H). Thus, the two $\text{PAX7}^+\text{MYOG}^-$ cell clusters represent cells of similar identity but in different phases of the cell cycle. Our data is consistent with most $\text{PAX7}^+\text{MYOG}^-$ cells representing a single population of slow cycling cells with a long G1 phase.

One-fifth of the $\text{PAX7}^+$ cells did not incorporate EdU, suggesting that they are either quiescent or post-mitotic. These cells likely include cluster 3 cells which express $\text{MYOG}$ (a gene expressed mostly in post-mitotic cells) and account for ~13% of the $\text{PAX7}^+\text{MYOG}^-$ cells. Cluster 3 (and cluster 4) cells also upregulate $\text{CDKN1C}$ (p57kip2), which controls exit from the cell cycle in myogenic cells in parallel with $\text{MYOG}$ (Zhang et al., 1999) (Fig. 3E). This supports the conclusion that clusters 3 and 4 (representing 21% of the $\text{PAX7Venus}^+$ total population) are composed of post-mitotic cells committed to myogenic differentiation. Cell cycle analysis of mouse myogenesis has shown that myogenic precursors exit the cell cycle in the G1 phase (Esteves de Lima et al., 2014). This therefore supports the idea that $\text{PAX7}^+$ precursors activate $\text{MYOG}$ and exit G1 phase to become post-mitotic and downregulate $\text{PAX7}$.

Notch signaling plays an important role in maintaining the pool of Pax7 cells in mouse, controlling the balance between progenitors and their post-mitotic descendants and maintaining the quiescence of adult SC (Bröhl et al., 2012; Low et al., 2018; Mourikis and Tajbakhsh, 2014). During fetal myogenesis, Notch signaling is activated in Pax7 precursors by ligands such as Dll1 presented by post-mitotic cells (Schuster-Gossler et al., 2007). This leads to expression of the transcription factors Hes1 and Hey1, which are important to prevent exit from the cell cycle of the Pax7 cells and thus maintain the progenitor pool (Zalc et al., 2014). We observed a very similar situation in differentiating human iPSC, with strong expression of the Notch receptors $\text{NOTCH2}$ and $\text{NOTCH3}$ (and to a lesser extent $\text{NOTCH1}$) and their targets $\text{HEY1}$ and $\text{HES1}$ (Bröhl et al., 2012; Mourikis and Tajbakhsh, 2014) in the $\text{PAX7}^+\text{MYOG}^-$ clusters. The ligand $\text{DLL1}$ is expressed by the differentiated $\text{MYOG}^+$ cells (Fig. 4A). This suggests that, as observed during development, Notch is activated in immature myogenic precursors in response to ligands presented by differentiating myocytes. Moreover, Notch signaling has been shown to repress $\text{CDKN1C}$ in muscle progenitor cells in mouse (Zalc et al., 2014), which is in line with our observation that its mRNA expression is lower in the $\text{PAX7}^+\text{MYOG}^-$ clusters 1 and 2 (Fig. 3E). To directly test the effect of Notch signaling on the differentiation of human Pax7 precursors in vitro, we treated replated cultures of the Pax7 reporter line in differentiation medium with the drug DAPT, which inhibits...
Notch cleavage. When compared with control, we could not detect any PAX7Venus+ cells in the treated cultures (Fig. 4B-G). These cultures presented large α-actinin-positive myofibers. Together, these data argue that Notch signaling is required for the maintenance of the pool of PAX7 progenitors. Thus, our data showing that Notch signaling inhibition is incompatible with the maintenance of the SC pool in vitro, raises concerns about the use of DAPT when differentiating myogenic cells from ESC/iPSC in culture (Choi et al., 2016; Selvaraj et al., 2019).

To determine their myogenic potential, we FACS-sorted 3-week-old human PAX7Venus+ cells generated in vitro followed by replating in SkGM medium for 1-2 days to allow cells to reach 80-90% confluency (Figs 1A and 5A). Cells were then transferred to differentiation medium, in which they soon acquired an elongated shape indicating their differentiation into myoblasts. Cell fusion was observed after 2 days, resulting in the formation of millimeter-long myofibers after 7 days. These fibers expressed fast MyHC, displayed typical striations (Fig. 5B,C) and showed spontaneous twitching, indicating a mature contractile apparatus (Movie 1). These striated myofibers were interspersed with mononucleated cells expressing PAX7-Venus (Fig. 5C), suggesting that the PAX7Venus+ cells are able to differentiate into myofibers and to self-renew in vitro. Similar observations were made with mouse Pax7 GFP+ FACS-sorted progenitors from 3-week old cultures, which were also able to differentiate into striated multi-nucleated fibers when cultured in the same conditions (Fig. 5D-F). Serial sorting and differentiation of human PAX7Venus+ cells showed that this population can be maintained over repeated sorting and replating rounds, further supporting the existence of a self-renewing population in the PAX7Venus+ cells (Fig. 5G).

