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

Ziad Al Tanoury1,2,3, Jyoti Rao2,3, Olivier Tassy1, Bénédicte Gobert1,6, Svetlana Gapon2, Jean-Marie Garnier1, Erica Wagner2, Aurore Hick6, Arielle Hall5, Emanuela Gussoni5 & Olivier Pourquié1,2,3,4*

1 Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS (UMR 7104), Inserm U964, Université de Strasbourg, Illkirch Graffenstaden, France.
2 Department of Pathology, Brigham and Women’s Hospital, 60 Fenwood road, Boston, Massachusetts, USA.
3 Department of Genetics, Harvard Medical School, 60 Fenwood Road, Boston, Massachusetts, USA.
4 Harvard Stem Cell Institute
5 Division of Genetics and Genomics, Boston Children’s Hospital, 3 Blackfan Circle, CLS 15021, Boston, Massachusetts, USA.
6 Anagenesis Biotechnologies, Parc d’innovation - BioParc 3, 850 Boulevard Sébastien Brandt, 67400 Illkirch Graffenstaden, France.

* Corresponding author: Olivier Pourquié
Department of Genetics, Harvard Medical School and Department of Pathology, Brigham and Women’s Hospital, 60 Fenwood road, Boston, Massachusetts, USA.
E-mail: pourquie@genetics.med.harvard.edu
SUMMARY
Satellite cells (SC) are muscle stem cells which can regenerate adult muscles upon injury. Most SC originate from PAX7-positive myogenic precursors set aside during development. While 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 (iPS) cell 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 iPS-derived PAX7-positive myogenic precursors. We show that the PAX7-positive 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 iPS cells, opening interesting avenues for muscular dystrophy cell therapy. This work provides significant insights into the development of the human myogenic lineage.
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 (SCs) (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 SCs become activated, proliferate and generate new fibers to reconstitute a healthy muscle. SCs 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.

SCs derive from the paraxial mesoderm, the embryonic tissue that forms the vertebral column and skeletal muscles (Chal and Pourquie, 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 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 Pourquie, 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 Pourquie, 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 Pourquie, 2017; Kassar-Duchossoy et al., 2005; Relaix et al., 2005). The Pax7+ cells become enclosed in the basal lamina of myofibers around day E16.5 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 SCs considerably slows down after birth and the SCs 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 (PSCs) such as embryonic stem (ES) cells or reprogrammed stem cells (induced pluripotent stem cells, iPS) (Chal et al., 2017; Magli and Perlingeiro, 2017). In mouse, epiblast-like cells can be efficiently induced
to a Presomitic Mesoderm fate characterized by \textit{Msgn1} and \textit{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 down-regulate Pax3 and activate Pax7 as observed in mouse myogenic precursors \textit{in vivo} (Chal et al., 2018; Chal et al., 2015; Hutcheson et al., 2009; Relaix et al., 2005). Long striated and multinucleated myofibers exhibiting characteristics of perinatal myofibers form after 3-4 weeks \textit{in vitro} (Chal et al., 2015). Such cultures recapitulate mouse embryonic and fetal skeletal myogenesis \textit{in vitro} (Chal et al., 2015). These cultures also generate Pax7\textsuperscript{+} myogenic precursors/fetal SCs (Chal et al., 2018; Chal et al., 2015). In cultures of a mouse Pax7\textsuperscript{GFP} ES reporter line, GFP expression appears around day 9 and peaks around 20\% at two to three weeks (Chal et al., 2018; Chal et al., 2015). After three weeks \textit{in vitro}, fast MyHC-positive (MyHC\textsuperscript{+}) fibers are interspersed with a population of small cells expressing Pax7. These cells are sometimes found enclosed in the basal lamina, resembling SC positioning \textit{in vivo} (Chal et al., 2015; Relaix et al., 2005; Sambasivan and Tajbakhsh, 2007). These dynamics of Pax7\textsuperscript{+} cells production \textit{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\textsuperscript{+} cells produced \textit{in vitro} can engraft in adult mouse muscles \textit{in vivo} and contribute Pax7\textsuperscript{+} satellite-like cells located under the basal lamina of myofibers (Chal et al., 2018; Chal et al., 2015). Thus, the mouse myogenic cultures spontaneously recreate a niche allowing the differentiation of Pax7\textsuperscript{+} cells exhibiting a similar regenerative potential to endogenous SCs.

