Supplemental information

Table of contents:

Figure S1

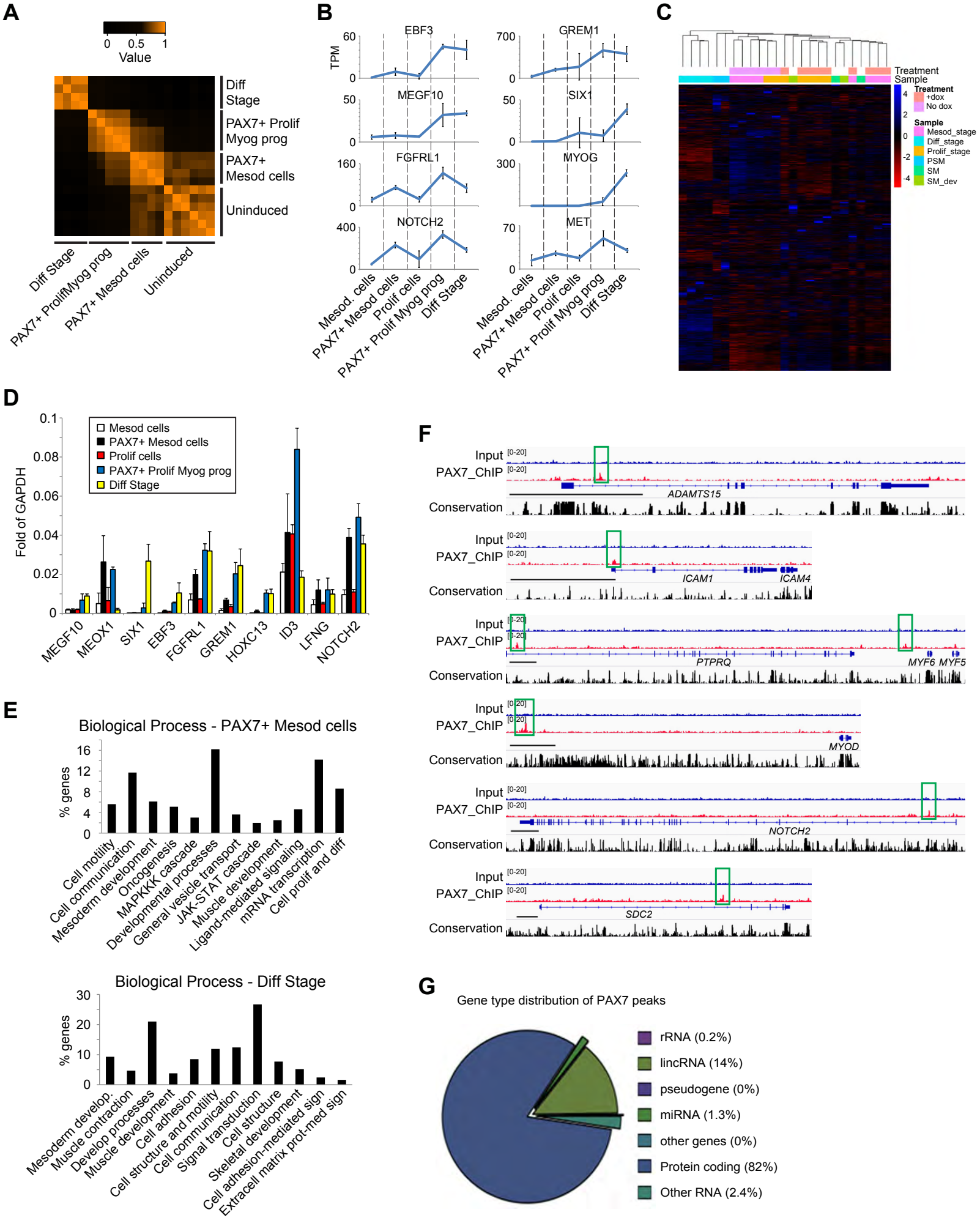
A
0 0.5 1
Value
Diff Stage
PAX7+ Prolif Myog prog
PAX7+ Mesod cells
Uninduced
Diff Stage
PAX7+ ProlifMyog prog
PAX7+ Mesod cells
Uninduced
B
TPM
EBF3
GREM1
MEGF10
SIX1
FGFRL1
MYOG
NOTCH2
MET
Mesod. cells
PAX7+ Mesod cells
Prolif cells
PAX7+ Prolif Myog prog
Diff Stage
C
Treatment
Sample
Treatment
+dox
No dox
Sample
Mesod_stage
Diff_stage
Prolif_stage
PSM
SM
SM_dev
D
Mesod cells
PAX7+ Mesod cells
Prolif cells
PAX7+ Prolif Myog prog
Diff Stage
Fold of GAPDH
MEGF10
MEOX1
SIX1
EBF3
FGFRL1
GREM1
HOXC13
ID3
LFNG
NOTCH2
E
Biological Process - PAX7+ Mesod cells
% genes
Cell motility
Cell communication
Mesoderm development
Oncogenesis
MAPKKK cascade
Developmental processes
General vesicle transport
JAK-STAT cascade
Muscle development
Ligand-mediated signaling
mRNA transcription
Cell prolif and diff
Biological Process - Diff Stage
% genes
Mesoderm develop.
Muscle contraction
Develop processes
Muscle development
Cell adhesion
Cell structure and motility
Cell communication
Signal transduction
Cell structure
Skeletal development
Cell adhesion-mediated sign
Extracell matrix prot-med sign
F
Input
PAX7_ChIP
Conservation
ADAMTS15
ICAM1
ICAM4
PTPRQ
MYF6
MYF5
MYOD
NOTCH2
SDC2
G
Gene type distribution of PAX7 peaks
rRNA (0.2%)
lincRNA (14%)
pseudogene (0%)
miRNA (1.3%)
other genes (0%)
Protein coding (82%)
Other RNA (2.4%)

