

Supplementary Materials for
**Integrative molecular roadmap for direct conversion of fibroblasts into
myocytes and myogenic progenitor cells**

Inseon Kim, Adhideb Ghosh, Nicola Bundschuh, Laura Hinte, Eduard Petrosyan,
Ferdinand von Meyenn, Ori Bar-Nur*

*Corresponding author. Email: ori.bar-nur@hest.ethz.ch

Published 6 April 2022, *Sci. Adv.* **8**, eabj4928 (2022)
DOI: [10.1126/sciadv.abj4928](https://doi.org/10.1126/sciadv.abj4928)

This PDF file includes:

Figs. S1 to S10
Table S1

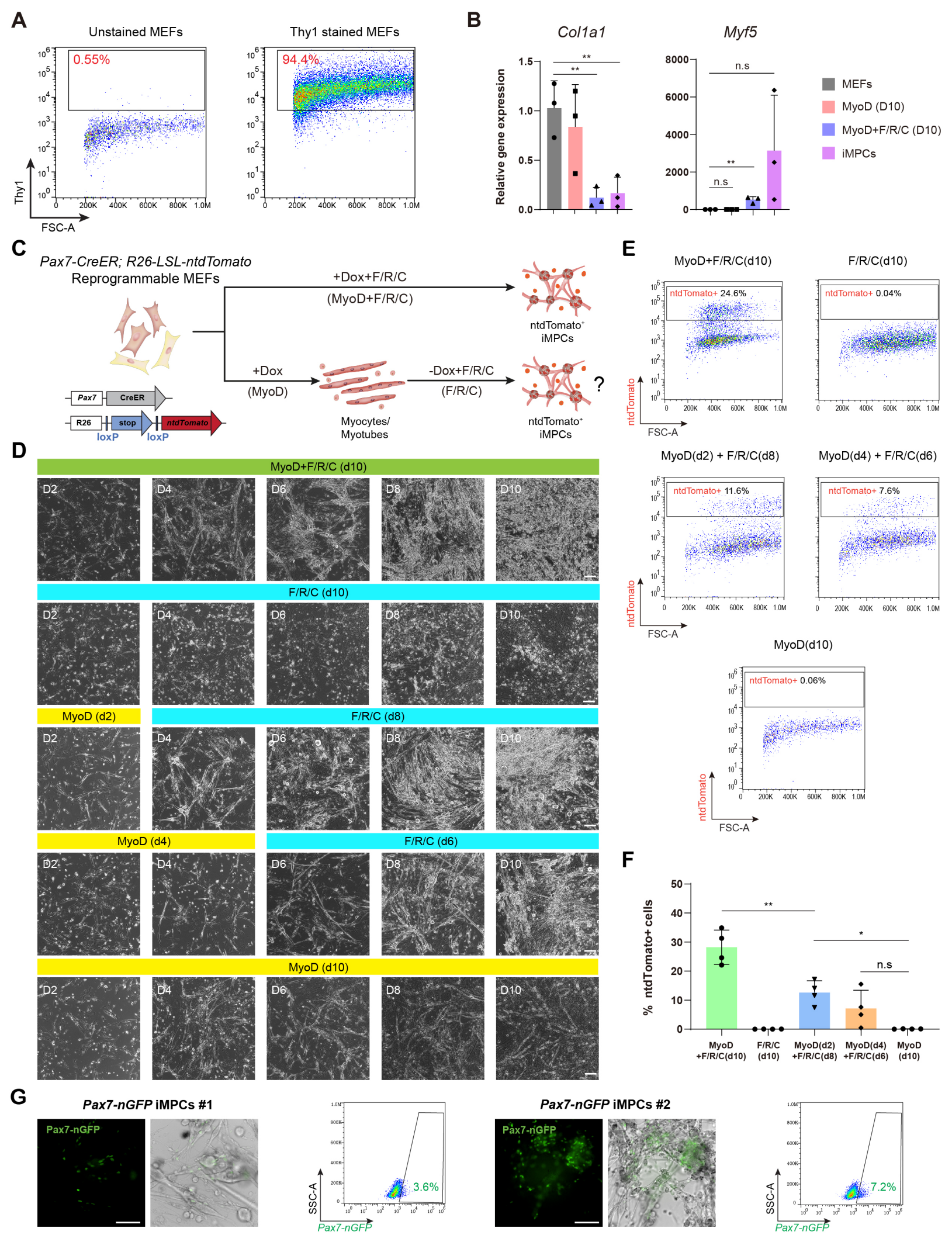


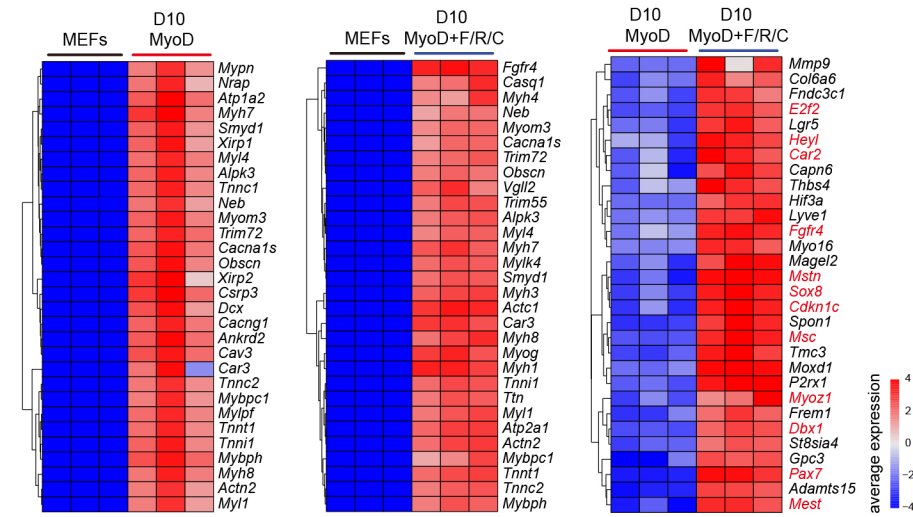
Fig. S1: MyoD and small molecules convert fibroblasts into iMPCs

(A) Flow cytometry analysis of MEF cultures stained for the fibroblast-specific surface marker Thy1. MEFs, mouse embryonic fibroblasts. **(B)** qRT-PCR analysis for *Col1a1* and *Myf5*. Data is shown as mean \pm S.D. N=3 cell lines per each group. Statistical significance was determined by a two-tailed unpaired *t*-test (** $p < 0.01$, n.s.