To determine the myogenic potential of the human PAX7Venus+ cells in vivo, we transplanted 10^5 FACS-sorted PAX7Venus+ cells from 3-week-old myogenic cultures into the tibialis anterior (TA) muscle of NOD; Rag1<sup>−/−</sup>; Dmd<sup>mdx-5Cv</sup> or RagB6 immunodeficient mice. Cardiotoxin treatment was performed 24 h before transplantation to induce myofiber injury but no X-irradiation of the TA muscles was...
performed before transplantation. Thus, endogenous mouse SC were also able to contribute to myofiber repair in vivo and potentially compete with the exogenous human PAX7Venus+ cells. Mice were sacrificed after 6-8 weeks. Human-derived myofibers and mononucleated cells were identified using species-specific antibodies for human spectrin, lamin B2 or lamin A/C, a combination of markers that simultaneously identifies human muscle fibers and human nuclei (Alexander et al., 2016). Most spectrin-positive myofibers contained at least one human lamin A/C-positive nucleus within the section analyzed, demonstrating that cells of human origin had successfully engrafted (Fig. 5H). Six weeks after transplantation, clusters of human-derived fibers expressing dystrophin and lamin B2 or spectrin and lamin A/C (Fig. 5I) were identified on sections along most of the length of the transplanted TA in the NOD; Rag1−/−; Dmdmdx-5Cv mice. Dystrophin+ muscle fibers were of varying diameter and often contained both peripheral and central human nuclei. Dystrophin expression most likely derives from the grafted human cells, as this strain of mdx mice exhibit very few revertants (Danko et al., 1992), and thus the probability of detecting such a revertant harboring a human nucleus is extremely low. FACS-sorted CD82+/CD56+ human fetal myogenic precursor cells were injected as a positive control and analyzed using similar antibody combinations (Alexander et al., 2016; Rozkalne et al., 2014). Up to 45 human fibers were identified in sections of PAX7Venus+ injected TA (34.81±11.52; mean±s.d.). A comparable number of engrafted fibers were identified when CD82+/CD56+ human fetal myogenic cells were injected (49.85±19.30), indicating that PAX7Venus+ cells show a comparable regenerative potential to these cells (Fig. 5J). These ratios of human fibers/injected cells are similar to those obtained when human iPSC differentiated to a myogenic fate are sorted based on ERBB3 or NGFR expression before injection in vivo (Hicks et al., 2018). Importantly, some human cells (human lamin A/C+) located underneath the myofiber basal lamina were also observed to express PAX7, indicating that human PAX7Venus+ transplanted cells can contribute to the SC pool (arrows, Fig. 5K-O). Therefore, the human PAX7Venus+ cells generated in vitro exhibit similar regenerative properties to mouse and human myogenic precursors/fetal SC (Chal et al., 2015; Tierney et al., 2016).

**Conclusion**

Here, we dissected the progression of human myogenesis in vitro using reporter iPSC lines expressing fluorescent proteins driven by the myogenic genes PAX7 and MYOG. We particularly focused on the differentiation of the PAX7-expressing lineage, which gives rise to adult muscle stem cells (SC) (Brack and Rando, 2012; Dumont et al., 2015). We previously reported that, in these culture conditions, a population of human PAX3+ cells is detected after 1 week of differentiation. This population subsequently expands during the second week in culture (Fig. 1Q) (Chal et al., 2016). Here, we observed that MYOG is first expressed in the myogenic cultures around day 10-12, before PAX7, which only appears at day 14. This is consistent with the existence of a first phase of primary or embryonic myogenesis initiated by PAX3+/PAX7− precursors as demonstrated for mouse in vivo and in vitro (Fig. 1Q) (Biressi et al., 2007b; Hutcheson et al., 2009; Kassar-Duchossoy et al., 2005; Relaix et al., 2005). This phase would be followed by a second phase of secondary or fetal

![Fig. 4. Notch signaling is required for the maintenance of the PAX7Venus+ cells in vitro.](image-url)
myogenesis, during which myofibers form from PAX7+ mononucleated cells, which are also able to self-renew (Hutcheson et al., 2009).

