Characterization of hiPSC-Derived Muscle Progenitors Reveals Distinctive Markers for Myogenic Cell Purification Toward Cell Therapy

Minas Nalbandian,¹ Mingming Zhao,¹,‡ Mitsuru Sasaki-Honda,¹ Tatsuya Jonouchi,¹ Antonio Lucena-Cacace,² Takuma Mizusawa,³ Masahiko Yasuda,³ Yoshinori Yoshida,² Akitsu Hotta,¹ and Hidetoshi Sakurai¹,‡

¹Department of Clinical Application, Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Shogoin-Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan
²Department of Cell Growth and Differentiation, Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Shogoin-Kawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan
³Central Institute for Experimental Animals, 3-25-12 Tonomachi, Kawasaki-ku, Kawasaki 210-0821, Japan

¹Correspondence: zhaoming@cira.kyoto-u.ac.jp (M.Z.), hsakurai@cira.kyoto-u.ac.jp (H.S.)
https://doi.org/10.1016/j.stemcr.2021.03.004

SUMMARY

The transplantation of muscle progenitor cells (MuPCs) differentiated from human induced pluripotent stem cells (hiPSCs) is a promising approach for treating skeletal muscle diseases such as Duchenne muscular dystrophy (DMD). However, proper purification of the MuPCs before transplantation is essential for clinical application. Here, by using MYF5 hiPSC reporter lines, we identified two markers for myogenic cell purification: CDH13, which purified most of the myogenic cells, and FGFR4, which purified a subset of MuPCs. Cells purified with each of the markers showed high efficiency for regeneration after transplantation and contributed to the restoration of dystrophin expression in DMD-immunodeficient model mice. Moreover, we found that MYF5 regulates CDH13 expression by binding to the promoter regions. These findings suggest that FGFR4 and CDH13 are strong candidates for the purification of hiPSC-derived MuPCs for therapeutic application.

INTRODUCTION

During skeletal muscle regeneration, a local resident stem cell population called satellite cells (SCs) (Mauro, 1961) plays a fundamental role in the maintenance and repair of the tissue (Relax and Zammit, 2012). The transplantation of healthy SCs has been studied as regenerative medicine for skeletal muscle disease (Cerletti et al., 2008; Marg et al., 2014; Montarras et al., 2005; Sacco et al., 2008; Tanaka et al., 2009; Xu et al., 2015). However, isolating human SCs from a human biopsy is challenging and invasive. In addition, in vitro expansion of SCs causes the loss of their regeneration potential, resulting in senescence (Gilbert et al., 2010; Montarras et al., 2005; Negroni et al., 2009). On the other hand, human induced pluripotent stem cells (hiPSCs) can be expanded and then differentiated to various cell lineages (Takahashi et al., 2007), making them an attractive cell source for generating muscle progenitor cells (MuPCs).

Indeed, several groups (Chal et al., 2015; Darabi et al., 2012; Sato et al., 2019; Shelton et al., 2014; Wu et al., 2018; Xi et al., 2017; Zhao et al., 2020) have developed in vitro differentiation systems to obtain MuPCs from hiPSCs (hiPSC-MuPCs). However, many of these methods artificially overexpress myogenic factors to induce the skeletal muscle differentiation (Darabi et al., 2012; Sato et al., 2019), which currently limits their clinical use because of potential tumorigenicity due to random integration. To avoid this risk, transgene-free differentiation protocols have also been developed (Chal et al., 2015; Shelton et al., 2014; Wu et al., 2018; Xi et al., 2017; Zhao et al., 2020).

Although hiPSC-MuPCs show myogenic potential and regenerative capacity after transplantation, there is large heterogeneity in the cell populations, including non-myogenic cells such as neuronal and mesenchymal cells (Xi et al., 2020), indicating that not all the hiPSC-derived cells are of myogenic lineage. The use of surface markers to purify cells presents an elegant solution to this problem. Because the developmental stage of hiPSC-MuPCs is known to be different from that of adult skeletal muscle tissues (Incitti et al., 2019; Xi et al., 2020; Zhao et al., 2020), previously known surface markers for human SCs may not be suitable for hiPSC-MuPC purification. Therefore, further investigation is required to identify bona fide cell-surface markers.

Recently, we reported a new myogenic induction protocol with prolonged Wnt agonist treatment to differentiate hiPSC-MuPCs (Zhao et al., 2020). This protocol showed improved dermomyotome differentiation efficiency compared with other protocols. In our previous study, we generated a MYF5-tdTomato reporter iPSC line and found that even the MYF5-tdTomato-negative cell population expresses PAX7, a widely accepted marker for quiescent SCs. This finding prompted us to compare the MYF5-positive and PAX7-positive cell populations to isolate hiPSC-MuPCs.