Figure S1. Transcriptomic and genome occupancy profiling during PAX7-induced human skeletal myogenesis, Related to Figure 1. (**A**) Spearman correlations map of sample vs sample. (**B**) Expression profiles of selected genes from RNA-seq. TPM: Transcripts Per Million. (**C**) Heat map representing global transcriptome comparison using Cufflinks between control, PAX7-induced differentiating human pluripotent stem cells and tissues isolated from human embryo (GEO data: GSE39613). PSM: presomitic mesoderm; SM: nascent somite; SM dev: developed somite. Expression values are expressed as FPKM (Fragments Per Kilobase of transcript per Million reads). (**D**) RT-qPCR validation of selected genes from RNA-seq. (**E**) Gene Ontology classification of up-regulated genes based on Biological Process. Histogram reports % genes on the y axis and categories ordered based on p value (ascending left to right) on the x axis. (**F**) Snapshot of tracks from the Integrative Genome Viewer (IGV) browser showing PAX7 peaks (green boxes) at the ADAMTS15, ICAM1, MYF5-MYF6-PTPRQ cluster, MYOD, NOTCH2 and SDC2 loci. Conservation (Phastcons scale 0-1). Bar: 10Kb. (**G**) Gene type distribution of PAX7 peaks following annotation using PAVIS web resource.

Figure S2

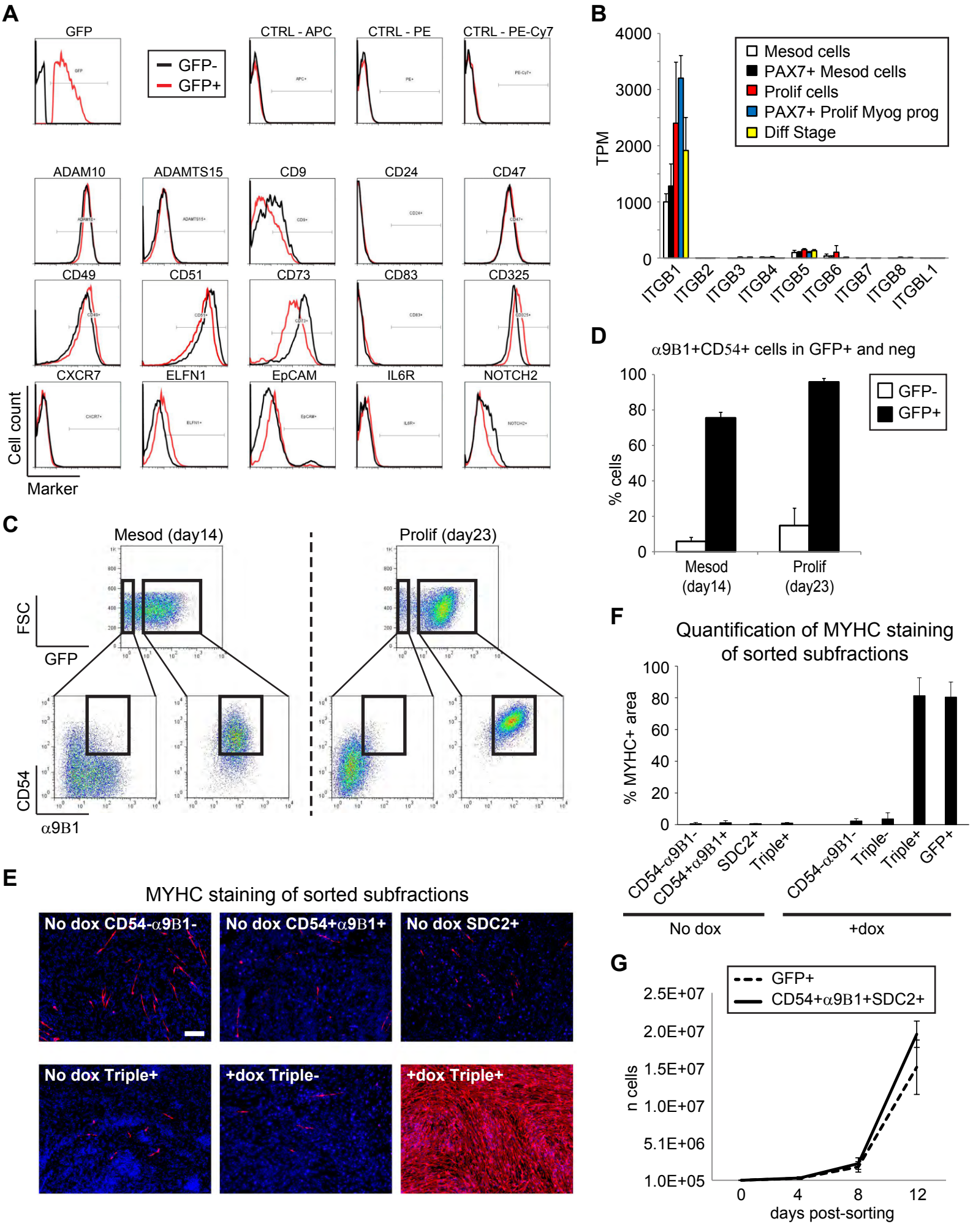
A
GFP
GFP-
GFP+
CTRL - APC
CTRL - PE
CTRL - PE-Cy7
ADAM10
ADAMTS15
CD9
CD24
CD47
CD49
CD51
CD73
CD83
CD325
CXCR7
ELFN1
EpCAM
IL6R
NOTCH2
Cell count
Marker
B
TPM
Mesod cells
PAX7+ Mesod cells
Prolif cells
PAX7+ Prolif Myog prog
Diff Stage
ITGB1
ITGB2
ITGB3
ITGB4
ITGB5
ITGB6
ITGB7
ITGB8
ITGBL1
C
Mesod (day14)
Prolif (day23)
FSC
GFP
CD54
α9B1
D
α9B1+CD54+ cells in GFP+ and neg
% cells
GFP-
GFP+
Mesod (day14)
Prolif (day23)
E
MYHC staining of sorted subfractions
No dox CD54-α9B1-
No dox CD54+α9B1+
No dox SDC2+
No dox Triple+
+dox Triple-
+dox Triple+
F
Quantification of MYHC staining of sorted subfractions
% MYHC+ area
CD54-α9B1-
CD54+α9B1+
SDC2+
Triple+
CD54-α9B1-
Triple-
Triple+
GFP+
No dox
+dox
G
GFP+
CD54+α9B1+SDC2+
n cells
days post-sorting