We also demonstrate using scRNA-seq, that our differentiation conditions can yield a pure myogenic human PAX7+ cell population free of neural contaminants (Kim et al., 2017). These cells express bona fide markers of fetal myogenic precursors/SC such as CD82, VCAM1, ITGA7, TNC and CAV1, and they can generate myofibers and self-renew both in vivo and in vitro, which are features of fetal SC in mouse (Alexander et al., 2016; Barruet et al., 2020; Biressi et al., 2007b; Tierney et al., 2016).

Analysis of the scRNA-seq data allowed us to characterize the developmental trajectory of the human PAX7 myogenic lineage in vitro. We have identified a large population of cycling PAX7+ myogenic precursors characterized by a long G1 phase. These PAX7+ cells subsequently activate MYOG and start expressing genes involved in the control of fusion, such as MYMX (Sampath et al., 2018) or CKB (Simionescu-Bankston et al., 2015). They eventually downregulate PAX7 and activate genes coding for sarcomeric proteins expressed in post-mitotic myocytes.

Together, our data demonstrate that myogenic commitment of human iPSC in vitro occurs through a sequence of stages resembling...
muscle development of mouse and chicken embryos in vivo (Esteves de Lima et al., 2014; Kassar-Duchossoy et al., 2005; Manceau et al., 2008; Murphy and Kardon, 2011; Picard and Marcellé, 2013; Relaix et al., 2005; Zalc et al., 2014). The PAX7+ cells produced in these cultures exhibit characteristics of fetal SC. In mouse, these cells exhibit a higher regenerative capacity compared with adult SC, although their ability to colonize the adult SC niche is reduced (Tierney et al., 2016). Therefore, these human PAX7+ cells produced in vitro could constitute an ideal substrate for the development of cell therapies targeting muscular dystrophies or volumetric muscle injuries.

MATERIALS AND METHODS
sgRNA design and Cas9 vector assembly
Cas9 target sites were identified using the online CRISPR design tool (http://tools.genome-engineering.org). Briefly, DNA sequences flanking the start codon (human PAX7) or the stop codon (MYOG) were used for designing the sgRNAs. Several pairs of sgRNAs targeting either the Watson or the Crick strand of genomic DNA were selected and tested. For each target, a specific Cas9 vector was made. Briefly, the Cas9 vector pSPCas9 (BB)-2A-GFP (pX458; Addgene plasmid 48138) was digested using BbsI, and a pair of phosphorylated and annealed oligos (20 bp target sequences) were cloned into the gRNA of phosphorylated and annealed oligos (20 bp target sequences) were cloned into the gRNA of gRNA vectors were sequenced to ensure the presence of the right sequence. GFP (pX458; Addgene plasmid 48138) was digested using BbsI, and a pair of phosphorylated and annealed oligos (20 bp target sequences) were cloned into the gRNA vectors were sequenced to ensure the presence of the right sequence. gRNA vectors were sequenced to ensure the presence of the right sequence. gRNA vectors were sequenced to ensure the presence of the right sequence.

Targeting vector design and construction
MYOG
The 5′-homology arm of 1 kb containing exons 2 and 3 was amplified using the primers forward 5′-CGCGGATCCGGAAGCCAGGATCGCACG-GAAGGAGGCCGAGCTG-GAACGGGCTGTCGCTGC-CCATGTGGGACGATTTTT-3′ and reverse 5′-CTGAATGGTGTTCGTCGCTGCCTTTGCCATGGTTTTT-3′ and then inserted in the pBSKS-2A-NLS-Venus Neomycin vector by ligation-recombination. The vector was then digested with BsrGI and ClaI in order to remove the selection cassette. The 3′-homology arm of 0.6 kb was amplified with the primers forward 5′-GCCGTGTACAACTGATTATGGTTCGTCGCTGCCTTTGCCATGGTTTTT-3′ and reverse 5′-ACGCGTCGACCACAGGAGACCTTGAGATGGGAGCACC-3′ and reverse 5′-ACGCGTCGACCACAGGAGACCTTGAGATGGGAGCACC-3′ and then inserted in the construction with the 5′-homology arm after being digested with BsrGI and Sall.