In the present study, we investigate the transcriptional differences between PAX7- and MYF5-expressing...
Please cite this article in press as: Nalbandian et al., Characterization of hiPSC-Derived Muscle Progenitors Reveals Distinctive Markers for Myogenic Cell Purification Toward Cell Therapy, Stem Cell Reports (2021), https://doi.org/10.1016/j.stemcr.2021.03.004
hiPSC-MuPC populations. Furthermore, to address the heterogeneity problem, we identified two specific and novel cell-surface markers for the hiPSC-MuPC purification, CDH13 (T-CADHERIN or H-CADHERIN) and FGFR4, which we confirmed are expressed in human fetal skeletal MuPCs. Cell populations positive for either marker showed a high regeneration capacity in vivo, including the restoration of dystrophin in Duchenne muscular dystrophy (DMD)-immunodeficient model mice. These markers are expected to benefit studies aiming to generate in vitro MuPCs and therapeutic applications.

RESULTS

**MYF5-tdTomato and PAX7-Venus Reporters Efficiently Identify hiPSC-MuPCs**

To generate hiPSC-MuPCs, we used our recently reported stepwise myogenic induction differentiation protocol (Zhao et al., 2020). Because the properties of the hiPSC-MuPCs depend on the culture time, in the present study, we decided to study hiPSC-MuPCs at day 42 (early stage) and day 84 (late stage) of differentiation (Figure 1A).

To identify and purify hiPSC-MuPCs, we established PAX7-Venus and MYF5-tdTomato reporter cell lines in two different hiPSC lines: 201B7 and S01 (Figures 1B, 1C, and S1A–S1D). After myogenic induction, we observed several twitching colonies (Video S1), in which many Venus-positive and tdTomato-positive cells were located (Videos S2 and S3). We were also able to observe and quantify the tdTomato-positive and Venus-positive cells by flow cytometry (Figure 1D). At the early stage, ~25% of mononuclear cells were PAX7-Venus positive and ~5% were MYF5-tdTomato positive, while at the late stage, it was ~35% and ~7% (Figure 1E).

qRT-PCR experiments revealed that PAX7 and MYF5 were respectively highly expressed in PAX7-Venus-positive and MYF5-tdTomato-positive cells compared with their negative counterparts (Figures 1F and 1G), indicating correlation with the reporter systems. Early- and late-stage MYF5-tdTomato-positive cells were also enriched for PAX7, but only late-stage PAX7-Venus-positive cells were enriched for MYF5. MYF5 and PAX7 staining (Figure 1H) of freshly sorted cells revealed that most PAX7-Venus-positive cells were positive for PAX7 and the majority of MYF5-tdTomato-positive cells were positive for MYF5. Furthermore, MYF5 and PAX7 double-positive cells were found in ~65% of MYF5-tdTomato-positive late-stage cells, but just ~15% of the other populations (Figure 1I). These data suggest that the majority of late-stage MYF5-tdTomato-positive cells were a sub-population of PAX7-positive cells.

All cell populations showed robust myogenic differentiation capacity in vitro (Figures 1E and 1F). We confirmed that more than 50% of the nuclei located outside of the myotubes were PAX7-venus positive (Figures S1G and S1H), suggesting that some of these cells can proliferate as a reserve myogenic cell population (Yoshida et al., 1998) and will not express the major histocompatibility complex (MHC). To study the regeneration capacity, we performed transplantation experiments in NOG-mdx mice (n = 4–6 per group) (Figure S1I). The early-stage cell engraftment capacity was very low compared with late-stage cells (Figures S1J and S1K), indicating that late-stage cells possess higher regeneration capacity. Moreover, the MYF5-tdTomato-positive cells showed the highest engraftment capacity.

**RNA Sequencing Reveals Maturation-Related Differences between hiPSC-MuPC Populations**

To transcriptomically compare different hiPSC-MuPC populations, we performed RNA sequencing (RNA-seq) of sorted early- and late-stage MYF5-tdTomato-positive and PAX7-Venus-positive cells. Principal component analysis

Figure 1. Myogenic Induction of hiPSCs and Validation of the PAX7-Venus and MYF5-tdTomato Reporter Cell Lines

(A) Schematic representation of the myogenic induction and the strategy for purifying hiPSC-MuPCs. PAX7-Venus-positive and MYF5-tdTomato-positive cell lines were sorted at the early (day 42) and late (day 84) stages.

(B) CRISPR-Cas9 knockin strategy to establish the PAX7-Venus reporter cell lines.

(C) CRISPR-Cas9 knockin strategy to establish the MYF5-tdTomato reporter cell lines.

