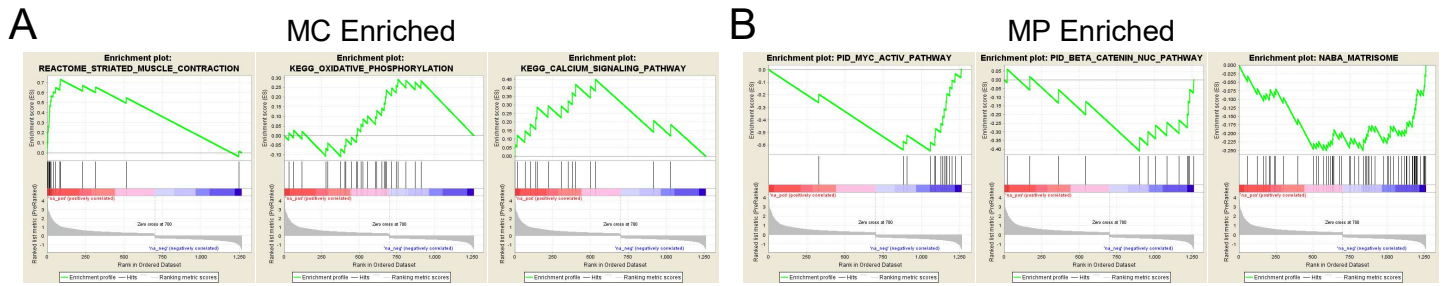
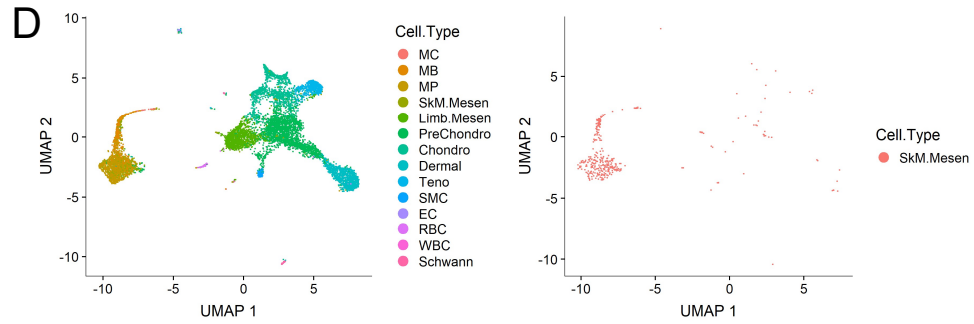
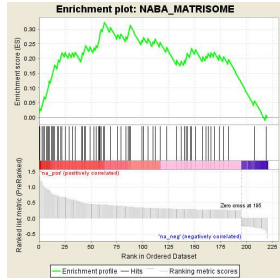


Figure S1. Related to Figure 1. Cell types present in limbs and skeletal muscle tissues at different human developmental stages.

(A-H) Single cells from all human biological replicates grouped by the indicated ages are plotted on tSNE space. Normalized expression levels of cell type marker genes are color-scaled (purple: high expression and gray: low expression). Small cell populations at certain developmental stages are boxed for easy visualization. **(I)** Proportions of SkM cells of each human biological replicates grouped by the indicated ages. Bars represent average SkM proportion of biological replicates within each age group and dots represent each sample.