PAX7
The vector was constructed in two steps. First, the 5′-homology arm of PAX7 (2 kb amplicon) was amplified by PCR using the primers forward 5′-GCCGAGATGCTCATGACGTCGCTGCCTTTGCCATGGTTTTT-3′ and reverse 5′-GAAAGCTTCTTGCCCTGTCCACATTCTCTTGCGAGCTG-3′ and reverse 5′-GAAAGCTTCTTGCCCTGTCCACATTCTCTTGCGAGCTG-3′. The PCR amplicon was digested with BglII and cloned by ligation/recombination into the pBSKS-2A-NLS-Venus Neomycin vector linearized with BamHI. The 3′-homology arm of PAX7 (1 kb amplicon) was amplified using the primers forward 5′-CCATCTGGGGGCTTCCCCTCCCCAACGCTGGAAT-3′ and reverse 5′-ACGCTATCGGTGCCCTCCCAAGGAGGAGTCCTCCTCCCCGATCGAGCTG-3′. The PCR amplicon was then digested with ClaI and Sall and ligated into the vector containing the 5′-homology arm after being digested with ClaI and Sall.

In a second step, a fragment containing exon 1 of PAX7 as well as the Neomycin selection cassette was excised using BsrGI/Sall restriction enzymes. Then by two successive PCRs, a 1.2 kb amplicon containing the 2A, PAX7 exon 1 (without the ATG) and the 3′-homology arm, were amplified using the 5′-AACGAGCGAGGGCAGTGGAGAGACAACC-AGGACACCGGCTTCCCTCCGGACGATCGAGA-3′, 5′-GAGAACGATCGATCTCCCAAGGAGGAGTCCTCCTCCCCGATCGAGCTG-3′, and reverse 5′-ACGCGTCGACCACAGGAGACCTTGAGATGGGAGCACC-3′ and reverse 5′-ACGCGTCGACCACAGGAGACCTTGAGATGGGAGCACC-3′. The PCR amplicon was then digested with BsrGI and Sall and ligated into the vector containing the 5′-homology arm and NLS-Venus, resulting in a construct of 2A-NLS-Venus in frame with exon 1 of PAX7 without the Neomycin selection cassette.

Generation of reporter human iPSC clones and validation
Generation
The NCRM1 human iPSC line (RUCDR, Rutgers University) was used for the generation of human reporter clones. Human iPSC were expanded on Matrigel in mTesR1 medium (Stemcell Technologies) for few days. Cells were dissociated using Accutase (Stemcell Technologies) and 8×10^5 cells were electroporated using 5 µg of total DNA (ratio 1:1) and the Amaxa Nucleofector kit as previously described (Ran et al., 2013). Then, 24 h after transfection, cells were dissociated and FACS-sorted for Cas9-GFP and expanded clonally at low densities. Later, clones were picked up for PCR screening and expansion.

Validation
PCR primers were designed outside the 5′- and the 3′-homology arms and in the insertion Cherry/Venus region to ensure proper site-specific targeting. All amplicons were sequenced to ensure in-frame positioning of the reporter/selection fragment and to rule out the presence of indels at the recombinase site.

iPSC maintenance and differentiation
Maintenance
Human iPSC were cultured as previously described (Chal et al., 2016, 2015). Briefly, cells were cultured on Matrigel-coated dishes (Corning, 354277) in mTesR1 medium. Cells were passaged as aggregates or as single cells. The NCRM1 human iPSC line and its engineered derivatives were tested and found to be mycoplasma-free.

Differentiation
Serum-free myogenic primary differentiation of human iPSC lines was performed as previously described (Chal et al., 2016, 2015). For FACS analysis and secondary differentiation purposes, 3-week-old primary myogenic cultures generated from hiPSC were dissociated as described and myogenic progenitors were replated at a density of 35-40 k/cm² onto Matrigel-coated (Corning, 354277) dishes in skeletal muscle growth medium (SKGM-2, Lonza, CC-3245) with 10 µM ROCK inhibitor. After 24 h, medium was changed to SKGM-2 medium without ROCK inhibitor. Cultures were allowed to proliferate for 1-2 days, at which point cultures reached ~90% confluence. Cultures were then induced for myogenic differentiation with DMEM/F12 supplemented with 2% knock-out serum replacement (Invitrogen, 10828028), 1 µM Chirion (Tocris Bioscience, 4423), 0.2% Pen/Strep (Life Technologies, 15140122) and 1× ITS (Life Technologies, 41400045). Following induction, differentiation medium was changed on days 1 and 2 and was then refreshed every other day for 1 week. To inhibit Notch signaling, 25 µM DAPT (Sigma-Aldrich, D5942) was added to KC medium for 1 week.