(D) Representative fluorescence of the reporter cell lines before dissociation. Left: PAX7-Venus staining at the early and late stages. Right: images of live early- and late-stage MYF5-tdTomato cells in culture. Scale bars, 200 μm.

(E) Quantification by FACS of PAX7-Venus-positive and MYF5-tdTomato-positive cells at the early and late stages. Error bars represent SEM.

***p < 0.001.

(F and G) Relative mRNA expressions of PAX7 and MYF5 in PAX7-Venus (F) and MYF5-tdTomato (G) positive and negative cells at the early and late stages normalized to GAPDH. Data from at least three independent experiments are shown as the mean ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001, and p****<0.0001.

(H and I) Immunohistochemical analysis of PAX7-Venus-positive and MYF5-tdTomato-positive cells. (H) Freshly sorted cells were seeded for 30 min on laminin 511-coated dishes and fixed with 4% paraformaldehyde. Fixed cells were stained with PAX7 and MYF5 antibodies. Scale bar, 50 μm. (I) PAX7 and MYF5 stained cells were quantified. Data shown are from three independent experiments. All data are from the 201B7 cell line. E, early stage; L, late stage.
Please cite this article in press as: Nalbandian et al., Characterization of hiPSC-Derived Muscle Progenitors Reveals Distinctive Markers for Myogenic Cell Purification Toward Cell Therapy, Stem Cell Reports (2021), https://doi.org/10.1016/j.stemcr.2021.03.004

(legend on next page)
revealed that hiPSC-MuPCs are transcriptomically very far from undifferentiated hiPSCs (Figure 2A). Undifferentiated hiPSCs migrated along a trajectory defined by PC1 to early-stage cells, while early-stage cells used a trajectory defined by PC2 to become late-stage cells. In addition, the sample correlation matrix indicated that late-stage cells correlated less with undifferentiated cells than did early-stage cells (Figure 2B). Consistently, a cluster dendrogram revealed that hierarchically late-stage cells grouped separate from early-stage cells (Figure 2C). These results confirmed the culture-time-dependent differences in hiPSC-MuPCs in vitro.

As shown in Figures 2D and 2E, we identified hundreds of DEGs (differentially expressed genes: ≥ 2-fold change and p < 0.05) between samples. The majority of DEGs between early and late stages consisted of upregulated genes in the late stage (Figure 2D, left), indicating that late-stage cells activated several stage-specific genes. We found that 422 genes were commonly upregulated in late over early stage cells (Figure 2F). GO (gene ontology) analysis revealed that late-stage cells were enriched for biological process terms related to muscle development (Figure 2G), suggesting a time-dependent skeletal muscle lineage specification.

To determine the developmental stage of the different hiPSC-MuPC populations, we plotted the gene signatures of markers related to different developmental stages of human MuPCs (personal communication by Dr. Xi, UCLA) (Figure 2H). Late-stage cells were enriched for fetal and post-natal (juvenile and adult) human markers. We confirmed these results by qRT-PCR (Figure 2I). Extended signatures (Xi et al., 2020) for genes upregulated in MuPCs during development and for the transcriptional program of each developmental stage of human skeletal MuPCs confirmed that late-stage cells were enriched for maturation markers as well as for fetal and post-natal transcription factors (Figures S2A and S2B). These data suggest that late-stage cells possess a more mature phenotype and transcriptomically resemble the fetal/post-natal developmental stage.

To understand the transcriptomic differences between PAX7-positive and MYF5-positive cells, we studied DEGs between MYF5-tdTomato-positive cells and PAX7-Venus-positive cells in both stages. Three hundred seventy-four DEGs were upregulated in the two MYF5-tdTomato-positive populations (Figure S2C), while 205 DEGs were upregulated in the two PAX7-Venus-positive populations (Figure 2J). GO analysis revealed that MYF5-tdTomato-positive cells were enriched for terms related to development and cell adhesion (Figure S2D). On the other hand, PAX7-Venus-positive cells were enriched for neuron development-related terms (Figures 2K and 2L). We further confirmed gene enrichment for neuron markers in PAX7-positive cells (Figures S2E and S2F). These results provide evidence that neuronal-lineage cells expressing PAX7 are derived from our hiPSC-MuPC differentiation protocol, suggesting that in our culture system, MYF5 possesses a higher efficiency than PAX7 to specify MuPCs.