Figure S2. Human pluripotent-derived myogenic progenitors express CD54, α9β1 and SDC2, Related to Figure 2. (**A**) Histograms from FACS analysis of differentiating pluripotent stem cells using the indicated antibodies. PAX7-expressing cells (GFP+; red line) are compared to control cells (GFP-; black line). (**B**) Expression profiles of beta integrin genes across all RNA-seq samples. (**C**) Representative FACS analysis of GFP+ and GFP- cells for CD54 and α9β1 at day 14 of differentiation and proliferating myogenic progenitors following PAX7 induction. (**D**) Quantification of FACS data (n=5) from panel **C**. (**E**) Assessment of the myogenic potential of different FACS-sorted fractions following staining using CD54, α9β1 and SDC2 antibodies. Representative immunostaining (n=4) for the skeletal myogenic differentiation marker MYHC (red) following terminal differentiation of uninduced (No dox) CD54-α9β1-, CD54+α9β1+, SDC2+ and Triple+ as well as dox-treated (+dox) CD54-α9β1-SDC2- (Triple-) and Triple+ sorted cells. Nuclei: blue. Bar 100μm. (**F**) Quantification of myogenic potential of sorted subfractions (n=4) from panel **E**. (**G**) Growth curve analysis of GFP+ and CD54+α9β1+SDC2+ sorted cells (n=4).

Figure S3

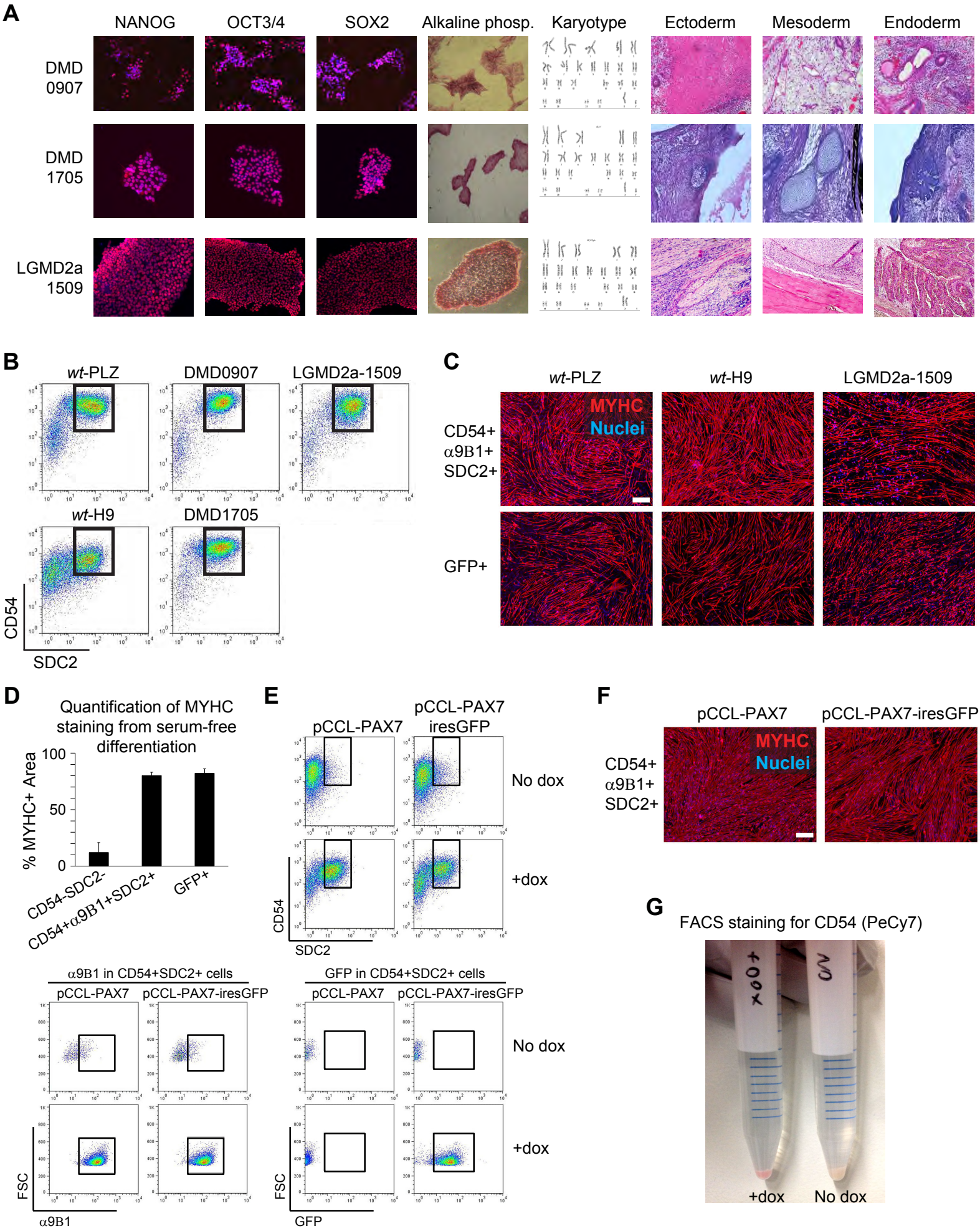
A
NANOG
OCT3/4
SOX2
Alkaline phosp.
Karyotype
Ectoderm
Mesoderm
Endoderm
DMD 0907
DMD 1705
LGMD2a 1509
B
wt-PLZ
DMD0907
LGMD2a-1509
wt-H9
DMD1705
CD54
SDC2
C
wt-PLZ
wt-H9
LGMD2a-1509
MYHC
Nuclei
CD54+ α9B1+ SDC2+
GFP+
D
Quantification of MYHC staining from serum-free differentiation
% MYHC+ Area
100
50
0
CD54-SDC2-
CD54+α9B1+SDC2+
GFP+
E
pCCL-PAX7
pCCL-PAX7 iresGFP
No dox
+dox
CD54
SDC2
α9B1 in CD54+SDC2+ cells
GFP in CD54+SDC2+ cells
FSC
α9B1
GFP
F
pCCL-PAX7
pCCL-PAX7-iresGFP
MYHC
Nuclei
CD54+ α9B1+ SDC2+
G
FACS staining for CD54 (PeCy7)
+dox
No dox