SCs 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. Due to the difficulty to access 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 \textit{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 \textit{PAX7}, \textit{MYOG} and Myosin Heavy Chain (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, while 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 ES/iPS cells demonstrated that long striated myofibers and PAX7\textsuperscript{+} cells can also be produced \textit{in vitro} from PSCs (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 \textit{in vitro} using two human iPS reporter lines (PAX7\textit{Venus} and \textit{MYOG\textit{Venus}}), which identify subpopulations of the differentiating myogenic lineage. Using single cell RNA sequencing (scRNAseq) of FACS-sorted PAX7\textit{Venus\textsuperscript{+}} populations, we characterize the developmental stages of the PAX7\textsuperscript{+} precursors produced \textit{in vitro}. We also
demonstrate the myogenic potential of the human PAX7+ cells in vivo and in vitro. Our analysis demonstrate that these cells exhibit characteristics of the fetal satellite cell 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 iPS cells towards the satellite cell lineage, we targeted a Venus (a YFP derivative (Nagai et al., 2002)) reporter cassette into the exon1 of the human PAX7 gene in the human NCRM1 iPS 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 exon1. This allows for a fusion transcript but distinct protein products (Supplementary Figure 1a). We next differentiated the targeted cells according to an established myogenic differentiation protocol (Chal et al., 2016) (Figure 1a). The Venus nuclear signal first appeared after 2 weeks of differentiation in a subpopulation of mononuclear cells (Figure 1b). After 21 days of primary differentiation and after replating and secondary differentiation in vitro, the reporter cells generated long striated myofibers (>1mm) intermingled with PAX7Venus+ mononucleated cells (Figure 1a, c-d). Using anti-GFP and anti-PAX7 antibodies, we show that Venus expression largely overlaps with that of PAX7 (Figure 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 (Figure 1b,h). Moreover, when the Venus fraction was FACS-sorted from 21-day cultures, the PAX7 transcript was exclusively detected in the Venus+ fraction thus validating the reporter specificity (Figure 1i). The percentage of Venus+ cells decreased after 3 weeks, consistent with the dynamics of PAX7+ cells in vivo (Figure 1b). When 3-4 weeks cultures were dissociated and replated in proliferation medium (SKGM) for 1-2 days before transferring into differentiation medium for a week (secondary differentiation, Figure 1a), 20-25% of the mononucleated cell fraction was Venus+ (Figure 1b). These results contrast with the recently reported differentiation of a human PAX7/MYF5 ES 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 scRNAseq 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 iPS reporter cells. PAX7Venus+ cells upregulated several known markers of quiescent satellite cells 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; Pietrosemoli et al., 2017) (Table S1). Some genes coding for sarcomeric proteins such as Myosin Light Chain1 (MYL1) which are normally associated with later stages of myogenic differentiation were also upregulated,
suggesting that more differentiated cells could be included in the PAX7$^{\text{Venus}^+}$ population due to the stability of the Venus protein. We also compared the transcriptome of human PAX7$^{\text{Venus}^+}$ myogenic progenitors from 3-week old cultures to mouse Pax7$^{\text{GFP}^+}$ 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, Itga7 and Ncam1 (Table S1, S2). Thus, this analysis shows that the human PAX7$^{\text{Venus}^+}$ cells differentiated in vitro share a common transcriptional signature with the mouse ES-derived Pax7$^{\text{GFP}^+}$ myogenic precursor cells. Together our data indicate that expression of the PAX7$^{\text{Venus}^+}$ reporter allows tracking the endogenous expression of PAX7 protein in human myogenic precursors/fetal SCs differentiating in vitro.

To analyze the commitment of differentiating human iPS cells 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 exon3 of MYOG (Supplementary Figure 1b). When these reporter cells were differentiated in myogenic conditions (Chal et al., 2016), we detected Venus$^+$ myocytes as early as 10 days in culture (Figure 1b), i.e. prior to 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 MYOG$^{\text{Venus}^+}$ cells progressively fused to produce long myotubes expressing sarcomeric proteins such as alpha-actinin (ACTA2)(Figure 1j-k). Venus protein expression overlapped extensively with the MYOG protein (Figure 1l-n). After 3 weeks of differentiation in vitro, Venus$^+$ cells peaked to 12-15 % of the mononucleated cell fraction (Figure 1b, o). The MYOG transcript was only detected in the Venus$^+$ population in FACS-sorted cells, thus supporting the specificity of the reporter (Figure 1p). Thus at 3-4 weeks of differentiation, the proportion of myogenic cells (PAX7 and MYOG-positive) in the mononucleated fraction of the culture is around 30%. The remaining fraction appears to be mostly composed of fibroblastic populations.

We next compared the transcriptome of the MYOG$^{\text{Venus}^+}$ cells to that of undifferentiated iPS reporter cells. In contrast to the PAX7$^{\text{Venus}^+}$ cells, the FACS-sorted MYOG$^{\text{Venus}^+}$ 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 chain (MYH) isoforms MYL1, MYL4, MYH3 (embryonic), MYH8 (perinatal), troponin C type 2 (TNNC2) and titin (TTN) and many muscle-specific proteins such as muscle creatine kinase (CKM) (Table S3). Thus, the MYOG$^{\text{Venus}^+}$ reporter line allows tracking the differentiation of human myocytes in vitro.

We next performed a Gene Ontology (GO) analysis of the genes differentially expressed between the PAX7$^{\text{Venus}^+}$ and the MYOG$^{\text{Venus}^+}$ cells. This identified distinct over-represented pathways between the two populations. The MYOG$^{\text{Venus}^+}$ cells showed enrichment in GO categories such as Muscle protein, Muscle filament sliding, Z disc or muscle organ development while extracellular matrix organization was in the top GO categories for the PAX7$^{\text{Venus}^+}$ cells (Table S1, S3). Several genes coding for ECM proteins such as lumican (LUM), decorin (DEC), and COL3A1 were
among the most strongly up-regulated genes in PAX7\textsuperscript{Venus+} as compared to undifferentiated iPS cells (Table S1, S3). Tenascin-C (TNC), previously described as one of the most highly upregulated genes in mouse fetal SCs when compared to adult SCs (Tierney et al., 2016) was also highly upregulated in PAX7\textsuperscript{Venus+} cells. Together, these data show that the MYOG\textsuperscript{Venus+} cells represent a more differentiated population of myogenic cells compared to the PAX7\textsuperscript{Venus+} cells.