Flow cytometry analysis and FACS
Sorting and analysis
Cells were purified by FACS from differentiated cultures using an Aria (BD Biosciences) or an S3 cell sorter (Bio-Rad). Gating was determined for each reporter line using corresponding undifferentiated culture as a baseline control. Sorted populations were processed either for replating, transplantation experiments, microarray or RNA-seq experiments. Biological triplicates were generated.

Microarrays generation and analysis
Microarrays
Biotinylated cRNA targets were prepared from total RNA using a double amplification protocol according to the GeneChip Expression Analysis Technical Manual: Two-Cycle Target Labeling Assay (P/N 701021 Rev.5, Affymetrix). Following fragmentation, cRNAs were hybridized on GeneChip Human Gene 2.0ST Arrays. Each microarray was then washed and stained on a GeneChip fluidics station 450 and further scanned with a GeneChip Scanner 3000 7G. Finally, raw data (CEL Intensity files) were
extracted from the scanned images using the Affymetrix GeneChip Command Console (AGCC) version 1.4.1.46.

**Microarray data analysis**

Initial filtering and preprocessing, including background correction, quantile normalization and summarization, was performed using both RMA and MAS with the R BioConductor package (R version 2.12.1, Bioconductor version 2.8). Expression sets were then filtered according to Calls information. Probe sets expression fold changes between conditions (biological triplicates) were calculated using the ‘Comparative Marker Selection’ module of GenePattern (Reich et al., 2006). Volcano and FCsc scatter plots were generated using the Multiploit application of GenePattern. Histogram expression profiles of probe gene sets were generated from MAS values. Further analysis was performed using the Mantea database (Tassy and Pourquié, 2014) and DAVID Functional Annotation tools (Huang et al., 2009). Hierarchical clustering was performed on Microarray RMA data. Clustering was computed with an Average linkage method and Euclidean distances. A frequency histogram was used to determine the cell cycle state. For 10% of the cells, cell cycle phase could not be determined. Both approximately unbiased (AU) and bootstrap probability (BP) values. Further analysis was performed using the Manteia database (Tassy Hospital, Institute of Genetics and Molecular and Cellular Biology), in agreement with national and international guidelines.

**Immunohistochemistry**

Cell cultures were fixed for 20 min in 4% paraformaldehyde (PFA). Cultures were rinsed three times in PBS, followed by blocking buffer composed of Tris-buffered saline (TBS) supplemented with 10% FBS and 0.1% Triton X-100. Primary antibodies were diluted in blocking buffer and incubated overnight at 4°C. Cultures were then washed three times with TBST (TBS supplemented with 0.5% Tween-20) and incubated with secondary antibodies (1:500) and DAPI (5 μg/ml) in blocking buffer for 1 h at room temperature. Cultures were ultimately washed with TBST followed by PBS, before analysis. Dissected TA muscles were prepared for cryosections (10 μm) as previously described (Mathew et al., 2011). Slides were fixed in 4% PFA in PBS for 15 min at room temperature followed by 3 min of permeabilization with 0.5% Triton X-100. After fixation, slides were washed twice with PBS and blocked in 10% FBS-PBS-0.1% Triton X-100 for 30 min at room temperature. Slides were incubated with primary antibodies diluted in blocking buffer at 4°C overnight. They were then washed three times for 5 min in PBST (PBS and Tween 0.1%), and the corresponding secondary antibodies were applied for 2 h at room temperature. Sections were mounted with SouthernBiotec mounting medium (0100-01).

Antibodies used in this study are rabbit anti-myogenin (Santa Cruz Biotechnology, M-225, 1:200), mouse anti-myogenin (Developmental Studies Hybridoma Bank, F5D, 1:200), mouse anti-Pax7 (Developmental Studies Hybridoma Bank, Clone Pax7, 1:100), chicken anti-laminin (LSBio, LS-C96142, 1:500), chicken anti-GFP (Abcam, ab13970, 1:800), mouse anti-Fast MyHC (Sigma-Aldrich, MY-32, 1:500), mouse anti-human spectrin (Leica, NCL-SPEC1, 1:100), mouse anti-dystrophin (Leica, NCL-DYS1, 1:20), rabbit anti-human lamin A/C (Abcam, ab108595, 1:50) and rabbit anti-human lamin B2 (Cell Signaling Technology, 12255S, 1:50). The secondary antibodies used in this study are donkey anti-mouse IgG Alexa Fluor 488 conjugate (Thermo Fisher Scientific, R37114, 1:1000), goat anti-chicken IgY (H+L) Alexa Fluor 647 (Thermo Fisher Scientific, A-21449, 1:500), anti-mouse DyLight 549 (Jackson Immuno Research, 1:500) and anti-rabbit Alexa flour-488 diluted (Jackson Immuno Research, 1:500).