=non-significant). **(C)** Experimental design to investigate whether transient MyoD overexpression followed by F/R/C supplementation can reprogram cells into iMPCs. **(D)** Representative bright-field images of *Pax7-CreER; Rosa26-LoxSTOPLox(LSL)-ntdTomato* MEFs subjected to the indicated conditions. Scale bar, 100 μ m. **(E)** Flow cytometry analysis for ntdTomato⁺ cells in *Pax7-CreER; Rosa26-LSL-ntdTomato* MEFs subjected to the indicated reprogramming conditions. A 4-Hydroxytamoxifen (4OHT) was added to the cells two days before the FACS analysis. **(F)** Bar graph showing the percentage of ntdTomato⁺ cells for the indicated conditions as demonstrated in Fig. S1E. The data is shown as mean \pm S.D. N=2 experimental repeats for 2 cell lines per each group. Statistical significance was determined by one-way ANOVA between conditions (* $p < 0.05$, ** $p < 0.01$, n.s.=non-significant). **(G)** Representative images and flow cytometry analysis of different *Pax7-nGFP* iMPC clones at passage 1. Note nuclear GFP expression only in mononucleated cells. Scale bar, 100 μ m.

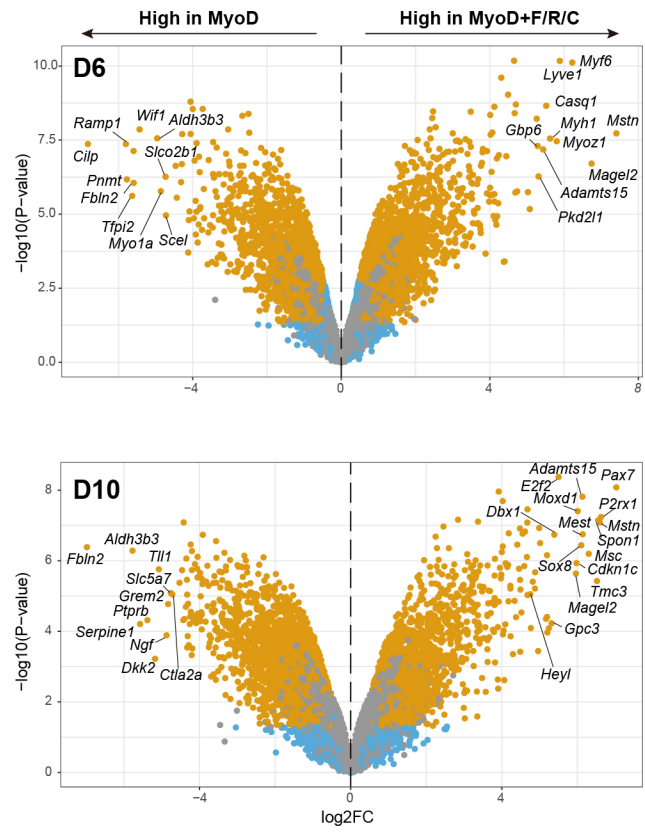
Fig. S2: Gene expression changes during myogenic lineage conversions