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

Two of the clusters are characterized by expression of PAX7 transcripts and lack of MYOG expression (Figure 2b). A first cluster (cluster 1) representing 18% of the PAX7\textsuperscript{Venus+} cells formed a distinct cluster of cycling cells which expressed genes involved in cell division such as MKI67 or PCNA (in green, Figure 2a-b, Supplementary Figure 3). A second cluster (cluster 2) contains the largest number of PAX7\textsuperscript{+} cells (60%) (in blue, Figure 2a-b, Supplementary Figure 2). We could not identify markers strictly specific for this cluster. The most specific markers such as MYF5 or CXCR4 were also shared with the cluster of cycling cells (Figure 2b, Supplementary Figure 3). Gene signatures specific for mouse and human satellite cells including VCAM1, MYF5, NFIA, ITGA7, and CAV1 were enriched in these two clusters (Barruet et al., 2020; Charville et al., 2015; Fukada et al., 2007; Liu et al., 2013) (Figure 2b, Supplementary Figure 3). These cells were also enriched in genes coding for extracellular matrix proteins such as COL6A1, COL3A1, TNC, DCN, GPC3 and for effectors of cell signaling (Supplementary Figure 2b-3). Members of the TGF beta pathway (such as TGFBI and TGFBR1/2) which has been involved in the control of myogenesis were upregulated in these two clusters (Biressi et al., 2007b; Hicks et al., 2018) (Supplementary Figure 3) along with targets of the FGF (SPRY1), and YAP pathway (CYR61) (Supplementary Figure 3) (Chakkalakal et al., 2012; Esteves de Lima et al., 2016; Flanagan-Steet et al., 2000).

A third cluster (cluster 3) includes cells expressing both PAX7 and MYOG (13% of total cells, in orange, Figure 2a, c). Genes enriched in this cluster include several genes associated to muscle differentiation and fusion such as the satellite cell marker SOX8 (Schmidt et al., 2003), the transcription factor HES6 (Gao et al., 2001), creatine kinase B (CKB) (Simionescu-Bankston et al., 2015), CDH15 (Marti et al., 2013) and the small fusogenic peptide myomixer (MYMX), which in mouse peaks in skeletal muscles at E14.5 (Bi et al., 2017) (Supplementary Figure 2 and 3). This cluster was located between the PAX7\textsuperscript{+}MYOG clusters and a fourth cluster (cluster 4) characterized by lack of PAX7 and expression of MYOG (8 % total cells in red, Figure 2a, c), suggesting that cluster 3 represents a transition state between PAX7\textsuperscript{+}MYOG- and PAX7\textsuperscript{-}MYOG\textsuperscript{+} cells. We also detected expression of the markers of human...
myogenic progenitors/fetal satellite cells *NGFR, ERBB3* and *CD82* in virtually all *PAX7*+ cells (Alexander et al., 2016; Hicks et al., 2018) (Figure 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, TNNC2, MYL1*, or *MYH8* indicating the differentiated status of these cells (Figure 2b, Supplementary Figure 3). 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 *PAX7*Venus+ 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 to muscle maturation and fusion described above to structural genes found in post-mitotic myocytes such as *MYH3* (embryonic MHC) and *MYH8* (Fetal MHC) (Figure 2b, Supplementary Figure 3).

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)(Figure 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; Zalc et al., 2014).

We next used scRNAseq data to analyze the cell cycle status of the *PAX7*Venus+ cells (Kowalczyk et al., 2015). Based on this analysis, around 68% of the cells were in G0/G1 state and the remaining cells were in G2/M and S phase (Figure 3a, b). We confirmed this data with an analysis of DNA content of FACS-sorted *PAX7*Venus+ cells by flow cytometry. This identified around 70% cells in the G0/G1 state (Figure 3 c, 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 *PAX7*Venus+ fraction, we labeled 21-day primary *PAX7*Venus- reporter cell cultures with EdU for 40 hours 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 (Figure 3e). Among the fraction of cells expressing the PAX7 protein, ~80% of the cells also expressed EdU indicating that they are actively proliferating (Figure 3e). We next performed cell cycle scoring and regressing out of the genes pertaining to S and G2/M phase from the cluster of *PAX7*MYOG- cyclic cells. This resulted in cells of cluster 1 to become mostly incorporated into cluster 2 (Figure 3f-g). Thus, the two *PAX7*MYOG- cell clusters represent cells of similar identity but in different phases of the cell cycle. Our data is consistent with most *PAX7*MYOG- cells representing a single population of slow cycling cells with a long G1 phase.

20% of the *PAX7*+ cells did not incorporate EdU arguing that they are either quiescent or post-mitotic. These cells likely include cluster 3 cells which express *MYOG* (a gene expressed mostly in post-mitotic cells) and account for ~13% of the
PAX7\textsuperscript{Venus+} cells. Cluster 3 (and cluster 4) also upregulate \textit{CDKN1C} (p57kip2), which controls exit from the cell cycle in myogenic cells in parallel with \textit{MYOG} (Zhang et al., 1999) (Figure 3h). This supports the conclusion that clusters 3 and 4 (representing 21% of the PAX7\textsuperscript{Venus+} 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 PAX7\textsuperscript{+} precursors activate \textit{MYOG} and exit G1 phase to become post-mitotic and down-regulate \textit{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 descendence and maintaining the quiescence of adult SCs (Brohl 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 iPS cells, with strong expression of the Notch receptors \textit{NOTCH2} and \textit{NOTCH3} (and to a lesser extent \textit{NOTCH1}) and their targets \textit{HEY1} and \textit{HES1} (Brohl et al., 2012; Mourikis and Tajbakhsh, 2014) in the PAX7/\textit{MYOG} clusters. The ligand Dll1 is expressed by the differentiated \textit{MYOG} positive cells (Figure 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 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 PAX7+/\textit{MYOG-} clusters 1 and 2 (Figure 3e). To directly test the effect of Notch signaling on the differentiation of human PAX7 precursors \textit{in vitro}, we treated replated cultures of the PAX7 reporter line in differentiation medium with the drug DAPT which inhibits Notch cleavage. When compared to control, we could not detect any PAX7\textsuperscript{Venus+} cells in the treated cultures (Figure 4b-g). These cultures presented large alpha-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 satellite cell pool in vitro, raises concerns about the use of DAPT when differentiating myogenic cells from ES/iPS in culture (Choi et al., 2016; Selvaraj et al., 2019).