RNA-Seq Reveals that CDH13 and FGFR4 Surface Markers Enrich Myogenic Capacity and Regenerative Potential In Vivo
To uncover new surface markers that could uniquely distinguish myogenic cell populations, we compared the RNA-seq data of MYF5-tdTomato-positive and -negative cells (DDBJ: DRA010291) differentiated from hiPSCs (Zhao et al., 2020) (Figure 3A). We identified a total of 993 DEGs (≥ 2-fold change and p < 0.05) between populations, of which 219 were upregulated in MYF5-tdTomato-positive and 774 in MYF5-tdTomato-negative cells (Figures 3B and 3C). Moreover, GO analysis of the DEGs upregulated in MYF5-tdTomato-positive cells indicated that the most distinguished GO term was skeletal muscle satellite cell differentiation (Figure 3D). Furthermore, from the DEGs upregulated in MYF5-tdTomato-positive cells, we could identify...
several cell-surface markers (Table S1). To screen these surface markers, we decided to contrast them with the Human Skeletal Muscle Atlas database (aprilpylelab.com), which is a single-cell RNA-seq database that provides gene expression during human development, as well as with hiPSC-MuPCs differentiated by one of three different protocols: MS (Shelton et al., 2014), JC (Chal et al., 2015) and HX (Xi et al., 2017). We identified $CDH13$ and $FGFR4$, which were mostly expressed in skeletal MuPCs across development, with a markedly higher expression during the embryonic and fetal stages (Figures S3A and S3B). To determine whether $CDH13$ and $FGFR4$ expression was specific...
to human MuPCs or common in other animals, we referred to Tabula Muris, a mouse single-cell RNA-seq database (https://tabula-muris.ds.czbiohub.org/; Schaum et al., 2018), which revealed that Cdh13 and Fgfr4 are robustly expressed in mouse MuPCs, too (Figures S3C and S3D), but that Cdh13 is also expressed in endothelial cells (Philipova et al., 2009).

To better understand the expression patterns of these genes in human tissue and whether they are predominantly expressed by skeletal muscle, we used an mRNA-seq database of human tissues surrounding skeletal muscle (GEO: GSE106292). We found that CDH13 and FGFR4 were highly expressed in skeletal muscle compared with surrounding tissues (Figures S3E and S3F). After verifying the expression of these markers in skeletal muscle by different databases, we decided to test them for the purification of hiPSC-MuPCs.

First, we confirmed by qRT-PCR CDH13 and FGFR4 enrichment in MYF5-tdTomato-positive cells (Figures 3E and 3F). Moreover, fluorescence-activated cell sorting (FACS) analysis revealed that ~12% of mononuclear cells were FGFR4 positive and ~23% were CD13 positive. Moreover, ~90% of FGFR4-positive cells were also CD13 positive, suggesting that FGFR4-positive cells are a subset of CD13-positive cells (Figures 3G and S4A). Then we studied the capacity of CD13 and FGFR4 to enrich the myogenic population of hiPSC-MuPCs. For this purpose, we used three different hiPSC lines: 201B7 (healthy donor), DMD (DMD patient-derived cells), and DMD CRISPR-Cas9-corrected isogenic control (DMD-corrected) (Figure 4A) (Li et al., 2015). CDH13-positive cells were enriched for MYF5, PAX7, and MYOD1, while FGFR4-positive cells were enriched for MYF5 and PAX7, but not MYOD1 (Figures 4B, 4C, and S4B).

To evaluate the myogenic capacity of CD13-positive and FGFR4-positive cells, we studied the differentiation capacity in vitro. CD13-positive and FGFR4-positive cells had higher in vitro differentiation capacity than their negative counterparts (Figures 4D, 4G, S4C, and S4D). Moreover, we could detect DYSTROPHIN protein in DMD-corrected cells, indicating that this purification method may be useful for in vitro disease modeling.

To determine the therapeutic potential of cells purified with CDH13 and FGFR4, we studied their regeneration capacity in vivo. Immunohistochemical analysis revealed a superior engraftment capacity in FGFR4-positive and CD13-positive cells over the negative sorted fractions (n = 3–4 per group). Moreover, FGFR4-positive cells accounted for a higher number of h-SPECTRIN-positive cells (Figures 5A, 5B, and S4E). In addition, in muscle transplanted with cells positive for either of the markers, in the interstitial zone we found hundreds of h-SPECTRIN-positive tiny fibers (i.e., the area of the fiber was not visible as larger than the nuclear area). Because these fibers seem to be de novo fibers that had not fused yet with host myofibers, we decided not to include these tiny fibers in the calculation of the engrafted cells (Figures S4F and S4G). Furthermore, DMD-corrected cells restored dystrophin expression in the engrafted fibers (Figure 5C). Interestingly, with either of the markers, we found that some of the transplanted cells were PAX7 positive and positioned beneath the basal lamina and outside of the DYSTROPHIN-positive sarcolemma, which histologically is consistent with SCs (Figures 5D and S4H). Moreover, double staining for PAX7 and MYOD1 indicated that many of the FGFR4-positive transplanted cells express myogenic markers (Figures 5E and 5F). These results suggest that the transplanted cells may also contribute to the muscle stem cell pool and that CDH13 and FGFR4 can be used to efficiently purify myogenic cells with therapeutic potential.