**Image acquisition and processing**

Live or fixed bright-field and fluorescent images were acquired on either a Zeiss Axiocamt, Evos FL or a Zeiss LSM780. Images were processed using Adobe Photoshop/Illustrator, and quantification and measurements were carried out using the Fiji software (Schindelin et al., 2012).

**DNA content analysis**

hiPSC-derived primary muscle cultures were dissociated using collagenase and trypsin as described above. PAX7<sup>Ymna+</sup> cells were isolated using flow cytometry. Propidium iodide staining was performed as previously described (Kim and Sederstrom, 2015). In brief, isolated PAX7<sup>Ymna+</sup> cells were fixed in ice-cold 70% ethanol followed by incubation at −20°C for 2 h. Cells were washed twice with PBS supplemented with 2% FBS and 1 mM EDTA. Cells were then stained in 50 μg/ml propidium iodide (ReadiDrop, Bio-Rad, 1351102) supplemented with 100 μg/ml RNase and 2 mM MgCl₂ for 20 min. Propidium iodide-stained cells were analyzed using flow cytometry (S3e™ Cell Sorter, Bio-Rad). Three independent experiments were performed and mean±s.d. was calculated.

**EdU staining**

Cultures of 19-day-old myogenic PAX7-Venus cells were incubated with 10 μM EdU in the culture medium for 40 h. On day 21 cells were dissociated using a mix of collagenase IV (Thermo Fisher Scientific, 17104019) and Trypsin-EDTA (Thermo Fisher Scientific, 25200-056) as previously described (Chal et al., 2015). Cell preparations were filtered through 70 μm and then 30 μm cell strainer and Venus<sup>+</sup> cells were FACs-sorted. Target cells were resuspended in phosphate-buffered saline (PBS) with tracking fluorescent beads (Molecular Probes). One day before transfection, 10 μl of Cardiotoxin (Sigma-Aldrich, 40 ng/ml) was injected in both TA muscles of 3- to 4-month-old Rag2<sup>−/−</sup> Rag1<sup>−/−</sup> males. Then 15 μl of cell preparation containing 10<sup>5</sup> cells was injected into TA muscles of 3- to 4-month-old mice. Injections were carried out under general anesthesia. Grafted TA muscles were collected 6-8 weeks after transplantation and processed for cryosection and immunofluorescence analyses. Experiments on mice were carried out according to local regulations (Boston Children’s Hospital, Brigham and Women’s Hospital, Institute of Genetics and Molecular and Cellular Biology), in agreement with national and international guidelines.

**Table 1. qRT-PCR human primers (5′-3′)**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAX7</td>
<td>ACCCTGCTTAAAACCACAC</td>
<td>GGGCCAAGAGATCTTGGACG</td>
</tr>
<tr>
<td>VENUS</td>
<td>ACGAATCCGACGAGCATC</td>
<td>ACGGAGAAAGCTGAGTACAC</td>
</tr>
<tr>
<td>GAPDH</td>
<td>AAGGTGAAAGTGCAGGATCACC</td>
<td>GGCTTCACTGATGCAATCCCTTA</td>
</tr>
<tr>
<td>MYOG</td>
<td>GCTGTATGAGACATCCCTCCTTA</td>
<td>GCATTCTTCCTACCACCTTAC</td>
</tr>
</tbody>
</table>

**Human fetal muscle cell isolation**

Human de-identified, discarded fetal tissue was collected under a protocol approved by the Committee of Clinical Investigation at Boston Children’s Hospital (IRB-P-00020286). Primary tissue was dissected into mononucleous cells, then cells were frozen and stored in a ultralow freezer at −140°C as previously described (Pakula et al., 2019). For purification of myogenic cells, frozen cells were thawed and plated overnight in DMEM-high glucose (4.5 g) media supplemented with 20% fetal bovine serum (FBS) and antibiotics. Cells were purified using FACS as described in Pakula et al. (2019) using APC anti-human CD56 antibody, Clone HCD56 (BioLegend, 318310) and PE anti-human CD92 antibody, Clone ASL-24 (BioLegend, 342103). Antibodies were added at a concentration of 5 μl antibody per million cells, as recommended by the manufacturer. Double-positive (CD56<sup>+</sup>/CD92<sup>+</sup>) sorted cells were plated overnight in DMEM high glucose media before injection in animals. For injections, sorted human fetal cells were trypsinized and resuspended in physiological-grade saline at a concentration of 100,000 cells/15 μl.