To determine their myogenic potential, we FACS-sorted 3-week old human PAX7\textsuperscript{Venus+} cells generated \textit{in vitro} followed by replating in SKGM medium for 1-2 days to allow cells to reach 80-90% confluency (Figure 1a, 5a). Cells were then transferred to differentiation medium where 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 (Figure 5b-c) and showed spontaneous twitching indicating a mature contractile apparatus (data not shown). These striated myofibers were interspersed with mononucleated cells expressing PAX7-Venus (Figure 5c), suggesting that the PAX7\textsuperscript{Venus+} cells are able to differentiate into myofibers and to self-renew \textit{in vitro}. Similar observations were made with
mouse Pax7<sup>GFP</sup>+ 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 (Figure 5d-f). Serial sorting and differentiation of human PAX7<sup>Venus+</sup> 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 PAX7<sup>Venus+</sup> cells (Figure 5g).

To determine the myogenic potential of the human PAX7<sup>Venus+</sup> cells <em>in vivo</em>, we transplanted 10<sup>5</sup> FACS-sorted PAX7<sup>Venus+</sup> cells from 3 weeks-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 24h before transplantation to induce myofiber injury but no X-irradiation of the TA muscles was performed before transplantation. Thus, endogenous mouse satellite cells were also able to contribute to myofiber repair <em>in vivo</em> and potentially compete with the exogenous human PAX7<sup>Venus+</sup> cells. Mice were sacrificed after 6-8 weeks. Human-derived myofibers and mono-nucleated 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 (Figure 5h). Six weeks after transplantation, clusters of human-derived fibers expressing dystrophin (DMD) and lamin B2 or spectrin and lamin A/C (Figure 5i) were identified on sections along most of the length of the transplanted TA in the NOD; Rag1<sup>−/−</sup>; Dmd<sup>mdx-5Cv</sup>. DMD-positive muscle fibers were of varying diameter and often contained both peripheral and central human nuclei. DMD expression most likely derives from the grafted human cells as this strain of <em>mdx</em> 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 antibodies combinations (Alexander et al., 2016; Rozkalne et al., 2014). Up to 45 human fibers were identified in sections of PAX7<sup>Venus+</sup> injected TAs (34.81 ± 11.52). A comparable number of engrafted fibers were identified when CD82+/CD56+ human fetal myogenic cells were injected (49.85 ± 19.30), indicating that PAX7<sup>Venus+</sup> cells show a comparable regenerative potential to these cells (Figure 5j). These ratios of human fibers/injected cells are similar to those obtained when human iPS differentiated to a myogenic fate are sorted based on ERBB3 or NGFR expression prior to injection <em>in vivo</em> (Hicks et al., 2018). Importantly, some human cells (human lamin A/C +) located underneath the myofibers basal lamina were also observed to express PAX7, indicating that human PAX7<sup>Venus+</sup> transplanted cells can contribute to the satellite cell pool (arrows, Figure 5k-o).

Therefore, the human PAX7<sup>Venus+</sup> cells generated <em>in vitro</em> exhibit similar regenerative properties to mouse and human myogenic precursors/fetal satellite cells (Chal et al., 2015; Tierney et al., 2016).
CONCLUSION
Here, we dissected the progression of human myogenesis in vitro using reporter iPS 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 called satellite cells (SCs) (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 (Figure 1q) (Chal et al., 2016). Here, we observed that MYOG is first expressed in the myogenic cultures around day 10-12, prior to 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 (Figure 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 myogenesis during which myofibers form from PAX7+ mononucleated cells which are also able to self-renew (Hutcheson et al., 2009).

We also demonstrate using scRNAseq, that our differentiation conditions can yield a pure myogenic human PAX7+ cells 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 satellite cells in mouse (Alexander et al., 2016; Barruet et al., 2020; Biressi et al., 2007b; Tierney et al., 2016).

Analysis of the scRNAseq data allowed us to characterize the developmental trajectory of the human PAX7 myogenic lineage in vitro. We identify 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 iPS cells 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 Marcelle, 2013; Relaix et al., 2005; Zalc et al., 2014). The PAX7+ cells produced in these cultures exhibit characteristics of fetal SCs. In mouse, these cells exhibit a higher regenerative capacity compared to adult SCs 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.
EXPERIMENTAL PROCEDURES

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 ID: 48138) was digested using BbsI and a pair of phosphorylated and annealed oligos (20 bp target sequences) were cloned into the guide RNA locus as described (Ran et al., 2013). All vectors were sequenced to ensure the presence of the right sequence. gRNA pairs were tested for their efficiency using the SURVEYOR assay as described (Ran et al., 2013).

Targeting vector design and construction

**MYOG-Venus.** The 5’ homology arm of 1kb containing exons 2 and 3 was amplified using the primers fwd 5’-CGCGGATCCGGAAGCCAGGAAGGCCTGGAAGCAGGGCCT-3’ and rev 5’-CTGAAGTTGGTTGCTCGCTGCCCATGTGGGGATGGTTT-3’ then inserted in the pBSKS-2A-NLS-Venus Neomycin vector by ligation-recombination. The vector was then digested with BsrGI and Clal in order to remove the selection cassette. The 3’ homology arm of 0.6kb was amplified with the primers fwd 5’-GCGGTGTACAAGTAAATCGATGATTGTCTTCCAAGCGGAGCTTCTGCGG-3’ and rev 5’-ACGCCTGACACAGGAGACCTTGGGATGGCAGAG-3’ then inserted in the construction with 5’ homology arm after being digested with BsrG1 and SalI.