CDH13 and FGFR4 Improve hiPSC-MuPC Purification Compared with Previously Reported Markers

Other surface proteins have been proposed for hiPSC-MuPC purification. To compare CDH13 and FGFR4 with those previously reported markers, we used the Human Skeletal Muscle Atlas to evaluate the single-cell transcriptomics of hiPSC-MuPCs differentiated from the MS, JC, and HX protocols (Figures S5A–S5C). We compared the expression patterns of FGFR4 and CDH13 with NCAM1 (Garcia et al., 2018); CD82 (Alexander et al., 2016; Uezumi et al., 2016); NGFR and ERBB3 (Hicks et al., 2018); ICAM1, ITGA9, and SDC2 (Magli et al., 2017); and MME (Wu et al., 2018) across the different protocols. ICAM1, ITGA9, and MME were not expressed by myogenic cells, and SD2, NCAM1, and NGFR were mostly non-specific for any cell type. On the other hand, CD82 was robustly expressed by myogenic cells but also marked non-myogenic cells; ERBB3 was robustly expressed by myogenic cells (mostly by MYOD1- and MYOGENIN-expressing cells) but enriched as well for other cell types, including Schwann cells; CD13 (which was less expressed than the other markers) marked myogenic cells but was expressed as well by other non-myogenic cells; and finally, FGFR4 was expressed almost exclusively in skeletal MuPCs, especially those expressing PAX7 and/or MYF5. This last finding suggests that as a marker, FGFR4 may give a higher efficiency for purifying hiPSC-MuPCs than the other markers.

Furthermore, the gene expression of these markers was evaluated in MYF5-tdTomato-positive and MYF5-tdTomato-negative cells. We found that CDH13, FGFR4, NGFR, ERBB3, and CD82 were enriched in MYF5-positive cells, while SDC2, NCAM1, ITGA9, ICAM1, and MME were enriched in MYF5-tdTomato-negative cells (Figure S5D).
To extrapolate these results to our hiPSC myogenic induction protocol, we tested the markers that were enriched in MYF5-tdTomato-positive cells for hiPSC-MuPC purification in our DMD and DMD-corrected cell lines. We found that, consistent with the single-cell RNA-seq data, NCAM1 and NGFR marked over 50% of the cells, while CD82, ERBB3, CDH13, and FGFR4 marked between 15% and 30% (Figures 6A and 6B, left). To characterize the sorted cells, we first studied the gene expression of MYF5, PAX7, and MYOD1. We found that all the markers enriched these three genes. Importantly, CDH13-positive cells and FGFR4-positive cells were the most enriched for MYF5 and PAX7 (Figures 6A and 6B). On the other hand, again consistent with single-cell RNA-seq data, FGFR4-positive cells were less enriched for MYOD1.
To compare the myogenic capacity between sorted cell populations, we performed in vitro differentiation (Figures S6A and S6B). All sorted cell populations could give rise to myotubes, but CDH13-positive and FGFR4-positive cells showed the highest differentiation capacity. By using a Pax7-Venus hiPSC reporter line, we also analyzed the percentage of Pax7-positive cells among the different sorted populations (Figures 6C and 6D). FACS analysis revealed...
that CD82 and NGFR enriched for ~40% of PAX7-Venus-positive cells, ERBB3 for ~60%, CDH13 for ~70%, and FGFR4 for ~95%.

To study possible contamination by non-myogenic cell lineages, we examined the gene expression of neuronal progenitor cell markers. Because the Human Skeletal
Please cite this article in press as: Nalbandian et al., Characterization of hiPSC-Derived Muscle Progenitors Reveals Distinctive Markers for Myogenic Cell Purification Toward Cell Therapy, Stem Cell Reports (2021), https://doi.org/10.1016/j.stemcr.2021.03.004
ChIP-RT-qPCR was performed for HA antibody on the predicted MYF5 binding sites in the CDH13 promoter region, with both sequences containing the E-box binding motif (Tierney et al., 2016). To confirm if MYF5 bound to these sites, we performed chromatin immunoprecipitation (ChIP) for MYF5 followed by qRT-PCR. qRT-PCR analysis revealed that DNA fragments precipitated by the hemagglutinin (HA) antibody were enriched for the predicted binding site sequences (Figure 7G), confirming that MYF5 promotes CDH13 expression by directly binding to the CDH13 promoter region at two different E-box sequences.