Figure S3. CD54, α9β1, SDC2 allow purification of myogenic progenitors from multiple cell lines, Related to Figures 2 and 3. (**A**) Characterization of the DMD0907, DMD1705 and LGMD2a-1509 human iPS cells. Immunofluorescence for pluripotency markers NANOG, OCT3/4 and SOX2; Alkaline phosphatase activity; karyotype; teratoma formation. (**B**) FACS analysis of PAX7-induced cultures from differentiating WT-PLZ (*wt* human iPS cells) and WT-H9 (human ES cells) cell lines, DMD1705 and DMD0907 (DMD human iPS cells) and LGMD2a-1509 human iPS cells. (**C**) Representative immunostaining (n=3) for the skeletal myogenic differentiation marker MYHC (red) following isolation of CD54+α9β1+SDC2+ and GFP+ cells from WT-PLZ, WT-H9 and LGMD2a-1509 cells. Nuclei: blue. Bar 100μm. (**D**) Quantification of myogenic potential of sorted subfractions (n=4) from differentiation in serum-free condition (panel **C** of Figure 3). (**E**) Representative FACS analysis of differentiating H9 cells (n=3) transduced with third generation lentiviral construct expressing PAX7 (pCCL-PAX7) or PAX7-ires-GFP (pCCL-PAX7-ires-GFP). Note the comparable expression pattern of CD54, α9β1 and SDC2 in dox-treated cultures from both cell lines, regardless of GFP expression. CD54, α9β1 and SDC2 are robustly up-regulated in the dox-group by PAX7. (**F**) Representative immunostaining (n=3) for the skeletal myogenic differentiation marker MYHC (red) following isolation of CD54+α9β1+SDC2+ from H9 pCCL-PAX7 and H9 pCCL-PAX7-ires-GFP. Bar 100μm. (**G**) CD54 is highly upregulated following PAX7 (+dox) induction. Upon immunostaining using anti-CD54-biotin followed by incubation with Streptavidin-PeCy7, cell pellet from dox-treated cultures is visibly positive to CD54 compared to uninduced (no dox) cells.

Figure S4

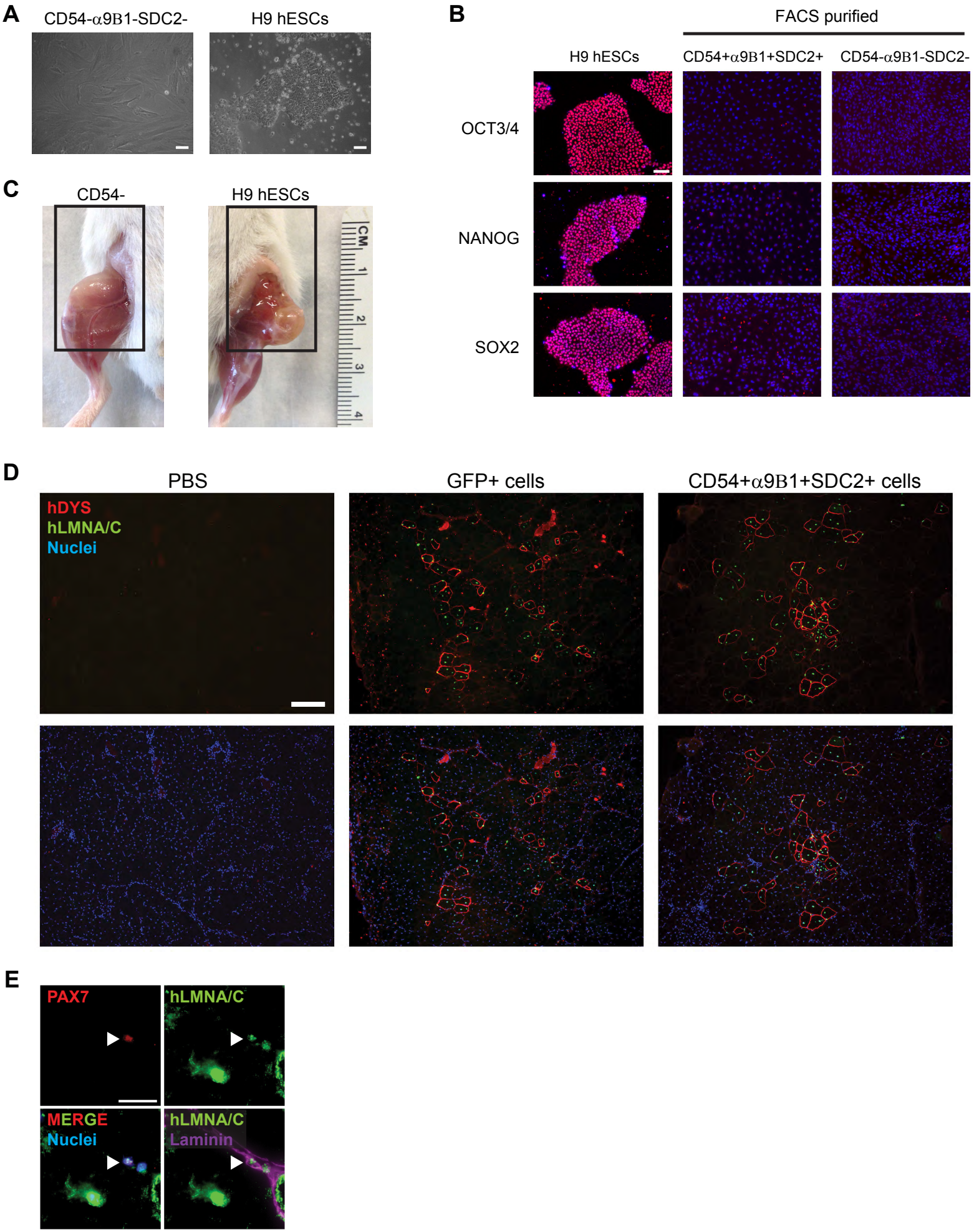
A
CD54-α9B1-SDC2-
H9 hESCs
B
FACS purified
H9 hESCs
CD54+α9B1+SDC2+
CD54-α9B1-SDC2-
OCT3/4
NANOG
SOX2
C
CD54-
H9 hESCs
D
PBS
GFP+ cells
CD54+α9B1+SDC2+ cells
hDYS
hLMNA/C
Nuclei
E
PAX7
hLMNA/C
MERGE
Nuclei
hLMNA/C
Laminin