**PAX7-Venus.** The vector was constructed in 2 steps: In a first step, the 5’-homology arm of PAX7 (2kb amplicon) was amplified by PCR using the primers fwd 5’-GCCAGATCTGTCAATGCTGCCAAGGGGCAAAGGTCACAG-3’ and rev 5’-GAACAGCTCTCCTGGCTCTACCATTCTGCGACGCGCGGGAGCAATCCTAAA-3’. The PCR amplicon was digested with BglIII and cloned by ligation/recombination into the pBSKS-2A-NLS-venus Neomycin vector linearized with BamHI. The 3’-homology arm of PAX7 (1kb amplicon) was amplified by using the primers fwd 5’-CCCATCGATGCGGCCCTTCCCGGCACGGTACCAGAATCCAAA-3’, rev 5’-ACGCATCGATGTCGACCGAAGGGAGCTTCCGGATCGGGCT-3’. The PCR amplicon was then digested with the Clal and Sall and ligated into the vector containing the 5’ homology arm after being digested with Clal and Sall. In a second step, a fragment containing the exon1 of PAX7 as well as the Neomycin selection cassette was excised by using BsrG1/Sall restriction enzymes. Then by 2 successive PCRs, a 1.2kb amplicon containing the 2A, PAX7 exon1 (without the ATG) and the 3’ homology arm, were amplified using the primers 5’-AAGCAGGCCAGCGTGAGGAGAAACCAACTTCAGCTTGGCAGGACGCGC-3’ and 5’-ACGCATCGATGTCGACCGAAGGGAGCTTCCGGGATCGGCT-3’ primers. The
amplicon was digested with BsrGI and Sall and ligated into the vector containing the 5'- homology arm and NLS-venus, thus resulting in a construction 2A-NLS-venus in frame with exon1 of PAX7 without the neo selection cassette.

**Generation of reporter human iPS clones and validation**

*Generation.* The NCRM1 human iPS line (RUCDR, Rutgers University) was used for generation of human reporter clones. Human iPS cells were expanded on Matrigel in mTesR1 medium for few days. Cells were dissociated using Accutase (Stemcell Technologies) and 8 x 10^5 cells were electroporated using 5 μg of total DNA (Ratio 1:1) and the Amaxa Nucleofector kit as described (Ran et al., 2013). 24h post transfection, cells were dissociated and FACS-sorted for Cas9-GFP^+ve 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 arm 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 recombination site.

**iPS cell maintenance and differentiation**

*Maintenance.* Human iPS cells were cultured as described previously (Chal et al., 2016; Chal et al., 2015). Briefly, cells were cultured on Matrigel (BD Biosciences)-coated dishes in mTesR1 media (Stem Cell Technologies). Cells were passaged as aggregates or as single cells. The NCRM1 human iPS line (RUCDR, Rutgers University) and its engineered derivatives were tested mycoplasma-free.

*Differentiation.* Serum-free myogenic primary differentiation of human iPS lines was performed as described previously (Chal et al., 2016; Chal et al., 2015). For FACS analysis and secondary differentiation purposes, 3-week old primary myogenic cultures generated from hiPSCs were dissociated as described and myogenic progenitors were replated at a density of 35-40k/cm^2 onto Matrigel (Corning, Cat#354277)-coated dishes in skeletal muscle growth media (SKGM-2, Lonza CC-3245) with 10 μM ROCK inhibitor. After 24 hours, medium was changed to SKGM-2 media 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% knockout serum replacement (Invitrogen, Cat. # 10828028), 1 μM Chiron (Tocris, Cat. # 4423), 0.2% Pen/Strep (Life Technologies, Cat. # 15140122) and 1x ITS (Life Technologies, Cat. # 41400045). Following induction, differentiation medium was changed on day 1 and 2 and then was refreshed every other day for 1 week. To inhibit Notch signaling, 25 μM DAPT (Sigma Aldrich cat. No. 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 a 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 sequencing 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, Santa Clara, USA). 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 FcFc scatter plots were generated using the Multiplot application of GenePattern. Histogram expression profiles of gene probe sets were generated from MAS values. Further analysis was performed using the Mantel database (Tassy and Pourquie, 2014) and DAVID Functional Annotation tools (Huang da et al., 2009). Hierarchical clustering was performed on Microarray RMA data. Clustering was computed with an Average linkage method and Euclidean distances. Both an Approximately unbiased (AU) and Bootstrap probability (BP) P values were calculated (pvclust, R package). GEO accession code for the microarray data is GSE149057.

**Cell preparation and transplantation into Tibialis Anterior (TA) muscles**

Cultures of 3-week old differentiated PAX7\textsuperscript{Venus} hiPS cells were dissociated using a mix of collagenase IV (Thermo Fisher Scientific #17104019) and Trypsin-EDTA (Thermo Fisher Scientific #25200-056) as described (Chal et al., 2015). Cell preparations were filtered through 70 μm and then 30 μm Cell Strainers and Venus positive cells were FAC5-sorted. Target cells were resuspended in phosphate-buffered saline (PBS) with tracking fluorescent beads (Molecular probes). One day prior to transplantation, 10μl of Cardiotoxin (Sigma, 40 ng/mL ) was injected in both *Tibialis Anterior* (TA) muscles of 3- to 4-month-old mice. Injections were done under general anesthesia. Grafted TA muscles were collected 6-8 weeks after transplantation and processed for cryosection and immunofluorescence analyses. Experiments on mice were done according to local regulations (Boston Children’s Hospital, Brigham and Women’s Hospital, IGBMC), in agreement with national and international guidelines.
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-P00020286). Primary tissue was dissociated into mononuclear cells, then cells were frozen and stored in an 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 (4.5g) media supplemented with 20% FBS and antibiotics. Cells were purified using the FACS as described in (Pakula et al., 2019) using APC anti-human CD56 antibody, Clone HCD56 (BioLegend, catalog number:318310) and PE anti-human CD82 antibody, Clone ASL-24 (BioLegend, catalog number: 342103). Antibodies were added at a concentration of 5µl antibody per million cells, as recommended by the manufacturer. Double positive (CD56+CD82+) 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) and run on a LightCycler 480II (Roche). Human GAPDH was used as a gene of reference and relative changes were calculated according the Δ∆Ct method.