(A) Dendrogram clustering based on bulk RNA-Seq data using all gene read counts for the indicated samples. N = 3 cell lines per each group. **(B)** Gene expression analysis based on bulk RNA-Seq for the listed genes in MEFs subjected to either MyoD or MyoD+F/R/C treatment. Established iMPCs and primary myoblasts served as positive controls. Relative gene expression was calculated by normalizing the RPKM values of each sample to that of parental MEFs (D0 MEFs). The data is shown as mean \pm S.D. N = 3 cell lines per each group. Statistical significance was determined by two-way ANOVA between conditions at each time point (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$). **(C)** Representative Integrative Genomic Viewer (IGV) tracks for the indicated genes, conditions and time points. **(D)** PCA based on bulk RNA-seq data for this study in comparison to Yagi et al., 2021 using log normalized gene counts. **(E)** Gene expression analysis during a reprogramming time course for the indicated genes as performed by Yagi et al., 2021 and this study. **(F)** Enriched network analysis using the STRING database for the genes shown in gene set 2 (MyoD+F/R/C treatment, Fig. 2E) that annotated with *Regulation of cell cycle* (left) and *Striated muscle tissue development* (right).

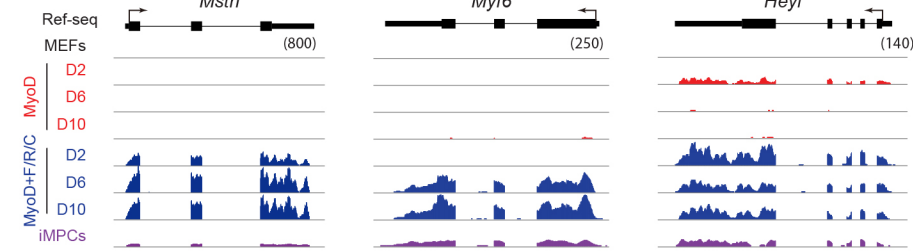
A Top 30 upregulated genes



B



C



D

Upregulated genes vs. MEFs ($\log_2FC > 1$, $p\text{-value} < 0.01$)