**qRT-PCR**

RNA was extracted from cells using Trizol (Invitrogen) or the NucleoSpin RNA XS RNeasy Mini Kit (Macherey-Nagel). RT-PCR was performed on total RNA using Quantifast SYBR Green RTPCR kit (Qiagen) and gene-specific primers (Primer-BLAST and Qiagen; Table 1) and run on a LightCycler 480II (Roche). Human GAPDH was used as a gene of reference and relative changes were calculated according the ΔΔCt method.

**Cell preparation and transplantation into TA muscles**

Cultures of 3-week-old differentiated PAX7<sup>Venus+</sup> hiPSC were dissociated using a mix of collagenase IV (Thermo Fisher Scientific, 17104019) and Trypsin-EDTA (Thermo Fisher Scientific, 25200-056) as previously described (Chal et al., 2015). Cell preparations were filtered through 70 μm and then 30 μm cell strainer and Venus<sup>+</sup> cells were FACs-sorted. Target cells were resuspended in phosphate-buffered saline (PBS) with tracking fluorescent beads (Molecular Probes). One day before transfection, 10 μl of Cardiotoxin (Sigma-Aldrich, 40 ng/ml) was injected in both TA muscles of 3- to 4-month-old Rag2<sup>−/−</sup> Rag1<sup>−/−</sup> males. Then 15 μl of cell preparation containing 10<sup>5</sup> cells was injected into TA muscles of 3- to 4-month-old mice. Injections were carried out under general anesthesia. Grafted TA muscles were collected 6-8 weeks after transplantation and processed for cryosection and immunofluorescence analyses. Experiments on mice were carried out according to local regulations (Boston Children’s Hospital, Brigham and Women’s Hospital, Institute of Genetics and Molecular and Cellular Biology), in agreement with national and international guidelines.

**HUMAN DEVELOPMENT**

Development (2020) 147, dev187344. doi:10.1242/dev.187344
and PAX7\textsuperscript{Venus\textsuperscript{+}} cells were sorted using flow cytometry (S3e\textsuperscript{TM} Cell Sorter, Bio-Rad). The sorted cells were plated in DMEM supplemented with Rock inhibitor (Tocris Bioscience, 1254) and after 30 min cells were fixed with 3.7% PFA (Electron Microscopy Sciences, 15710) for 15 min at room temperature. EdU staining was performed using the Click-it\textsuperscript{TM} EdU Imaging Kit (Invitrogen, C10340) following the manufacturer’s instructions. After EdU staining, cells were stained with PAX7 antibody (Abcam, ab34360, 1:200) and Hoechst 33342 (Thermo Fisher Scientific, H3570) using immunostaining as described above. The number of PAX7 and EdU double-positive cells was determined using Fiji (Schindelin et al., 2012). Three independent experiments were performed and mean±s.d. was calculated.

**scRNA-seq**

**Preparation of single cell suspensions for scRNA-seq**

The PAX7-Venus reporter line was differentiated according to our previously published method (Chal et al., 2016). On day 30, cells were dissociated using 2.5 mg/ml collagenase, (Type IV, Thermo Fisher Scientific, 17104019) and 0.05% Trypsin EDTA (Thermo Fisher Scientific, 25200-056) in PBS. The enzymes were neutralized by 5% FBS in DMEM (Thermo Fisher Scientific, 11965-118). Dissociated cells were passed through a 70 µm nylon cell strainer (Celltreat, 229483), spun down at 300 g for 5 min and resuspended in 0.1% FBS in PBS. PAX7-Venus\textsuperscript{+} cells were then passed through a 30 µm filter (Falcon, 352235) and sorted using flow cytometry (S3e\textsuperscript{TM} Cell Sorter, Bio-Rad). Sorted cells were spun down and resuspended in 0.1% BSA in PBS. Cell density was quantified manually and was adjusted to 150,000 cells/ml for encapsulation. The dissociation method and flow cytometry were optimized to achieve >90% viable cells and minimize cell doublets. Cells were collected from two independent experiments and 1100 cells were encapsulated from each experiment. Sequencing data from the replicates were combined for data analysis.