C SkM.Mesen Enriched



E

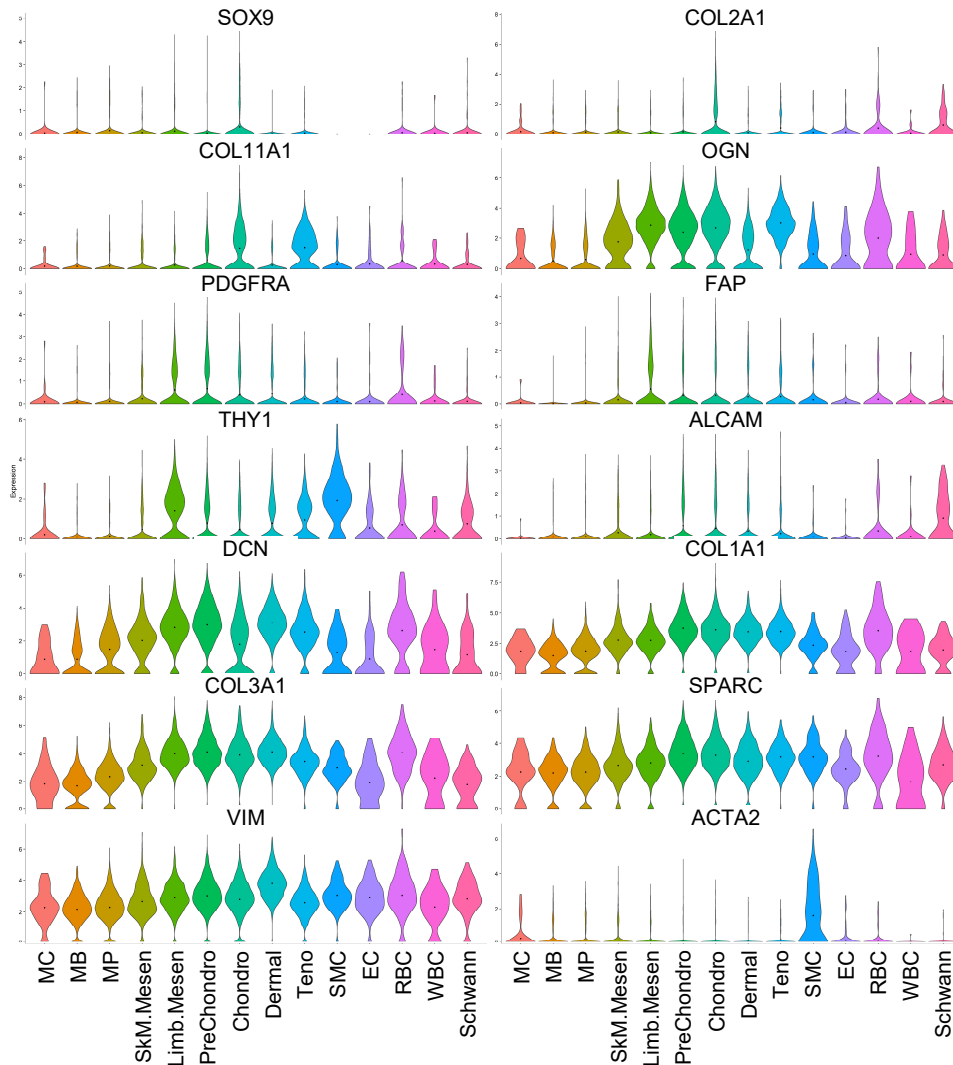


Figure S2. Related to Figures 2 and 3. Characterization of skeletal myogenic subpopulations in human fetal limbs.

(A and B) Selected enriched GSEA pathways from genes upregulated in MC vs. MP **(A)** or MP vs. MC **(B)** from fetal week 9 human SkM cells. Spikes in the red left side regions represent genes enriched in MC and those in the blue right side regions enriched in MP. **(C)** Selected enriched GSEA pathway from genes upregulated in SkM.Mesen vs. the main SkM subpopulations (MP, MB and MC) from fetal week 9 human SkM cells. Spikes in the red left side region represent genes enriched in SkM.Mesen and those in the blue right side region enriched in the main SkM subpopulations. **(D)** Left panel: single cells from all biological replicates from fetal week 9 human SkM cells are plotted on UMAP space in Monocle3 and colored by their cell types and myogenic subtypes assigned by Seurat. Right panel: only cells from the SkM.Mesen population are plotted to show their overlap with the main SkM populations. **(E)** Violin plots generated by Monocle3 showing normalized expression counts of chondrogenic and mesenchymal/fibroblastic markers across cell types and myogenic subtypes. Dots of each violin represent mean expression levels and width is proportional to the number of cells expressing the markers at a certain level.

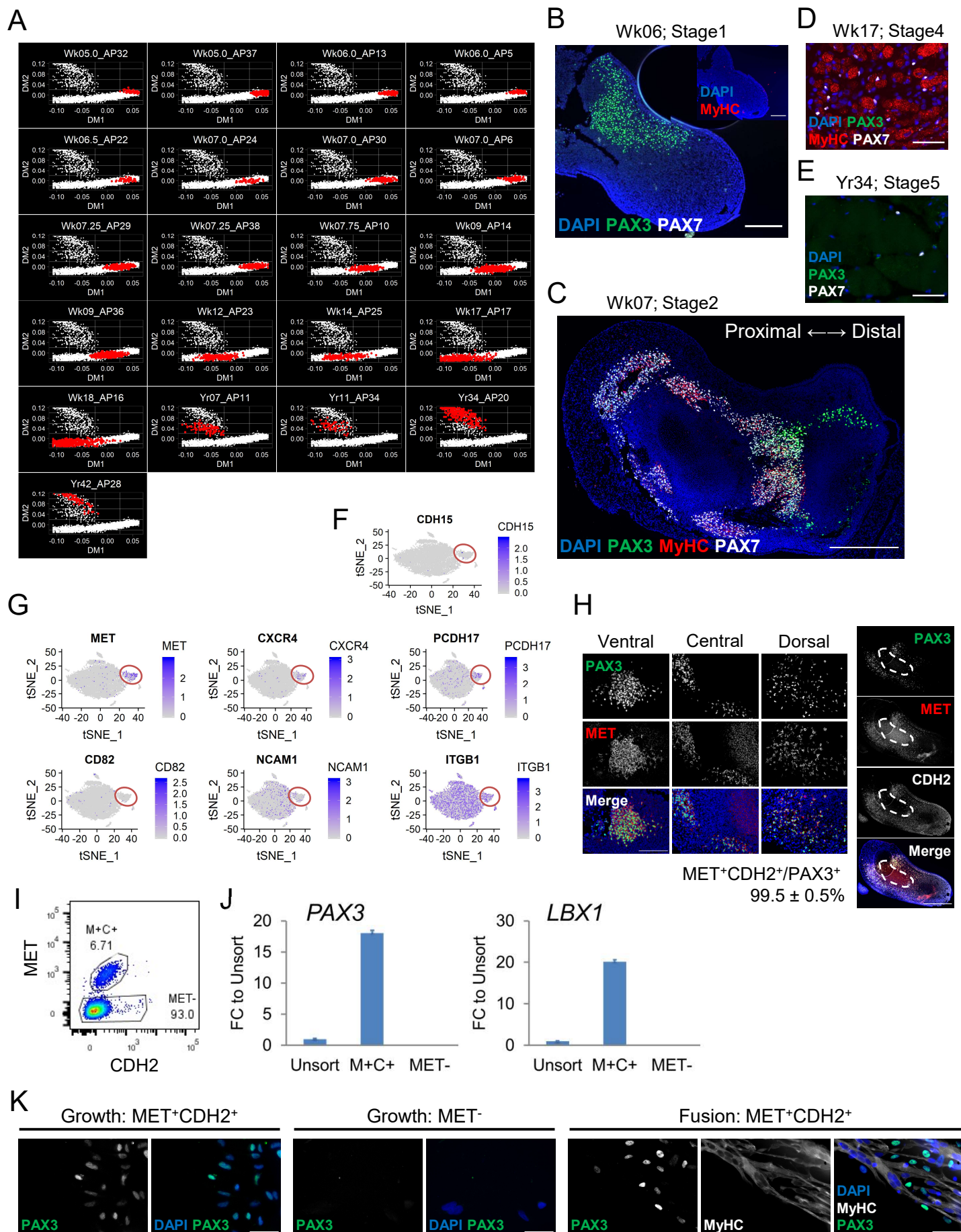


Figure S3. Related to Figure 4. Distinct SMPC and SC populations across human development and isolation of myogenic cells from early human embryonic limbs.