Figure S4. Evaluation of safety and skeletal muscle regeneration potential of human pluripotent-derived myogenic progenitors, Related to Figures 3 and 4. (**A**) Phase-contrast pictures of human ES cells (H9 hESC) and CD54-α9β1- sorted cells. CD54-α9β1- cells were isolated at the proliferation stage from PAX7-induced cultures. (**B**) Immunostaining for the pluripotency markers OCT3/4 (upper panels), NANOG (middle panels) and SOX2 (lower panels). For the analysis, CD54+α9β1+SDC2+ and CD54-α9β1- sorted cells were compared to pluripotent cells (H9 hESCs) (n=3). Specific staining are shown in Red; Nuclei in blue. Bar 100µm. (**C**) Assessment of teratoma formation upon intramuscular injection of H9 ES cells (n=2, right) and CD54- FACS sorted cells (n=4, left). Both CD54- and CD54+ cells (n=4, data not shown) gave no visible sign of teratoma. (**D**) Representative immunostaining of cryosections from PBS-injected or GFP+ sorted and CD54+α9β1+SDC2+ cell-transplanted muscles (n=4) using human-specific antibodies for DYSTROPHIN (red) and LAMIN A/C (green). Nuclei (blue). Bar 100µm. (**E**) Representative immunostaining of cryosections from CD54+α9β1+SDC2+ cell-transplanted muscles (n=4) using antibodies for Pax7 (red), human-specific LAMIN A/C (green) and Laminin α-2 (Violet). Nuclei (blue). Arrowhead indicates a donor derived satellite cell. Bar 20µm.

II. SUPPLEMENTAL EXPERIMENTAL PROCEDURES

FACS analysis, sorting and MACS. Cultures from day 14 or day 23 (proliferating myogenic progenitors) were washed once with PBS and then harvested using Enzyme-free cell dissociation buffer (Gibco) for ~10 min at 37°C followed by gentle resuspension with pipette. Cells were collected, centrifuged, washed with PBS and filtered through a 70μm strain to remove cell clumps. Cell pellets were then resuspended in PBS supplemented with 10% FBS (PBSF) in the presence of Fc Block (1μl/million cells – BD bioscience) and incubated for 5 minutes. Staining was performed by adding 1μl of antibody/million cells followed by 20 min incubation on ice. For biotin-conjugated primary antibodies, samples were washed once with PBS then followed by 5 min incubation with PE-Cy7 conjugated Streptavidin (0.5μl/million cells diluted in PBSF). Cells were washed with PBS and then resuspended in PBSF containing propidium iodide to exclude dead cells. Samples were analyzed and sorted using a FACSAria II (BD Biosciences). For MACS purification, cells were collected and processed as described above and stained using Biotin-conjugated anti-CD54 antibody followed by incubation with Streptavidin-conjugated microbeads (Miltenyi) following the manufacturer's instructions. Magnetic cell separation was performed on MS columns (Miltenyi).