Immunohistochemistry

Cell cultures were fixed for 20 mins in 4% paraformaldehyde (PFA). Cultures were rinsed three times in PBS, followed by blocking buffer composed of Tris-buffered saline (TBS) supplemented with 10% Fetal Bovine Serum (FBS) and 0.1% Triton X-100. Primary antibodies were then 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 1h at Room Temperature. Cultures were ultimately washed with TBST followed by PBS, before analysis. Dissected TA muscles were prepared for cryosections (10 µm) as 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% fetal bovine serum (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 3 times 5 min in PBS-T (Tween 0.1%), and the corresponding secondary antibodies were applied for 2 h at room temperature. Sections were mounted with SouthernBiotech mounting medium (Cat# 0100-01).

Antibodies used in this study are rabbit anti-Myogenin (Santa Cruz, M-225), mouse anti-Myogenin (DSHB, F5D), mouse anti-Pax7 (DSHB, Clone Pax7), chicken anti-Laminin (LS bio, catalog LS-C96142), Rabbit anti-Laminin (Sigma, L9393), rabbit anti-Laminin (Abcam, ab11575), chicken anti-GFP (Abcam, ab13970), mouse anti-Fast
MyHC (Sigma, MY-32), mouse anti-human spectrin (Leica, NCL-SPEC1), mouse anti-dystrophin (Leica, NCL-DYS1), rabbit anti-human lamin A/C (Abcam, ab108595) and rabbit anti-human lamin B2 (Cell Signaling, Cat#122555).

**Image acquisition and processing**

Live or fixed brightfield and fluorescent images were acquired on either a Zeiss Axiovert, Evos FL, or a Zeiss LSM780. Images were processed with Adobe Photoshop/Illustrator, and quantification and measurements were done with 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\textsuperscript{Venus+} cells were isolated using Flow cytometry. Propidium iodide staining was performed as described (Kim and Sederstrom, 2015). In brief, isolated PAX7\textsuperscript{Venus+} cells were fixed in ice cold 70% ethanol followed by incubation at -20°C for two hours. 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 minutes. Propidium iodide stained cells were analyzed using Flow cytometry (S3e™ Cell Sorter, Bio-Rad). Three independent experiments were performed and Mean±SD was calculated.

**EdU staining**

19-day old myogenic cultures of PAX7-Venus cells were incubated with 10μM EdU in the culture medium for 40 hours. On day 21 cells were dissociated and PAX7\textsuperscript{Venus+} cells were sorted using flow cytometry (S3e™ Cell Sorter, Bio-Rad). The sorted cells were plated in DMEM supplemented with Rock inhibitor (Tocris, 1254) and after 30 minutes, cells were fixed with 3.7% Paraformaldehyde (Electron Microscopy Sciences, 15710) for 15 minutes at room temperature. EdU staining was performed using the Click-it® EdU Imaging Kit (Invitrogen, C10340) following the manufacturer’s instructions. After Edu staining, cells were stained with PAX7 antibody (Abcam, ab34360) and Hoechst 33342 (ThermoFisher Scientific, H3570) using immunostaining as described above. The number of PAX7 and EdU double positive cells was determined using Fiji (Schindelin, J. et al. 2012). Three independent experiments were performed and Mean±SD was calculated.

**Single cell RNA sequencing**

*Preparation of single-cell suspensions for single cell RNA-sequencing:* PAX7-Venus reporter line was differentiated according to our published method (Chal et al., 2016). On day 30, cells were dissociated using with 2.5mg/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.
Dissociated cells were passed through a 70 µm nylon Cell Strainer (Celltreat, 229483), spun down at 300g for 5 min and resuspended in 0.1% FBS in PBS. PAX7-Venus positive cells were then passed through a 30 µm filter (Falcon, 352235) and sorted using flow cytometry (S3e™ 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. 1100 cells were encapsulated from each experiment. Sequecing data from the replicates were combined for data analysis.

**Barcoding and sequencing:** PAX7-Venus positive 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, 8 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 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 (\(\text{-e } 200\)).

A weighted histogram of transcript counts per cell barcode vs cell barcode abundance was used to identify transcripts originating from abundant cell barcodes. Only transcripts 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 3 cells. The filtered counts were normalized by total number of counts for each biological sample. Top 2000 variable genes were identified according to Satija (Satija et al., 2015). For each cell, fraction of counts due to mitochondrial genes was determined and cells with >0.2 fraction were filtered out. Source of variation between the libraries were regressed out using Seurat’s regressout function (Satija et al., 2015). 1427 cells passed the filtering criteria and were further analyzed.

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 Louvain graph-clustering method (community detection based on optimizing modularity) (Blondel et al., 2008). 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. Differentially expressed genes table report top-ranking differentially expressed genes, ranked by FDR adjusted p-values.

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-positive 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 Pourquie’s laboratory, Christophe Marcelle and Stephane Vincent for their feedback and comments. We thank Amélie Freismuth, Marion Humbert and Betty Heller from the IGBMC cell culture service for hiPSC culture assistance. We thank the IGBMC cell sorting facility, the microarray-sequencing platform and DANA Farber Molecular Biology Core Facilities. We like to thank Marie Knockaert, Tania Knauer-Meyer, Hannah Nelson, Lai Ding and Caroline Mursch for their help. This work was supported by an advanced grant from the European Research Council to O.P., by the FP7 EU grant Plurimes (agreement no. 602423), by HFSP (award no. RGP0052/2018), by a strategic grant from the French Muscular Dystrophy Association (AFM) to O.P. and by an NIH R01 AR074526 research grant to O.P and E.G.