(A) DM plots highlighting cells from each *in vivo* human samples. Highlighted cells from a specific sample are encoded with red, while all the other cells colored white. (B-E) Representative IHC images of PAX3, PAX7 and MyHC on human tissue sections from embryonic week 6 hindlimb (B), embryonic week 7 forelimb (C), fetal week 17 quadriceps (D) and adult year 34 quadriceps (E). Nuclei in all images were counterstained with DAPI. In (B), images were taken in a mosaic mode and stitched together for the whole hindlimb and inset shows a stitched MyHC mosaic image taken from a neighboring section. Scale bars: 250 μm . In (C), images were taken in a mosaic mode and stitched together for the whole forelimb. Scale bar: 500 μm . In (D), scale bar: 50 μm . In (E), no specific PAX3 signals were detected and the green color represent autofluorescence emitted from myofibers. Scale bar: 50 μm . (F-G) tSNE plots showing normalized expression levels of CDH15 (F) or additional cell surface molecules (G) in cells from embryonic week 5-6 hindlimbs. The SkM population was red-circled. (H-K) Identification of cell surface markers to isolate myogenic cells from human week 5-6 limbs. (H) Representative IHC images of co-expression of PAX3 and MET from hindlimbs of a human week 6 embryo. Nuclei were counterstained with DAPI. Left panel: Representative images taken at different section levels. Scale bar: 150 μm . Right panel: representative images of the whole hindlimb at the central section level (images were taken in a mosaic mode and stitched together). Area enclosed by white dotted circle contains cells expressing low levels of MET, but not PAX3 or CDH2. Scale bar: 500 μm . (I) Representative FACS plot of sorting MET⁺CDH2⁺ (M⁺C⁺) and MET⁻ populations from hindlimbs of a week 5 human embryo. Numbers indicate the percentage of each population. (J) qRT-PCR of PAX3 and LBX1 of the sorted populations. Expression values are normalized to GAPDH and fold change of unsorted cells is set to 1. Data represent mean+SD of technical triplicates of 1 out of 2 representative embryos. (K) Images of IF staining of sorted cells cultured under growth conditions for PAX3 and fusion conditions for PAX3 and MyHC. Since MET⁻ cells were not supported by the skeletal muscle growth medium, they were not further subjected to the fusion medium. Nuclei were counterstained by DAPI. Scale bars: 50 μm .