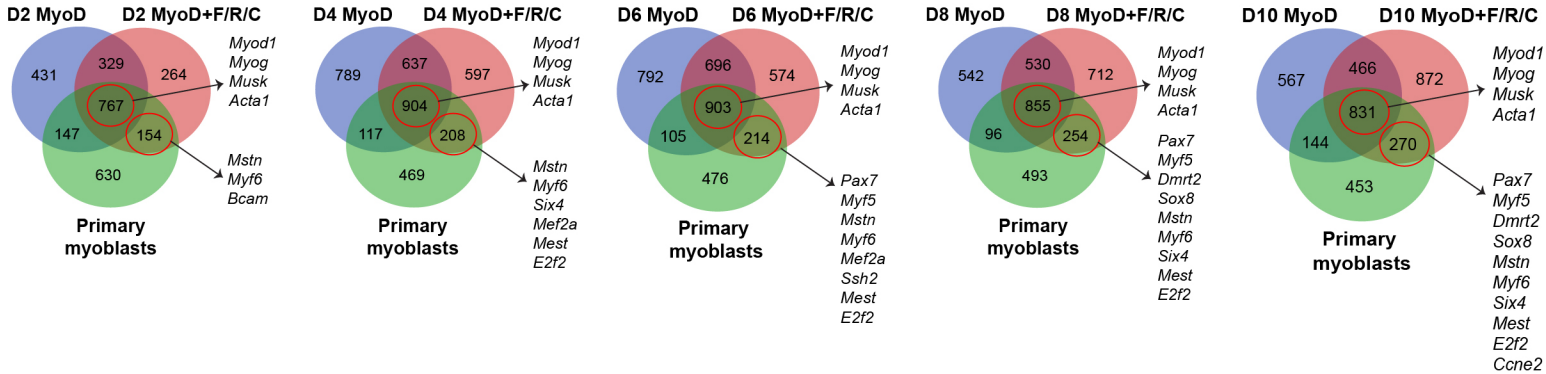
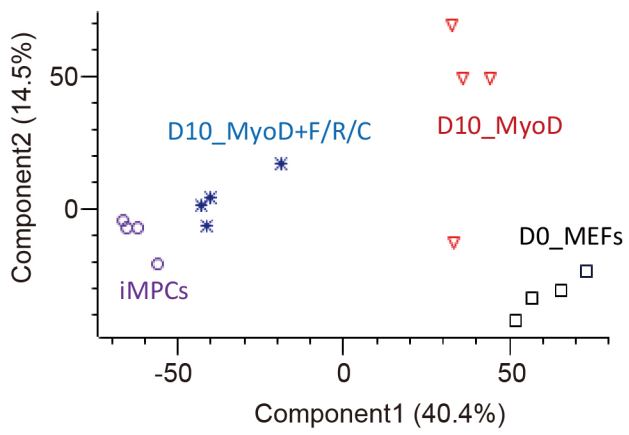
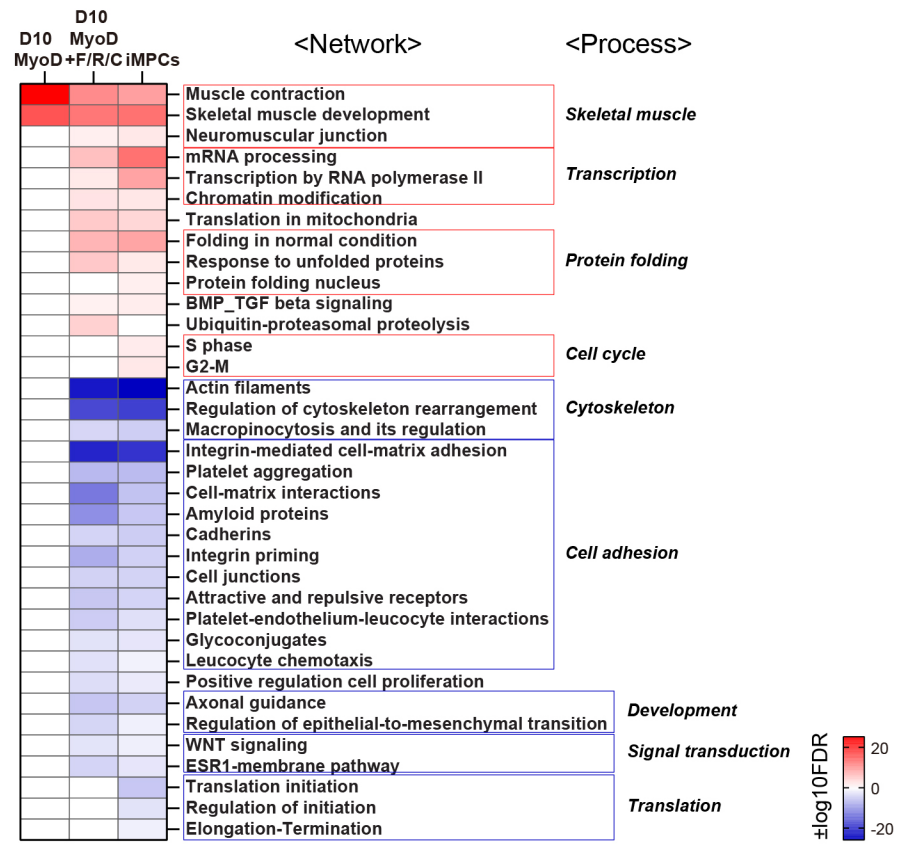


Fig. S3: Transcriptional differences between myogenic cell conversions

(A) Heatmaps based on bulk RNA-Seq data displaying the top 30 upregulated genes for the indicated cell lines. The average expression is presented as gradient. N = 3 cell lines per group. Unique myogenic genes are highlighted in red color. **(B)** Volcano plots based on bulk RNA-Seq data for differentially expressed genes (DEGs) between MyoD and MyoD+F/R/C conditions at days 6 and 10. Significant DEGs ($|\log_2FC| > 0.5$, $p\text{-value} < 0.05$) are shown as yellow dots. N = 3 cell lines per each group. **(C)** IGV tracks for selected myogenic genes. **(D)** Venn diagrams based on bulk RNA-Seq data showing the overlap of upregulated genes in primary myoblasts and MEFs subjected to either MyoD or MyoD+F/R/C conditions at each time point vs. parental MEFs ($\log_2FC > 1$, $p\text{-value} < 0.01$). N = 3 cell lines per each group.