Antibodies used for FACS:

Antibody - FACS	**Manufacturer**	**Catalog number**
Streptavidin PE-Cyanine7	eBioscience	25-4317-82
PE anti-human Integrin α9β1 Antibody	BioLegend	351605
APC anti-human CD156c (ADAM10) Antibody	BioLegend	352705
Human ADAMTS15 MAb (Clone 561819)	R&D systems	MAB5149
Anti-Human CD9 PE	eBioscience	12-0098-41
Anti-Human CD24 PE-Cyanine7	eBioscience	25-0247-41
Anti-Human CD47 APC	eBioscience	17-0479-41
Anti-Human CD49e (Integrin alpha 5) Biotin	eBioscience	13-0496-80
Biotin anti-human CD51 Antibody	BioLegend	327906

Anti-Human CD54 (ICAM-1) APC	eBioscience	17-0549-41
Anti-Human CD54 (ICAM-1) Biotin	eBioscience	13-0549-82
APC anti-human CD73	eBioscience	17-0739-41
Anti-Human CD83 PE-Cyanine7	eBioscience	25-0839-41
PE/Cy7 anti-human CD325 (N-Cadherin) Antibody	BioLegend	350811
Biotin anti-human/mouse CXCR7 Antibody	BioLegend	331111
ELFN1 Antibody (N-12)	Santa Cruz Biotech.	sc-136660
Anti-Human CD326 (EpCAM) PE-Cyanine7	eBioscience	25-9326-41
PE anti-IL-6R alpha antibody	BD Biosciences	561696
Anti-Human/Mouse Notch2 PE	eBioscience	12-5786-80
Human Syndecan-2/CD362 APC-conjugated Antibody	R&D systems	FAB2965A

RNA isolation, library preparation and sequencing. Cells were resuspended in Trizol (Invitrogen) and purified using the Purelink RNA mini kit (Life Technologies) following the manufacturer's instructions. DNase treatment was performed on column during the RNA isolation procedure. RNAs were retro-transcribed using Superscript Vilo (Invitrogen). Gene expression analyses were performed using an amount of cDNA solution corresponding to 12.5 ng of starting RNA for each reaction. RT-qPCR analysis was performed using TaqMan Universal PCR Master Mix and TaqMan probes (Applied Biosystems).

Sequencing libraries were generated from 100 ng of total RNA using the Ligation Mediated Sequencing (LM-Seq) protocol (Hou et al., 2015) quantitated using the Qubit fluorometer (Life Technologies) following the author's instructions. The libraries were then pooled using 13 ng/sample for a 51+10 cycle Single Read run on the HiSeq 2500 (Illumina) by high output sequencing. The sequencer outputs were processed using Illumina's CASAVA-1.8.2 basecalling software. Demultiplexing assigned ~170 million reads across the 19 samples, ranging from 7 million to 11 million reads per sample. Of the assigned reads, about 1 million were discarded for low quality or the presence of sequencing adaptors in the reads.

Transcriptome analysis. Each sample's reads were then processed using RSEM version 1.2.3 (with bowtie-0.12.9 for the alignment step) (Langmead et al., 2009; Li and Dewey, 2011).

Percentage of reads mapped to the transcriptome ranged from 53% to 65%. Genes were filtered to select only those with median-normalized TPM (Transcripts Per Million reads) greater than 64 in at least one sample, and a fold change greater than 4 between the median-normalized Expected Counts (EC) of any two samples. Heat map of relative expression (displayed in Figure 1B) display for each gene G and sample S the value log2 ((1+median-normalized EC(G,S))/(1+mean over all Si median-normalized EC(G, Si)). Data were uploaded in NCBI Gene Expression Omnibus and are available at the accession number GSE98976.

Raw sequences were also analyzed using a customized pipeline (gopher-pipelines; https://bitbucket.org/jgarbe/gopher-pipelines/overview) developed and maintained by the Minnesota Supercomputing Institute (MSI). Briefly, quality controls were performed on each FASTQ files using FastQC (version 0.11.5) before and after adapter trimming with Trimmomatic (version 0.33) (Bolger et al., 2014). Between 7,105,374 to 10,951,120 reads (mean 8,929,324) passed filtering. Post-trimming sequences were aligned to GRCh38/hg38 reference genome using HISAT2 (version 2.0.2) (Kim et al., 2015). FPKM values were then generated using Cufflinks (version 2.2.1) (Trapnell et al., 2012).

Public data: RNA-seq data of developing human paraxial mesoderm from 4.5-5 weeks of gestation human embryos were downloaded from NCBI Gene Expression Omnibus (GSE90876). Sequence analysis was performed using the same method described above. Between 11,735,971 to 16,786,367 reads (mean 13,745,343) were generated per library and the average quality score passing quality filter were Q39. After adapter trimming, between 11,699,048 to 16,732,983 reads (mean 13,705,360) passed filtering.

Heat map (relative to Figure S1C): FPKM values were used to generate heat maps in R (version 3.2.3) using *pheatmap* (version 1.0.8) package. Briefly, FPKM values of the two datasets were

filtered independently to include transcripts with FPKM ≥ 1 in at least 10% of the samples. The FPKM values of the union of transcripts (21,918) were then log2 transformed and mean centered within each dataset. Then the two datasets were combined to generate heat maps. Hierarchical clustering was performed using euclidean distances and average linkage clustering method.

Chromatin-immunoprecipitation, library generation, sequencing and data analysis. Chromatin-immunoprecipitation was performed following the protocol described by Young and colleagues (Boyer et al., 2005) with minor modifications. Briefly, day 23 Proliferating Myogenic progenitors (+dox) were collected using trypsin and reaction was inhibited by adding 10%FBS/PBS. Single cells were washed once with PBS, resuspended in 10%FBS/PBS and supplemented with formaldehyde (final concentration 1%) for crosslinking of protein-DNA complexes (10 minutes at RT) followed by quenching with glycine. Cells were snap-frozen in liquid nitrogen and stored at -80°C if not processed immediately. Cell pellets were incubated in lysis buffer LB1 supplemented with protease inhibitors (50mM HEPES KOH pH7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP40, 0.25% Triton X-100 + Complete-mini - Roche) for 10 minutes at +4°C followed by incubation in buffer LB2 supplemented with protease inhibitors (10mM TRIS HCl pH 8, 200mM NaCl, 1mM EDTA, 0.5mM EGTA + Complete-mini - Roche) for 10 minutes at +4°C. Cell pellet was then resuspended in LB3 supplemented with protease inhibitors (10mM TRIS HCl pH 8, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% Sodium Deoxycholate, 0.5% N-lauroylsarcosine + Complete-mini - Roche) and then sonicated with a Branson sonicator at 18% power for 1 minute with intervals of 10 sec ON-10 sec OFF. Each sample was subjected to 5-6 cycles of sonication to reach an average chromatin size of 300bp. After shearing, samples were centrifuged for 10 minutes at 16000g and snap frozen in liquid nitrogen if not processed immediately. For each ChIP, 20μg of chromatin (diluted to