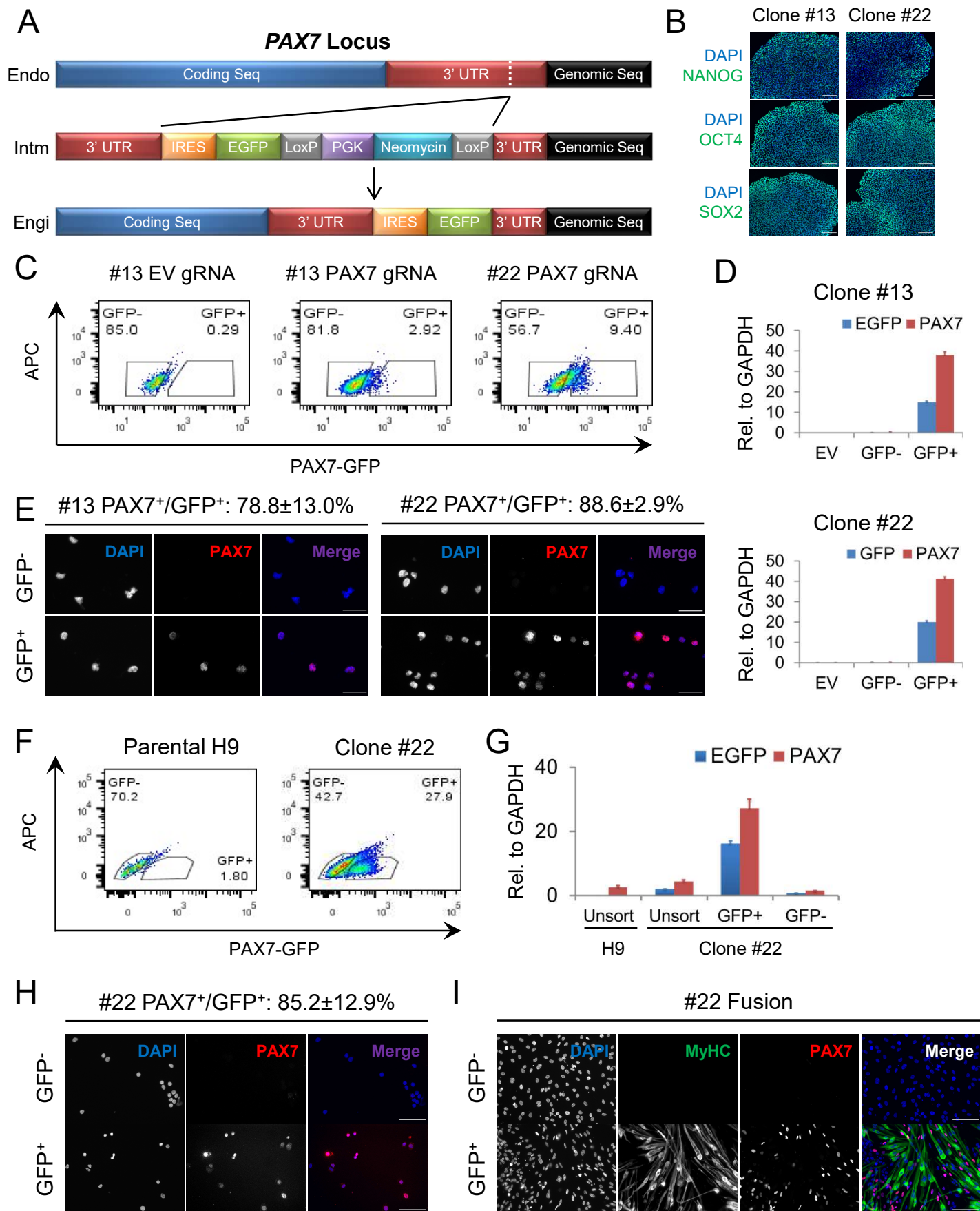


Figure S4. Related to Figures 5 and 6. Construction of the PAX7-GFP reporter cell lines.

(A) Schematic of reporter targeting strategy. The white dotted line in the top track indicates Cas9 cutting site. Endo, endogenous; Intm, intermediate; Engi, engineered. **(B)** Pluripotency marker IF staining of generated reporter cell clones #13 and #22. Nuclei were counterstained by DAPI. Scale bars: 150 μm . **(C-E)** Verification of reporter cells using dCas9-VPR system. **(C)** Reporter cells from clone #13 and #22 transfected by plasmids encoding dCas9-VPR along with those encoding *PAX7* promoter region targeting gRNAs or empty vectors (EV), were sorted into GFP⁺ and GFP⁻ fractions. **(D)** qRT-PCR of *GFP* and *PAX7* expression on sorted cells. Data are normalized to GAPDH and represent mean+SD of technical triplicates of a representative experiment. **(E)** Representative images of sorted cells subjected to cytospin followed by IF of PAX7. Numbers represent average percentage of GFP⁺-sorted cells expressing PAX7 with SEM, from 2 independent experiments. Nuclei were counterstained by DAPI. Scale bars: 50 μm . **(F-I)** Verification of reporter cells using directed differentiation (HX protocol). **(F)** Representative FACS plots of clone #22 reporter cells differentiated using HX protocol to sort for GFP⁺ and GFP⁻ fractions. H9 parental cells from the same differentiation batch were used as GFP gating controls. **(G)** qRT-PCR of *GFP* and *PAX7* expression on sorted #22 cells as well as unsorted #22 or parental H9 cells. Data are normalized to GAPDH and represent mean+SD of technical triplicates of 1 out of 2 representative experiments. **(H)** Representative images of sorted cells subjected to cytospin followed by IF of PAX7. Numbers represent average percentage of GFP⁺-sorted cells expressing PAX7 with SEM, from 2 independent experiments. Nuclei were counterstained by DAPI. Scale bars: 100 μm . **(I)** Images of IF staining of sorted cells cultured under fusion conditions for PAX7 and MyHC. Nuclei were counterstained by DAPI. Scale bars: 100 μm .