**Barcoding and sequencing**

PAX7-Venus\textsuperscript{+} single cells were encapsulated and barcoded using inDrops (Klein et al., 2015) as previously reported (Zilionis et al., 2017), using ‘V3’ sequencing adapters. The encapsulation and library preparation were performed by the Harvard Single Cell Core as described in Zilionis et al. (2017). inDrops libraries were sequenced on an Illumina NextSeq 500 using the NextSeq 75 High Output Kits. Standard Illumina sequencing primers were used: 61 cycles for Read1, 14 cycles for Read2, and eight cycles each for IndexRead1 and IndexRead2.

**Mapping, processing and clustering of single cell transcriptomes**

Raw sequencing data were processed using the inDrops.py bioinformatics pipeline available at github.com/indrops/indrops. Single cell transcriptome libraries were mapped to a human reference transcriptome built from the GRCh38.p12 (GCF_000001405.38) genome assembly. Bowtie version 1.2.2 was used with maximum permitted total of quality values of 200 at all mismatched read positions throughout the entire alignment (~200).

A weighted histogram of transcript counts per cell barcode versus cell barcode abundance was used to identify transcripts originating from abundant cell barcodes. Only transcript counts originating from abundant cell barcodes were included in downstream analysis. Basic filtering parameters were used to exclude cells expressing <250 genes and genes expressed in less than three cells. The filtered counts were normalized by the total number of counts for each biological sample. The top 2000 variable genes were identified according to Satija et al. (2015). For each cell, the fraction of counts due to mitochondrial genes was determined and cells with >0.2 fraction were filtered out. Any source of variation between the libraries was regressed out using Seurat\textsuperscript{’s} regressout function (Satija et al., 2015).

Single cell data were projected into a low dimensional space by principal component analysis (PCA). The nearest-neighbor graph (k=10) used Euclidean distance and 30 PCA dimensions. UMAP (Becht et al., 2018) and Force-directed graph (Jacomy et al., 2014) were used to embed the neighborhood graph. Cell clusters were identified using the Louvain graph-clustering method (community detection based on optimizing modularity) (Blondel et al., 2008). The top 500 differentially expressed genes were identified using a Wilcoxon rank-sum test by comparing cells of each cluster with cells of all the other clusters. The table of differentially expressed genes reports top-ranking differentially expressed genes, ranked by FDR-adjusted P-values (Table S4).

The cell cycle was scored as in Kowalczyk et al. (2015). Each cell was given a cell cycle score based on the expression of G2/M and S phase markers. The cells not expressing the markers from G2/M and S phase were identified to be in G0/G1 stage.

Coarse-grained layout of the data was generated using PAGA (threshold=0.05) (Wolf et al., 2019). Pseudotemporal orderings were constructed by selecting dividing PAX7\textsuperscript{+} cells as root. Diffusion pseudotime was calculated for all the remaining cells relative to the root. Cellular trajectories were assembled for paths through specified clusters, with cells ordered by DPT values (Wolf et al., 2019).

**Acknowledgements**

We are thankful to the Pourque lab, Christophe Marcelle and Stephane Vincent for their feedback and comments. We thank Amélie Freismuth, Marion Humbert and Betty Heller from the Institute of Genetics and Molecular and Cellular Biology (IGBMC) cell culture service for hiPSC culture assistance. We thank the IGBMC cell sorting facility, the microarray-sequencing platform and DNA Farber Molecular Biology Core Facilities. We thank Marie Klockaert, Tania Knauer-Meyer, Hannah Nelson, Lai Ding and Caroline Mursch for their help.

**Competing interests**

Olivier Pourque is a founder and shareholder of Anagenesis Biotechnologies.

**Author contributions**


**Funding**

This work was supported by an advanced grant from the European Research Council to O.P., by the European Union’s Seventh Framework Programme grant PluriMes (agreement no. 602423), by the Human Frontier Science Program (RGP0025/2016), by a strategic grant from the French Muscular Dystrophy Association (AFM) to O.P. and by a National Institutes of Health research grant (R01 AR074526) to O.P. and E.G. Deposited in PMC for release after 12 months.

**Data availability**

Microarray data have been deposited in GEO under accession number GSE149057. Single cell RNA-sequencing data have been deposited in GEO under accession number GSE149451.

**Supplementary information**

Supplementary information available online at https://dev.biologists.org/lookup/doi/10.1242/dev.187344.supplemental

**Peer review history**

The peer review history is available online at https://dev.biologists.org/lookup/doi/10.1242/dev.187344 reviewer-comments.pdf

**References**


muscle formation by the fusogenic micropeptide myomixer. Science 356, 323-327. doi:10.1126/science.aam361