400µl) were precleared for 4h at 4°C with 15µl of BSA-blocked Protein G-conjugated sepharose beads (GE healthcare) and then supplemented with 1/10 volume of 10% Triton X-100. Immunoprecipitation was performed by overnight incubation with normal mouse IgG (8µg – Santa Cruz Biotechnology) or anti-Pax7 antibody (1:50 – Developmental Studies Hybridoma Bank). Immune complexes were recovered by incubation with 15µl of BSA-blocked Protein G-conjugated sepharose beads for 4h at 4°C and then washed 5 times with RIPA wash buffer (50mM HEPES KOH pH7.5, 500mM LiCl, 1mM EDTA, 1% NP40, 0.25% Triton X-100, 0.7% Sodium Deoxycholate) and one time with TEN buffer (10mM TRIS HCl pH 8, 1mM EDTA, 50mM NaCl). Immunoprecipitated chromatin was recovered by incubating beads with 200µl of Elution buffer (50mM TRIS HCl pH 8, 10mM EDTA, 1% Sodium Dodecyl Sulfate) for 20 minutes at 65°C. Chromatin from IP and Input (equivalent to 1% of starting material) was reverse crosslinked overnight at 65°C, then diluted 1:1 with TE (10mM TRIS HCl pH 8, 1mM EDTA) supplemented with 4µl of RNaseA 20mg/ml and incubated for 2 hours at 37°C followed by Proteinase K treatment (4µl of 20mg/ml stock for each sample) for 30 minutes at 55°C. DNA was purified by Phenol-chloroform-isoamyl alcol extraction (twice) followed by chloroform extraction, then supplemented with 1/10 of volume of 3M Sodium Acetate pH 5 and 1.5µl of Glycogen and precipitated with 2 volumes of 100% Ethanol at -80°C for >1 hour. Followed 30 minutes centrifuge at 16000g, pellet were washed with 75% ethanol, air dried and dissolved in 45µl H2O.

Libraries were generated following a gel-free protocol using AMPure XP beads (Beckman Coulter) for all the purification and size selection steps. 10ng or less of DNA were end repaired using End-it DNA end repair (Epicentre), then A-tailed using Klenow Fragment (3'→5' exo- NEB) followed by adapter-barcode ligation using T4 DNA ligase (Enzymatics). Illumina

compatible adapter-barcodes were purchased from BIOO scientific. After ligation, DNAs were negatively size selected using 0.5x Ampure XP beads and unbound DNAs were positively size selected by adding 0.4x Ampure XP beads (this step allows for retention of DNA fragments ranging 200-500bp). Libraries were amplified using Phusion High Fidelity PCR master mix 2x (NEB) with a 16 cycles program. Purified libraries were then submitted to the University of Minnesota Genomic Center (UMGC) for quantification, quality control and sequencing. Libraries were pooled and sequenced on a lane of Single-End run on a HiSeq2500 operated at High Output mode (Illumina). The sequencer outputs were processed using the Galaxy platform (Goecks et al., 2010) available at the Minnesota Supercomputing Institute (MSI). Demultiplexing assigned ~60 million reads across the 2 samples (~30 million reads per sample). Each sample's reads were then aligned to the human genome (hg38) using Bowtie2 (Langmead and Salzberg, 2012) followed by removal of PCR duplicates using the SAM tool rmdup 1.0.1 (Li et al., 2009). Peak calling was performed using MACS 1.0.1 (Zhang et al., 2008) (parameters: -s 51 -bw 300, -p 1e-05 -m 16). Bigwig files for visualization on IGV (Thorvaldsdottir et al., 2013) were generated by converting the wig files obtained from MACS using the Galaxy tool Wig-to-BigWig. Motif analysis was performed using the SeqPos tool (Liu et al., 2011). For this analysis the Top 3000 PAX7 peaks (based on the MACS score) were filtered to exclude regions overlapping peaks detected in the Input and the human blacklist dataset (Consortium, 2012). Annotation of PAX7 peaks was performed using the web resource PAVIS (Huang et al., 2013). Data were uploaded in NCBI Gene Expression Omnibus and are available at the accession number GSE98976.

Immunostaining. Cultured cells were fixed with 4% PFA for 10 min at +4°C followed by permeabilization with 0.1% Triton in PBS for 10 min at RT. After PBS wash, cells were blocked

for 30 min at RT with 3% BSA in PBS and then incubated with primary antibodies diluted in blocking solution overnight at +4°C. Cells were then washed and incubated with secondary antibody diluted in blocking solution supplemented with DAPI for 1 hour at RT. After PBS washes, cells were maintained in PBS until final analysis. Pictures were acquired using an inverted fluorescence microscope (Zeiss). Tissue cryosections were permeabilized with 0.3% Triton X-100 in PBS for 20 min at RT, then blocked and incubated with primary antibodies overnight as described. Sections were then washed and incubated with secondary antibodies as described for cells. Antibodies: Myosin Heavy Chain (clone MF20 – Developmental Studies Hybridoma Bank); human specific Lamin A/C (Abcam); human specific Dystrophin (EMD Millipore); M-cadherin (BD Biosciences), Pax7 (Developmental Studies Hybridoma Bank), OCT3/4, NANOG, SOX2 and Laminin α-2 (all from Santa Cruz Biotechnology); Alexa-555 goat anti-mouse, Alexa-488 goat anti-rabbit and Alexa-647 goat anti-rat (Thermo Fisher). Antibodies used for Immunostaining and Chromatin-immunoprecipitation.