Previously, MYF5 and MYOD1 ChIP-sequencing experiments were reported for mouse SCs (GEO: GSE75370) (Conerly et al., 2016). By using the available database, we found that the MYF5-binding signal was strong in the promoter region of Cdh13, suggesting that the regulation of Cdh13 by MYF5 is conserved between mice and humans (Figure 7H). In addition, we analyzed previously published ChIP-sequencing data for chromatin histone modifications in mouse quiescent SCs (GEO: GSE103163) (Machado et al., 2017). We found an increased signal in H3K27ac and H3K4me3 next to the MYF5-binding binding signal, suggesting again a regulatory site for MYF5 (Figure 7H, bottom). We contrasted these results with other single-cell RNA-seq data of activated and quiescent mouse SCs (De Micheli et al., 2020a). In that study, single-cell RNA-seq was performed before the injury and at 2, 5, and 7 days after the injury. By analyzing the available data (GEO: GSE143437), we found that when activated (i.e., after injury), the number of Cdh13-positive and Fgfr4-positive SCs was decreased. Moreover, 7 days after the injury, the gene expression returned to the level before the injury (Figure 7I). These data suggest that quiescent SCs in mouse express Cdh13 and Fgfr4, which is consistent with our results for hiPSC-MuPCs, where CDH13-positive and FGFR4-positive cells were enriched for Pax7 and MYF5.

**CDH13 Transcription Is Directly Regulated by MYF5**

Because CDH13 and FGFR4 were found to be enriched in MYF5-tdTomato-positive hiPSC-MuPCs, we hypothesized that MYF5 might control the expression of these genes. Indeed, we found a strong correlation between MYF5 and CDH13 expression and between MYF5 and FGFR4 expression in hiPSC-MuPCs (Figure 7A). To elucidate the possible regulatory mechanism of MYF5 on CDH13 and FGFR4 expression, we investigated whether MYF5 overexpression may induce CDH13 and/or FGFR4 expression. We transfected hiPSCs (201B7) with a doxycycline-inducible mCherry-HA-MYF5 expression vector (Figure 7B) and observed that most of the cells were mCherry positive, indicating the efficacy of the transfection (Figure 7C). Moreover, we confirmed by qRT-PCR that MYF5 was efficiently overexpressed in doxycycline-treated cells (Figure 7D). Further, CDH13 expression was increased in MYF5-overexpressing undifferentiated hiPSCs, but FGFR4 expression was unchanged (Figure 7E), suggesting that CDH13 gene expression, but not FGFR4, is regulated by MYF5. Previously, Fgfr4 has been shown to be downstream of PAX3 in mouse embryonic myogenic progenitor cells (Lagha et al., 2008) and in hiPSC-MuPCs (Magli et al., 2019), which could explain why FGFR4 is expressed in hiPSC-MuPCs.

To verify the MYF5 control over CDH13 expression, we analyzed possible MYF5 binding motifs in the promoter region of CDH13. We found putative binding sites in the CDH13 promoter region, with both sequences containing the E-box binding motif (Tierney et al., 2016). To confirm if MYF5 bound to these sites, we performed chromatin immunoprecipitation (ChIP) for MYF5 followed by qRT-PCR. qRT-PCR analysis revealed that DNA fragments precipitated by the hemagglutinin (HA) antibody were enriched for the predicted binding site sequences (Figure 7G), confirming that MYF5 promotes CDH13 expression by directly binding to the CDH13 promoter region at two different E-box sequences.

Figure 7. MYF5 Regulates CDH13 Gene Expression

(A) Plots showing correlations between CDH13 and MYF5 relative expression (left) and between FGFR4 and MYF5 relative expression (right). Cells used were MYF5-tdTomato (201B7) positive (red dots) or negative (black dots). The gray areas represent 95% confidence bounds.
(B) The PB-MYF5-HA vector construct.
(C and D) MYF5 overexpression in hiPSCs (201B7). (C) Representative bright-field (PHASE) and mCherry images of live cultures of hiPSCs transfected with the PB-MYF5-HA vector and treated for 3 days with doxycycline (+dox) or a negative control without doxycycline (−dox). Scale bar, 200 μm. (D) Relative mRNA expression of MYF5 in transfected cells.
(E) Relative mRNA expression of CDH13 and FGFR4 in MYF5-overexpressing cells.
(F) Representative scheme of the predicted MYF5 binding sites in the CDH13 gene promoter region.
(G) ChIP-qPCR was performed for HA antibody on the predicted MYF5 binding sites in the CDH13 promoter region. The percentage of input was normalized to an IgG control.
(H) Integrative Genomics Viewer snapshot displaying the ChIP-seq tracks for MYF5, MYOD1, H3K4me3, and H3K27ac in mouse satellite cells (Cdh13 region). The black arrow represents the transcription start site and green boxes represent the intergenic region.
(I) Analysis of Cdh13 and Fgfr4 gene expression based on single-cell RNA-seq for the regeneration process of mouse TA. Data shown in (A), (D), and (E) represent gene expression normalized to GAPDH. Data shown in (D), (E), and (G) are from at least three independent experiments and shown as means ± SEM. *p < 0.05, **p < 0.01, and ***p < 0.001. Gene expression data are normalized to GAPDH.
and after transplantation gave rise to PAX7-positive cells located under the basal lamina-like SCs.