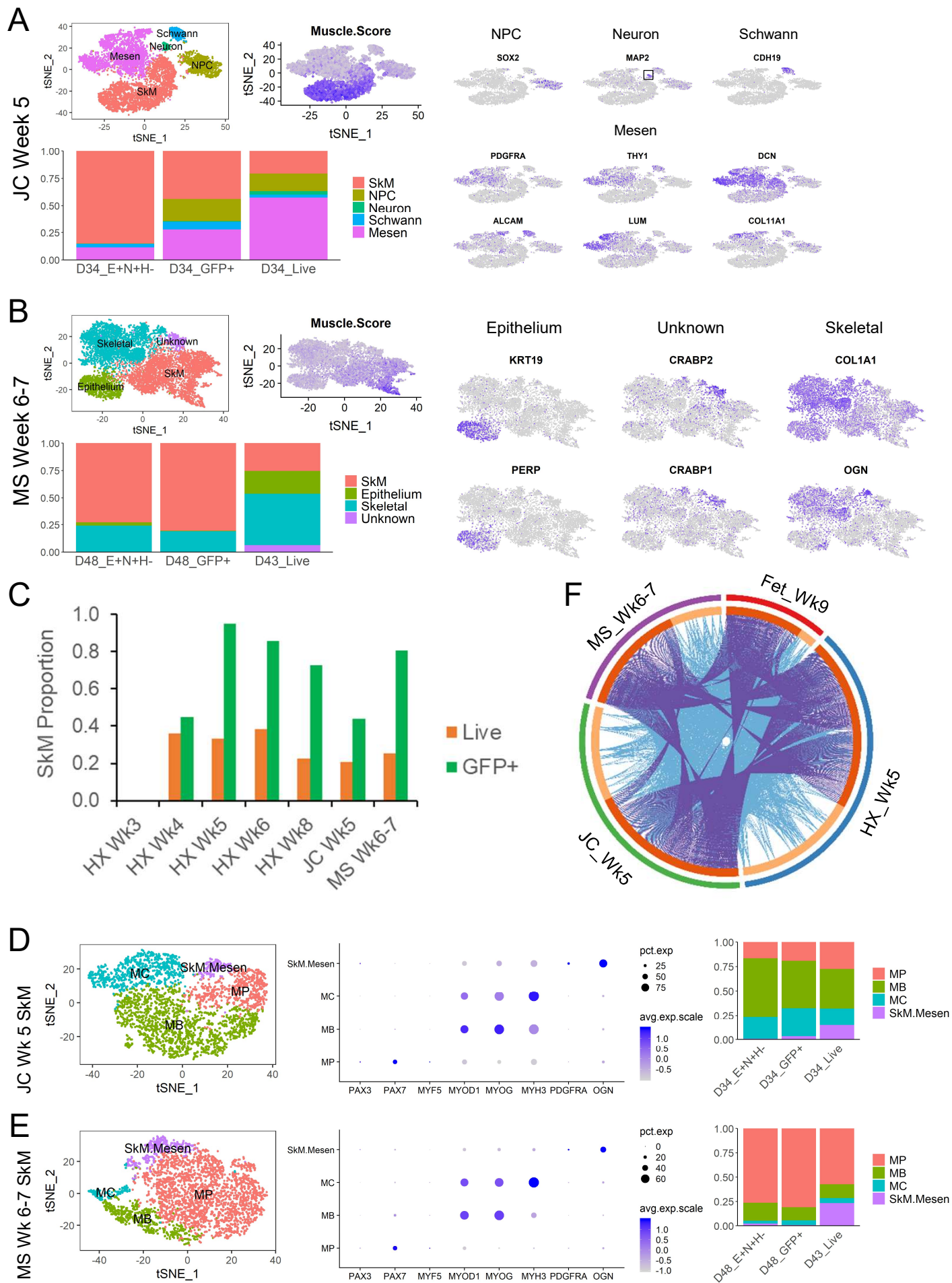


Figure S5. Related to Figures 5 and 6. scRNA-seq reveals heterogeneous cell types and skeletal muscle subpopulations from additional hPSC myogenic differentiation protocols.

(A) Single cells from all hPSC-derived samples (both unenriched and enriched) using JC protocol at week 5 of differentiation are plotted on tSNE space. They are color-coded by assigned cell types, or color-scaled based on normalized expression levels of “Muscle.Score” and other cell type gene markers (purple: high expression and gray: low expression). The small *MAP2*⁺ neuron population is boxed for easy visualization. Bottom left bar plots showing the proportion of cell types distributed in each enriched or unenriched samples. **(B)** Similar to **(A)** with week 6-7 cells differentiated by MS protocol. **(C)** Quantification of SkM cell proportion in live- or GFP⁺-sorted samples examined by scRNA-seq from 3 protocols. **(D)** Left panel: single cells (classified as the “SkM” cell type from **(A)**) from enriched and unenriched samples at week 5 under JC protocol are plotted on tSNE space and colored by their assigned myogenic subtypes. Middle panel: dot plots of selected marker genes in each cell subtype. Normalized gene expression counts are averaged across all cells in given cell clusters. Sizes of dots represent percentage of marker expressing cells within each population and colors show levels of markers in expressing cells scaled across populations. Right panel: bar plots showing the proportion of myogenic subtypes distributed in each enriched or unenriched samples. **(E)** Similar to **(D)** with week 6-7 cells differentiated by MS protocol. **(F)** Circos plot of shared and linked genes enriched in the SkM.Mesen subpopulations over the main SkM subpopulations from all three *in vitro* differentiation protocols and from *in vivo* fetal week 9 samples. Outer ring is color-coded by sample identities. Red regions of inner ring denote genes shared by at least two gene sets and orange regions contain genes that are unique. Purple lines connecting exactly same genes and light blue lines connecting different genes belonging to similar biological processes.