AUTHOR CONTRIBUTIONS

Z.A.T. designed, performed and analyzed the biological experiments with O.P. Z.A.T. generated the CRISPR/Cas9-reporter cell lines and analyzed data. J.R. performed and analyzed single cell RNA Sequencing experiments, DNA content and EdU analysis. O.T. helped with microarray analysis. B.G carried out transplantation experiments with Z.A.T. J.M.G helped in generating targeting constructions. S.G. helped with the in vitro differentiations. E.W. carried out PCR experiments with Z.A.T and EdU staining with J.R. E.G. performed transplantation in Nod/Rag1null mdx3cv mice. Z.A.T. and O.P. performed the final data analysis and wrote the manuscript. O.P. supervised the project. All authors discussed and agreed on the results and commented on the manuscript.
REFERENCES


stem cells to muscle fiber to model Duchenne muscular dystrophy. Nat Biotechnol 33, 962-969.


Figure 1: Generation of human iPS reporter cell lines

(a) Schematics of the differentiation protocol highlighting key factors and time scale of primary and secondary (following dissociation and replating) differentiation of myogenic progenitors from iPSCs, as previously described (Chal et al., 2016). C: Chir; L: LDN193189; F: FGF-2, H: HGF; I: IGF; K: KSR; d: day in culture. hiPS: undifferentiated human iPS cells.
(b) FACS-analysis showing the percentage of PAX7Venus and MYOGVenus positive cells in the mononucleated fraction during primary differentiation and after 7 days of secondary differentiation (SD7).
(c-g) PAX7Venus reporter cells at SD7 stained with antibodies against GFP and α-actinin (c-d) or with anti-GFP and anti-PAX7 antibodies (e-g). Scale bar, 100μm.
(h) Flow cytometry analysis of PAX7Venus cultures. PAX7Venus reporter cells were differentiated as described above for primary differentiation. Cultures were dissociated after 3 weeks and FACS sorted.
(i) RT-qPCR analysis for Venus and PAX7 in undifferentiated human iPSCs (PSC) and PAX7Venus+ FACS-sorted cells from 3-week old primary cultures.
(j-n) MYOGVenus reporter cultures at SD7 stained with antibodies against GFP and α-actinin (j-k) or with anti-GFP and anti-MYOG antibodies (l-n). Nuclei were counterstained with DAPI. Scale bar, 100μm.
(o) Flow cytometry analysis of MYOGVenus cultures. MYOGVenus reporter cells were differentiated for 2 weeks in primary differentiation, dissociated and FACS sorted.
(p) RT-qPCR analysis for Venus and MYOG in undifferentiated human iPS and FACS-sorted MYOGVenus+ cells isolated from 3-week old primary cultures.
(q) Schematic recapitulating the differentiation timeline of human iPS cells differentiating to skeletal muscle in vitro. Green bars represent expression windows of the fluorescent reporter lines. NMP (neuro-mesodermal precursors), PSM (Presomitic Mesoderm).
Figure 2: Single cell analysis of PAX7\textsuperscript{Venus+} cells

(a) ForceAtlas2 force-directed layout (k=10, 30 PC dimensions, 1427 cells) of single cell transcriptomes from 30-day old FACS-sorted PAX7\textsuperscript{Venus+} cells from primary cultures. Colors indicate Louvain cluster IDs.

(b) Genes indicative of cluster 2 (PAX7, MYF5), cluster 1 (MKI67), cluster 3 (MYOG, MYMX) and cluster 4 (MYH3), as well as markers of human myogenic precursors (MYOD1, CD82, NGFR, ERBB3) shown in ForceAtlas2 layouts colored by log-normalized transcript counts.

(c) Pseudo-temporal ordering of PAX7\textsuperscript{Venus+} cells along a path towards a differentiated state. Top, colors indicate Louvain IDs. Bottom, heatmap of selected markers showing gene changes along PAGA path (root = cluster 1).

(d) Coarse-grained layout of PAX7\textsuperscript{Venus+} cells using Partition-based graph abstraction (PAGA).
Figure 3: Cell cycle analysis of PAX7\textsuperscript{Venus\textsuperscript{+}} cells

(a, b) Cell cycle analysis of the PAX7\textsuperscript{Venus\textsuperscript{+}} fraction by scRNAseq. PCA plot showing distribution of cell cycle states in PAX7\textsuperscript{Venus\textsuperscript{+}} cells identified based on scRNAseq (a). Quantification of cells in different cell cycle states (1427 cells, Mean±SD, n= 2) (b).

(c, d) Cell cycle analysis of the PAX7\textsuperscript{Venus\textsuperscript{+}} fraction by DNA content analysis. Propidium iodide flow cytometry analysis of cell cycle in PAX7\textsuperscript{Venus\textsuperscript{+}} cells. Representative frequency histogram reflecting DNA content (c). Quantification of cell cycle state in PAX7\textsuperscript{Venus\textsuperscript{+}} cells (Mean±SD, n= 3) (d)

(e) Percent of PAX7\textsuperscript{Venus\textsuperscript{+}} cells expressing the PAX7 protein (left). Quantification of PAX7 protein expressing cells labeled by EdU after 40 hours (right) (Mean±SD, n= 3).
(f) ForceAtlas2 layout and Louvain clustering after performing the regression of cell cycle genes. Colors indicate Louvain cluster IDs.

(g) After cell cycle regression cycling cells collapse with cluster 1 (MYF5). Cell cycle indicator MKI67 and other genes shown in ForceAtlas2 layouts colored by log-normalized transcript counts.

(h) Force Atlas2 layouts colored by log-normalized transcript counts of CDKN1C (p57kip2).
Figure 4: Notch signaling is required for the maintenance of the PAX7\textsuperscript{Venus\textsuperscript{+}} cells \textit{in vitro}

(a) Distribution and expression of the Notch receptors (\textit{NOTCH1}, \textit{NOTCH2} and \textit{NOTCH3}), the ligand \textit{DLL1}, and the targets (\textit{HEY1} and \textit{HES1}) in the PAX7\textsuperscript{+}/MYOG\textsuperscript{-} clusters.