**DISCUSSION**

By using MYF5 and PAX7 hiPSC reporter lines, we identified two new markers for myogenic cell purification: CDH13 and FGFR4. Both markers could sort cells that have high *in vivo* regeneration capacity and *in vitro* differentiation potential. In addition, we showed that CDH13 expression is regulated by MYF5.

MYF5 is known to be the first of the four myogenic regulatory factors (MYF5, MYOD1, MYOGENIN, and MRF4) expressed during mammalian skeletal muscle development (Ott et al., 1991). Notably, one study suggested that MYF5-positive fetal MuPCs are the primary source of adult SCs, but MYF5-positive embryonic MuPCs made a much smaller contribution (Blessi et al., 2013). This finding is consistent with late-stage MYF5-tdTomato-positive cells having higher regenerative and therapeutic potential. On the other hand, PAX7, which has been suggested to promote the progression of somatic progenitors into the myogenic lineage and is recognized as a marker of SCs (Seale et al., 2000), is also a non-skeletal muscle-specific gene and known to be expressed in the developing central nervous system and neural crest lineage cells (Buckingham and Relaix, 2007). Consistently, we found that PAX7-Venus-positive cells were also enriched for genes related to neuron development, suggesting that in our culture system, PAX7-Venus-positive cells might include neuronal lineage cells. Similarly, single-cell RNA-seq analysis of PAX7-positive cells prepared from different myogenic induction protocols indicated the contamination of neuronal lineage cells in the sorted cell populations (Xi et al., 2020).

Cell heterogeneity and non-myogenic lineage have been described in several myogenic induction protocols (Xi et al., 2020). Moreover, the transplantation of unsorted hiPSC-MuPCs resulted in extremely poor regeneration capacity (Hicks et al., 2018). Previously reported markers for adult SCs, like NCAM1 and CD82 (Figures 6 and S5) or ITGA7, MCAM, C-MET, CXCR4, and CD11b (Hicks et al., 2018), have been shown to be unsuitable for purifying myogenic cells, probably because the developmental stage of hiPSC-MuPCs is rather fetal and different from that of adult skeletal muscle tissues (Incitti et al., 2019; Xi et al., 2020; Zhao et al., 2020).

We are not the only group to have reported surface markers specific for hiPSC-MuPCs (Hicks et al., 2018; Magli et al., 2017; Wu et al., 2018). Among them, the most used are ERBB3 and NGFR (Hicks et al., 2018; Kim et al., 2020; Sakai-Takemura et al., 2018), which were identified by analyzing human fetal tissues. We studied these markers in the Human Skeletal Muscle Atlas and found that they are expressed not only by skeletal MuPCs but also by Schwann cells across several myogenic induction methods (Figures S5B and S5C). In our culture system, we confirmed the enriched gene expression of Schwann cell progenitor cell markers on ERBB3-positive and NGFR-positive cells (Figure S6C), suggesting other markers for higher purification are desired.

We identified FGFR4 and CDH13 by characterizing hiPSC-MuPCs. FGFR4 and CDH13 have been consistently shown to be expressed by quiescent SCs in human and mouse based on RNA-seq experiments (De Micheli et al., 2020a, 2020b; Pietrosemoli et al., 2017; Schaum et al., 2018; Xi et al., 2020). Our results showed that the new markers improved the efficacy of hiPSC-MuPC purification, as indicated by the increased expression of MYF5 and PAX7 and improved *in vitro* differentiation across different cell lines. Moreover, we confirmed that ~95% of FGFR4-positive cells were PAX7-positive cells (Figures 6C and 6D), indicating that FGFR4 can be used to highly purify myogenic precursor cells expressing PAX7. Interestingly, the expression of MYOD1 in FGFR4-positive cells was lowest compared with cells positive for the other tested markers. These findings were supported by our *in vitro* differentiation and transplantation studies, indicating that FGFR4 marks a subset of MuPCs with higher regeneration capacity, while CDH13 marks most myogenic cells.