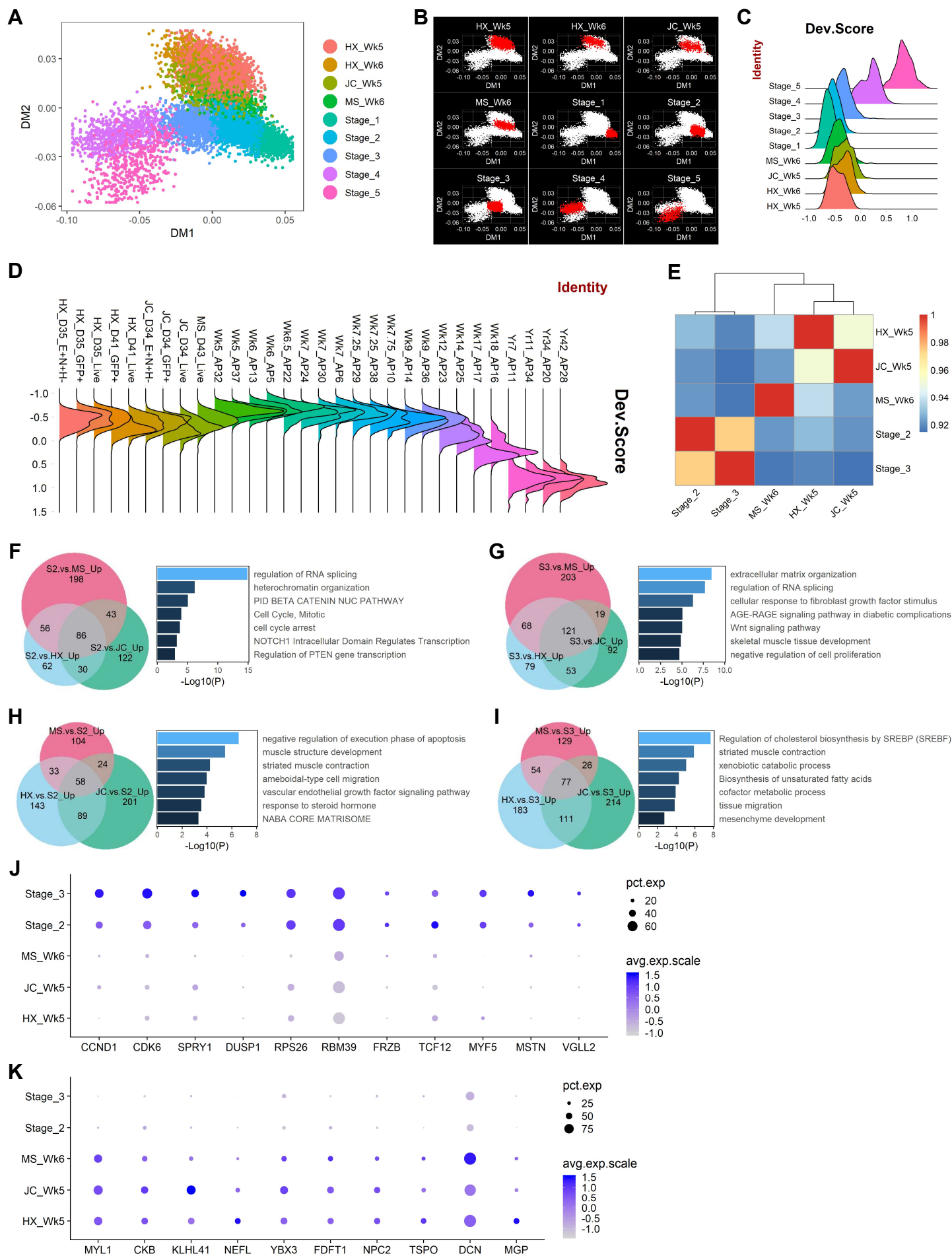


Figure S6. Related to Figure 7. *In vitro* SMPCs derived from multiple hPSC myogenic differentiation protocols are different from *in vivo* human myogenic progenitor cells during embryonic-to-fetal transition.

(A-C) Single cell trajectory **(A)** and individual stage highlighting **(B)** on DM space, and ridge plot “Dev.Score” distribution **(C)** as similar to Figures 6A-6C with SMPCs and SCs from *in vivo* human samples and *in vitro* derived from hPSCs using HX, JC and MS protocols at the indicated differentiation time points. **(D)** Ridge plot of “Dev.Score” distribution similar to **(C)** but grouped by individual samples. **(E)** Spearman correlation coefficients between each group of cells were calculated. Heatmap showing the groups’ correlations and hierarchical clustering. **(F-I)** Differential gene expression analyses were performed to identify genes commonly upregulated from *in vivo* stage 2/3 SMPCs compared to *in vitro* hPSC-SMPCs derived from three different protocols **(F and G)**, and vice versa **(H and I)**. Left panels show Venn diagrams of overlapping enriched genes and right panels depict selected top enriched biological processes and signaling pathways from the overlapping gene lists. **(J and K)** Dot plots showing expression of selected genes from top ranked GO terms that are enriched in *in vivo* stage 2/3 SMPCs **(J)** or in hPSC-derived SMPCs from multiple protocols **(K)**.