Antibody	**Manufacturer**	**Catalog number**
Myosin Heavy Chain	DSHB	Clone MF20
Pax7	DSHB	Clone Pax7
Human Dystrophin, clone 2C6 (MANDYS106)	EMD	MABT827
Human Lamin A + C antibody [EPR4100]	Abcam	ab108595
M-cadherin	BD Biosciences	611101
Laminin α-2 (4H8-2)	Santa Cruz Biotech.	sc-59854
OCT3-4	Santa Cruz Biotech.	sc-5279
NANOG	Santa Cruz Biotech.	sc-374103
SOX2	Santa Cruz Biotech.	sc-17320
Alexa-555 goat anti-mouse	ThermoFisher	A-21424
Alexa-488 goat anti-rabbit	ThermoFisher	A-11008
Alexa-647 goat anti-rat	ThermoFisher	A-11006

Quantification and Statistical Analysis. Analyses were performed using the ImageJ distribution Fiji (Schindelin et al., 2012). Quantification of *in vivo* engraftment was performed by counting the number of hLMNA-C positive fibers in 4 representative pictures for each transplanted mouse

using the plugin *Cell counter*. Data represent mean+standard error of n= 3 independent transplanted muscles. Analysis of myogenic differentiation potential of sorted subfractions was performed as follow: color channels were separated and threshold level for the red and blue channels were adjusted in order to select the area positive respectively to MYHC (red) and DAPI (blue). The area positive for each channel was analyzed using *Analyze Particle* using 0-Infinity as Size parameter. Finally, the percentage of MYHC+ area for each image was normalized based on nuclear staining (DAPI+). Data represent mean ± standard deviation of at least 2 representative pictures for each independent experiments (n=3).

III. ASSOCIATED REFERENCES

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy, T. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol *11*, R86.

Hou, Z., Jiang, P., Swanson, S.A., Elwell, A.L., Nguyen, B.K., Bolin, J.M., Stewart, R., and Thomson, J.A. (2015). A cost-effective RNA sequencing protocol for large-scale gene expression studies. Sci Rep *5*, 9570.

Huang, W., Loganantharaj, R., Schroeder, B., Fargo, D., and Li, L. (2013). PAVIS: a tool for Peak Annotation and Visualization. Bioinformatics *29*, 3097-3099.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Methods *12*, 357-360.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol *12*, R83.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B.*, et al.* (2012). Fiji: an open-source platform for biological-image analysis. Nat Methods *9*, 676-682.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*, 178-192.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

IV. Table S2. Genes in BP00124: Cell adhesion (From DAVID bioinformatics), Related to Figure 1.

List of genes associated with cell adhesion, which were found up-regulated following PAX7 induction in PAX7+ proliferating myogenic progenitors.

GENE_SYMBOL	GENE NAME
ASAP1	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1
CD151	CD151 molecule (Raph blood group)
CD47	CD47 molecule
CDON	Cdon homolog (mouse)
FAT1	FAT tumor suppressor homolog 1 (Drosophila)
FAT4	FAT tumor suppressor homolog 4 (Drosophila)
GIT2	G protein-coupled receptor kinase interacting ArfGAP 2
ARHGEF12	Rho guanine nucleotide exchange factor (GEF) 12
ROCK2	Rho-associated, coiled-coil containing protein kinase 2
ARVCF	armadillo repeat gene deletes in velocardiofacial syndrome
ATRNL1	attractin-like 1
CDH15	cadherin 15, type 1, M-cadherin (myotubule)
CDH2	cadherin 2, type 1, N-cadherin (neuronal)
CLSTN3	calsyntenin 3
CERCAM	cerebral endothelial cell adhesion molecule
COL2A1	collagen, type II, alpha 1
COL4A2	collagen, type IV, alpha 2
COL5A2	collagen, type V, alpha 2
GPC1	glypican 1
GPC4	glypican 4
HSPG2	heparan sulfate proteoglycan 2
ITGA1	integrin, alpha 1
ITGA5	integrin, alpha 5 (fibronectin receptor, alpha polypeptide)
ITGAV	integrin, alpha V (vitronectin receptor, alpha polypeptide, antigen CD51)
ICAM1	intercellular adhesion molecule 1
JAM2	junctional adhesion molecule 2
KIRREL	kin of IRRE like (Drosophila)
LAMA4	laminin, alpha 4
LRRC4B	leucine rich repeat containing 4B
LOX	lysyl oxidase
LOXL1	lysyl oxidase-like 1
LOXL2	lysyl oxidase-like 2
MXRA5	matrix-remodelling associated 5
MAGED1	melanoma antigen family D, 1
MAGED2	melanoma antigen family D, 2

NCAM1	neural cell adhesion molecule 1
NRCAM	neuronal cell adhesion molecule
NID1	nidogen 1
PCDH7	protocadherin 7
PCDHA10	protocadherin alpha 13; protocadherin alpha 10; protocadherin alpha subfamily C, 1; protocadherin alpha subfamily C, 2
PCDHB2	protocadherin beta 2
PCDHGA11	protocadherin gamma subfamily A, 11
PCDHGB4	protocadherin gamma subfamily B, 4
PCDHGB7	protocadherin gamma subfamily B, 7
PCDHGC3	protocadherin gamma subfamily C, 3; protocadherin gamma subfamily C, 5; protocadherin gamma subfamily C, 4; protocadherin gamma subfamily A, 12
SDC2	syndecan 2
SDC3	syndecan 3
TNS3	tensin 3
TSPAN9	tetraspanin 9
TRO	trophinin