One of the advantages of using hiPSCs instead of embryonic stem cells is the possibility of transplanting autologous gene-corrected cells (Hallett et al., 2015; Hanna et al., 2007; Shiba et al., 2016). By using CDH13 and FGFR4 as markers, we purified myogenic cells from DMD and DMD-corrected hiPSC-MuPCs. The DMD-corrected cells were able to differentiate *in vitro* into myotubes expressing DYSTROPHIN, and after transplantation they could restore DYSTROPHIN expression in engrafted myotubes, strengthening the suitability of CDH13 and FGFR4 as markers for clinical applications. Moreover, some of the transplanted cells were found to be PAX7 positive and to be localized in the SC position, suggesting a possible contribution to the SC pool. Further studies should address the capacity of these cells to contribute to future rounds of regeneration. Finally, in muscle transplanted with cells expressing either of the markers, we could identify hundreds of h-SPECTRIN-positive tiny fibers, which we speculate are de novo fibers that can fuse later to form new fibers, increasing the regeneration potential of the transplanted cells.

In summary, this study identified two new cell-surface markers for the purification of hiPSC-MuPCs: CDH13, which marks the majority of myogenic cells, and FGFR4, which marks a subset of MuPCs with higher regeneration capacity. We believe that these two markers will have
important implications for cell therapy for DMD and for the study of basic muscle development.

**EXPERIMENTAL PROCEDURES**

**Human iPSCs**
The hiPSC lines 201B7 (Takahashi et al., 2007), Ff-WJ14s01, DMD, and DMD-corrected were used for all the experiments. Ff-WJ14s01 (an HLA-homozygous iPSC line with the most frequent haplotype in Japan, abbreviated as S01 in this article) was established from cord blood cells by using episomal vectors (Okita et al., 2013) at the Facility for iPSC Cell Therapy, CIRA, Kyoto University. S01 was generated under written consent with approval of the Kyoto University Graduate School and Faculty of Medicine Ethics Committee (approval nos. E1762, G567, and Rinsho71). The DMD iPSC line (clone ID: CiRA00111) was prepared from dermal fibroblasts of a DMD patient with a deletion of exon 44 by using episomal vectors (Li et al., 2015). The DMD-corrected iPSC line was generated by knocking in exon 44 into CiRA00111 (Li et al., 2015). CiRA00111 was generated under written consent with approval of the Kyoto University Graduate School and Faculty of Medicine Ethics Committee (approval nos. R0091 and G259).

201B7 and S01 were used to generate the PAX7-Venus and MYF5-tdTomato reporter lines. All hiPSCs were cultured and maintained under feeder-free culture conditions as previously described (Nakagawa et al., 2014). Briefly, the cells were passaged once a week and cultured on an iMatrix-511 (Nippon) pre-coated dish in StemFit AK02N medium (Ajinomoto).

**Statistical Analysis**
All statistical analyses were performed using GraphPad Prism version 8.4.1 for Mac OS X (GraphPad Software, San Diego, CA, USA). Differences between two groups were analyzed with two-tailed Student’s t test, and differences between three or more groups were analyzed by ANOVA with Tukey’s range test for multiple comparisons. Differences were considered significant when the p value was <0.05. For the RNA-seq analysis, genes with at least a 2-fold enrichment and p < 0.05 were considered DEGs.

**OTHER MATERIALS AND METHODS**
Other materials and methods can be found in the Supplemental Experimental Procedures section.

**Data and Code Availability**
The accession number for the RNA-seq reported in this paper is DDBJ: DRA010568.

**SUPPLEMENTAL INFORMATION**
Supplemental Information can be found online at https://doi.org/10.1016/j.stemcr.2021.03.004.

**AUTHOR CONTRIBUTIONS**

**CONFLICT OF INTERESTS**
The authors declare that there is a patent being processed in Japan that is related to this work.

**ACKNOWLEDGMENTS**
We would like to thank Knut Woltjen for sharing some of the plasmids used in this study; April Pyle, Haibin Xi, and Peggie Chien for providing the developmental markers list; Takuya Yamamoto for helping with the data curation; Jun Otomo for valuable scientific discussions; Kanae Mitsunaga for her constant help with the FACS experiments; Masae Sato and Megumi Goto for technical assistance; and Peter Karagiannis for proofreading the manuscript. M.N. especially appreciates his family for continuous support. This work was mainly supported by a grant from The Core Center for iPSC Cell Research (19hm0104001h0007) and partially by a grant from The Acceleration Program for Intractable Diseases Research Utilizing Disease-Specific iPSC Cells (19bm0804005h0003), both of which are programs in the Research Center Network for Realization of Regenerative Medicine provided by the Japan Agency for Medical Research and Development, AMED (to H.S.).

Received: December 8, 2020
Revised: March 2, 2021
Accepted: March 2, 2021
Published: April 1, 2021

**REFERENCES**
and divergent features of quiescent and activated skeletal muscle stem cells. Skelet. Muscle 7, 28.