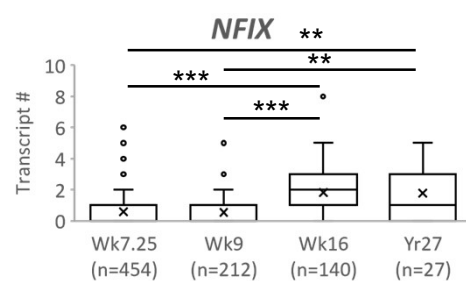
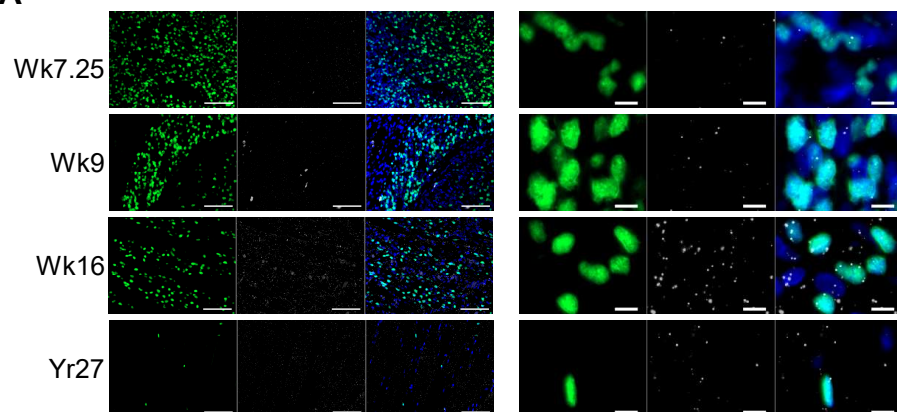
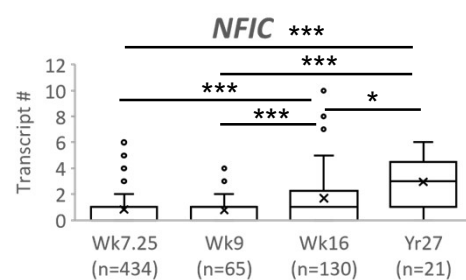
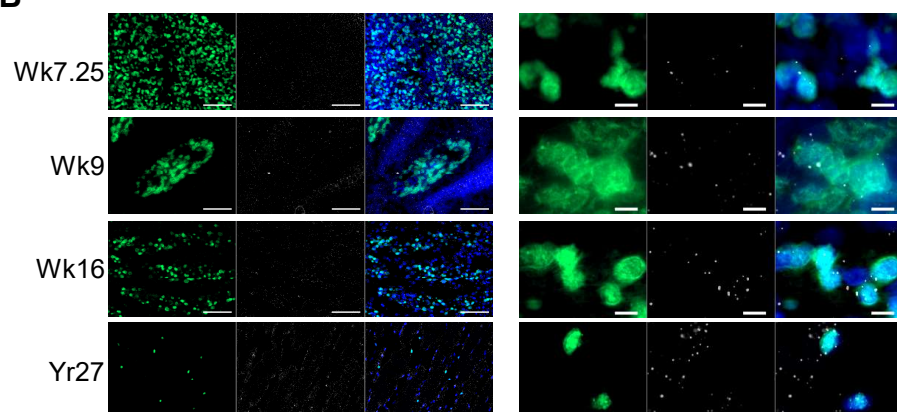
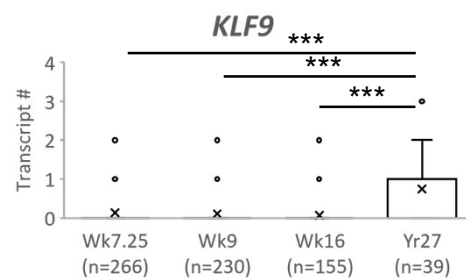
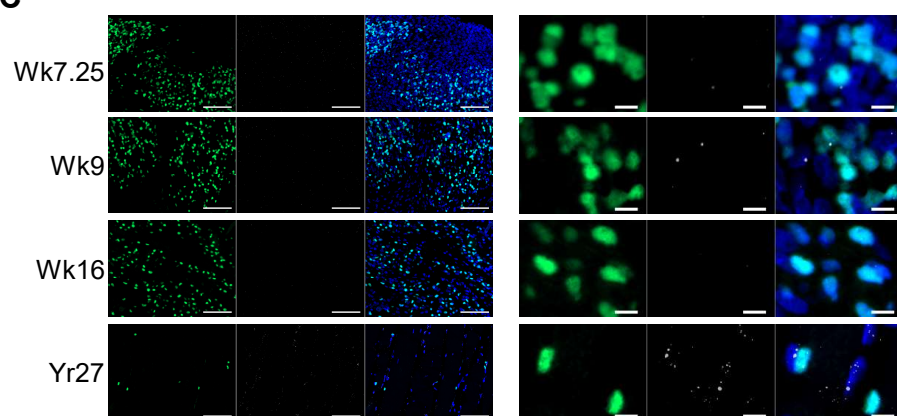
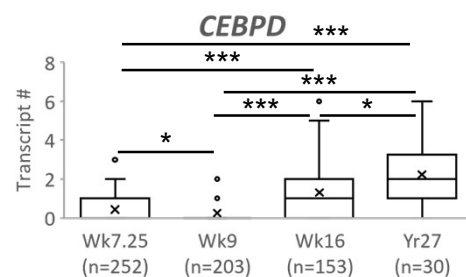
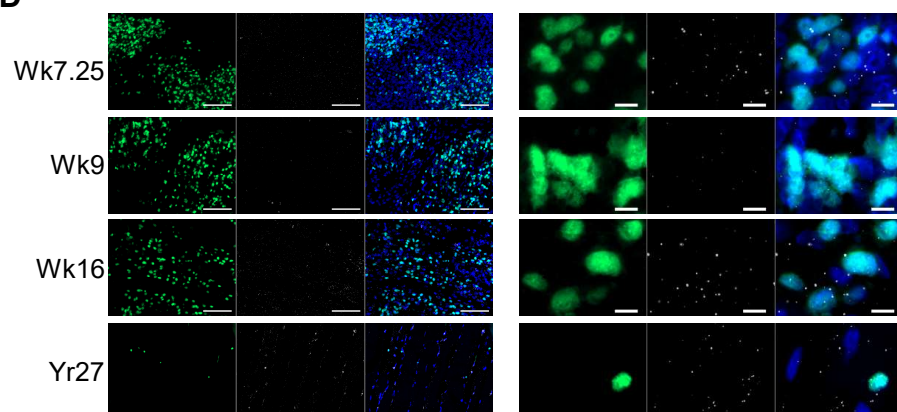
A**B****C****D**

Figure S7. Related to Figure 7. Validation of expression of TFs differentially expressed across human developmental stages.

Tissue sections from different human developmental stages were examined by IHC for PAX7 and RNAscope for indicated TFs: *NFIX* (A), *NFIC* (B), *KLF9* (C) and *CEBPD* (D). TF transcripts (RNAscope signal positive dots; grey) were quantified within PAX7 positive cells (green). Left panels: low magnification images (scalebars: 100 μ m). Middle panels: zoomed in areas (scalebars: 10 μ m). Right panels: box and whisker plots showing quantification of dots per PAX7⁺ cell from different stage samples. n denotes number of PAX7⁺ cells quantified. Two-tailed unpaired Student's *t*-tests were performed between each sample stages followed by Bonferroni multiple test correction. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

Table S1. Related to Figure 1. Information of *in vivo* human samples used for scRNA-seq.