(b-g) Effect of Notch signaling on the differentiation of human PAX7 precursors \textit{in vitro}. PAX7\textsuperscript{Venus\textsuperscript{+}} cells were FACS-sorted after 3 weeks of primary differentiation and replated as described above and differentiated for 1 week in KC differentiation medium in absence or presence of the \textit{γ}-secretase inhibitor DAPT (25μM) that blocks Notch signaling. Cells were fixed and stained with α-actinin and GFP antibodies. Scale bar: b-e, 1000 μm. f-g, 100 μm.
Figure 5: Myogenic potential of iPS-derived PAX7-Venus

(a-f) Myogenic potential of purified human (a-c) PAX7\textsuperscript{Venus\textsuperscript{+}} and mouse (d-f) Pax7\textsuperscript{GFP\textsuperscript{+}} cells \textit{in vitro}. Human iPS and mouse ES cells were differentiated for 3 weeks as previously described (Chal et al., 2016) and fluorescent cells were FACS-isolated and replated in SKGM medium for 1-2 days (a, d) to allow 80-90\% confluency. Cells were then induced for differentiation and immunostained for GFP and FAST MyHC (b-c, e-f). Nuclei were counterstained with DAPI. Scale bar, 200\(\mu\)m (a,d,c,f) and 1000\(\mu\)m (b,e).

(g) Serial sorting and differentiation of human PAX7\textsuperscript{Venus\textsuperscript{+}} cells. PAX7\textsuperscript{Venus\textsuperscript{+}} cells were differentiated as described for primary differentiation, dissociated at 3 weeks and FACS sorted (1\textsuperscript{st} round). FACS-sorted cells were then replated in SKGM medium for 1-2 days to
allow 80-90% confluency and then differentiated for 1 week in a differentiation medium containing 1µM CHIR and 2% KSR. Cultures were then dissociated and FACS-sorted for Venus expression (2nd Round). Replating and sorting experiments were repeated as indicated. The percentage of PAX7-Venus+ cells is shown at each round (n=3).

(h-o) Transplantation of human PAX7-Venus+ cells in immune-deficient mice. 10^5 FACS-sorted PAX7-Venus+ were injected into the TA of cardiotoxin-injured Rag2B6 WT (h) or NOD-Rag1^-/-Dmd^-/-Scv mice (i-j). In vivo contribution of the transplanted PAX7-Venus+ human cells to muscle fibers is visualized on transverse sections of the grafted Tibialis Anterior muscle. Sections were stained with a combination of human-specific anti-spectrin (green)/laminA-C (green) (h), which respectively detect human fibers and human nuclei but not mouse myofibers, and with anti-dystrophin (red) and laminB2 (green) antibodies that co-detects dystrophin -positive myofibers and human-derived nuclei, respectively (i). Transverse section of Tibialis Anterior muscle grafted with FACS-sorted CD82+/CD56+ primary fetal human myogenic precursor cells and stained with anti-dystrophin and anti-human lamin A-C (j). Scale bar, 100µm.

(k-o) Higher magnification of grafted TA section showing a human PAX7-Venus+ cell (white arrowhead) expressing the PAX7 protein (l) and human specific Lamin A/C (m) localized under the laminin-positive basal lamina (n). Laminin antibody in these images is a chicken anti laminin. Nuclei are labeled with DAPI (o). Red arrowhead shows a human laminA/C positive PAX7 negative myofiber nucleus. Scale bar, 50µm.
### 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>ACCCCTGCCTAACCACATC</td>
<td>GCGGCAAAAGAATCTTGGAGAC</td>
</tr>
<tr>
<td>VENUS</td>
<td>GGGCACAAGCTGGAGTACAAC</td>
<td>ACGAACTCCAGCAGGACCAT</td>
</tr>
<tr>
<td>GAPDH</td>
<td>AAGGTGAAGGTCGGAGTCAAC</td>
<td>GGGGTATTGATGGCAACATA</td>
</tr>
<tr>
<td>MYOG</td>
<td>GCTGTATGAGACATCCCCCTA</td>
<td>CGACTTCCTTTACACCTTAC</td>
</tr>
</tbody>
</table>
SUPPLEMENTARY FIGURES

**Figure S1: Targeting strategy for iPS reporters**

Targeting strategies for hPAX7<sup>Venus</sup> (a), and hMYOG<sup>Venus</sup> (b) with CRISPR/CAS9. HA, homology arm; nls, nuclear localization signal; HA, homology arm; H2B for nuclear localization.
Figure S2: Single cell analysis of PAX7-Venus cells

(a) ForceAtlas2 directed layout showing overlap between PAX7-Venus positive cells analyzed from two myogenic cultures.

(b) Top 20 enriched transcripts for the four Louvain clusters relative to all other clusters detected by Wilcoxon rank-sum test. Transcripts are ranked by FDR-corrected p-values (Benjamini-Hochberg). See also Table Differentially expressed genes.

(c) Top 20 enriched transcripts for the three Louvain clusters identified after regression of cell cycle genes relative to all other clusters detected by Wilcoxon rank-sum test. Transcripts are ranked by FDR-corrected p-values (Benjamini-Hochberg). See also Table Differentially expressed genes.
Figure S3: Single cell analysis of PAX7-Venus cells

ForceAtlas2 directed layout showing Louvain clusters and representative transcripts for each cluster. Scale represents log-normalized transcript counts of indicated genes.
Table S1

Click here to Download Table S1

Table S2

Click here to Download Table S2

Table S3

Click here to Download Table S3

Table S4

Click here to Download Table S4