ID	Age	Carnegie Stage	Tissue Type	Sorting/Enrichment Strategy	Cells Sequenced	Mean nGene	Mean nUMI	Mean Mapped Exon Reads
AP32	embryonic week 5	CS14	whole hindlimb	none	815	1835	3276	5215
AP37	embryonic week 5	CS14	whole hindlimb	none	7352	1512	2728	7501
AP5	embryonic week 6	CS16	whole hindlimb	none	2253	1511	2588	4381
AP13	embryonic week 6	CS16	whole hindlimb	none	2609	1585	2770	5387
AP22	embryonic week 6.5	CS17	whole hindlimb	none	1866	1575	2662	4941
AP6	embryonic week 7	CS18	whole hindlimb	none	3779	1321	2216	4605
AP24	embryonic week 7	CS18	whole hindlimb	none	856	1542	2724	5011
AP30	embryonic week 7	CS18	whole hindlimb	none	4229	1504	2651	7988
AP29	embryonic week 7.25	CS19	whole hindlimb	none	7532	1327	2238	9136
AP38	embryonic week 7.25	CS19	whole hindlimb	none	5108	989	1659	5767
AP10	embryonic week 7.75	CS21	whole hindlimb	none	1227	1606	2716	6254
AP14	fetal week 9	not applicable	whole hindlimb	none	4127	1496	2850	12251
AP36	fetal week 9	not applicable	whole hindlimb	none	5579	1371	2623	7181
AP23	fetal week 12	not applicable	hindlimb muscles	live and CD235a ⁺ /CD31 ⁻ /CD45 ⁻	939	1346	2401	4260
AP25	fetal week 14	not applicable	hindlimb muscles	live and CD235a ⁺ /CD31 ⁻ /CD45 ⁻	606	1276	2227	4971
AP17	fetal week 17	not applicable	hindlimb muscles	live and CD235a ⁺ /CD31 ⁻ /CD45 ⁻	3350	1260	2323	5660
AP16	fetal week 18	not applicable	hindlimb muscles	live and CD235a ⁺ /CD31 ⁻ /CD45 ⁻	3120	1326	2387	5817
AP11	juvenile year 7	not applicable	gastrocnemius	live and CD235a ⁺ /CD31 ⁻ /CD45 ⁻	1351	733	1140	5680
AP34	juvenile year 11	not applicable	quadriceps	live and CD235a ⁺ /CD31 ⁻ /CD45 ⁻	631	639	1058	4087
AP20	adult year 34	not applicable	quadriceps	live and CD235a ⁺ /CD31 ⁻ /CD45 ⁻	2259	496	836	6295
AP28	adult year 42	not applicable	quadriceps	live and CD235a ⁺ /CD31 ⁻ /CD45 ⁻	502	556	982	5223

Table S4. Related to Figure 5. Information of *in vitro* hPSC differentiation samples used for scRNA-seq.

ID	Length of Differentiation	Protocol	Sorting/Enrichment Strategy	Cells Sequenced	Mean nGene	Mean nUMI	Mean Mapped Exon Reads
HX5	20 days (3 weeks)	Xi H, et al; 2017; Cell Rep	live and PAX7-GFP ⁺	491	1655	3039	11730
HX16	21 days (3 weeks)	Xi H, et al; 2017; Cell Rep	live sorted	2533			
HX7	27 days (4 weeks)	Xi H, et al; 2017; Cell Rep	live and PAX7-GFP ⁺	322	1588	2868	9319
HX8	27 days (4 weeks)	Xi H, et al; 2017; Cell Rep	live sorted	1240	1842	3399	6459
HX19	35 days (5 weeks)	Xi H, et al; 2017; Cell Rep	live and PAX7-GFP ⁺	1387	1859	4470	12813
HX20	35 days (5 weeks)	Xi H, et al; 2017; Cell Rep	live and ERBB3 ⁺ /NGFR ⁺ /HNK1 ⁻	2031	1803	3883	8909
HX21	35 days (5 weeks)	Xi H, et al; 2017; Cell Rep	live sorted	3652	1967	4508	13844
HX9	41 days (6 weeks)	Xi H, et al; 2017; Cell Rep	live and PAX7-GFP ⁺	877	1899	3808	6209
HX10	41 days (6 weeks)	Xi H, et al; 2017; Cell Rep	live sorted	2198	1591	3042	4986
HX11	55 days (8 weeks)	Xi H, et al; 2017; Cell Rep	live and PAX7-GFP ⁺	1170	1667	3368	7196
HX12	55 days (8 weeks)	Xi H, et al; 2017; Cell Rep	live sorted	2344	1571	3187	5768
JC1	34 days (5 weeks)	Chal J, et al; 2015; Nat Biotechnol	live and PAX7-GFP ⁺	2526	1969	4404	13400
JC2	34 days (5 weeks)	Chal J, et al; 2015; Nat Biotechnol	live and ERBB3 ⁺ /NGFR ⁺ /HNK1 ⁻	1951	1854	4363	13905
JC3	34 days (5 weeks)	Chal J, et al; 2015; Nat Biotechnol	live sorted	3531	1412	2587	6145
MS2	48 days (7 weeks)	Shelton M, et al; 2014; Stem Cell Reports	live and PAX7-GFP ⁺	1390	1616	4102	18194
MS3	48 days (7 weeks)	Shelton M, et al; 2014; Stem Cell Reports	live and ERBB3 ⁺ /NGFR ⁺ /HNK1 ⁻	1827	1164	2328	8337
MS5	43 days (6 weeks)	Shelton M, et al; 2014; Stem Cell Reports	live sorted	5027	1212	2417	6438

Methods S1. Related to Figure 3. Primer pairs for qRT-PCR.

Gene Name	NCBI Gene ID	Primer Sequences	
		Forward	Reverse
<i>BGLAP</i>	632	CACTCCTCGCCCTATTGGC	CCCTCCTGCTTGGACACAAAG
<i>CKM</i>	1158	CTGACAAGCACAAGACTGACC	CTGCTGAGCACGTAGTTAGGG
<i>DCN</i>	1634	ATGAAGGCCACTATCATCCTCC	GTCGCGGTCATCAGGAACTT
<i>eGFP</i>	20473140	GGCAAGCTGACCCTGAAGTT	CTTCATGTGGTCGGGGTAGC
<i>IBSP</i>	3381	CACTGGAGCCAATGCAGAAGA	TGGTGGGGTTGTAGGTTCAA
<i>MYH8</i>	4626	AATGCAAGTGCTATTCCAGAGG	ACAGACAGCTTGTGTTCTTGTT
<i>OGN</i>	4969	CTACTTGGACCATAATGCCCTG	GTCCCGGATGTAAGTGGTGTC
<i>RPL13A</i>	23521	GAAGTACCTGGCTTTCCTCCG	TGGTTTTGTGGGGCAGCATA