A**B**

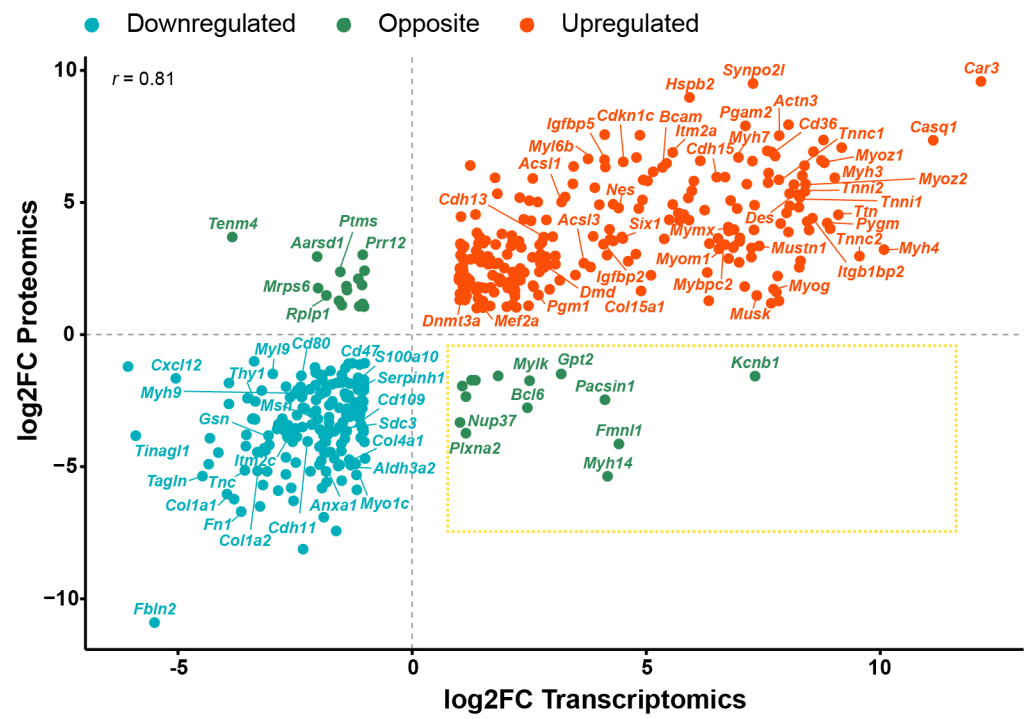
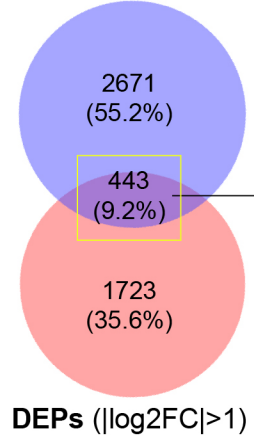
Enrichment analysis (Process networks)

**C**

Transcriptome/Proteome correlation

iMPCs vs. MEFs

DEGs ($|\log_2 \text{FC}| > 1$)

**D**

Anti-correlated markers (iMPCs vs. MEFs)

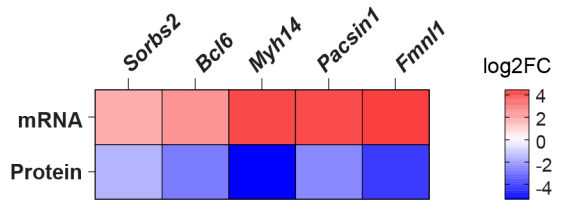


Fig. S4: Proteomic analysis during myogenic lineage conversions

(A) PCA of global protein expression based on LC-MS data across all samples. N = 4 cell lines per each group. **(B)** Enriched process networks of DEPs in each condition vs. MEFs ($|\log_2FC| > 1$, $p\text{-value} < 0.05$) as analyzed via Metacore. Only significantly enriched process networks are presented ($FDR < 0.05$). Upregulated (Red) and downregulated (Blue) process networks are shown using $-\log_{10}(FDR)$ and $\log_{10}(FDR)$, respectively. **(C)** Left - A Venn diagram based on transcriptome and proteome datasets showing the number of significant overlapping DEGs and DEPs in iMPCs vs. MEFs ($|\log_2FC| > 1$, $p\text{-value} < 0.05$). Right - scatterplot showing the correlation between the transcriptome and proteome datasets of iMPCs vs. MEFs. In total, 443 overlapping DEGs / DEPs are projected on the plot, corresponding to the Venn diagram analysis shown on the left. Red and blue dots indicate upregulated and downregulated genes or proteins, respectively. **(D)** Heatmap showing mRNA and protein expression levels (\log_2FC) between iMPCs and MEFs for the indicated genes. \log_2FC is presented as a color gradient. Note an anti-correlation trend between gene and protein expression.

D2 MyoD+F/R/C vs. MEFs

● Downregulated ● Opposite ● Upregulated

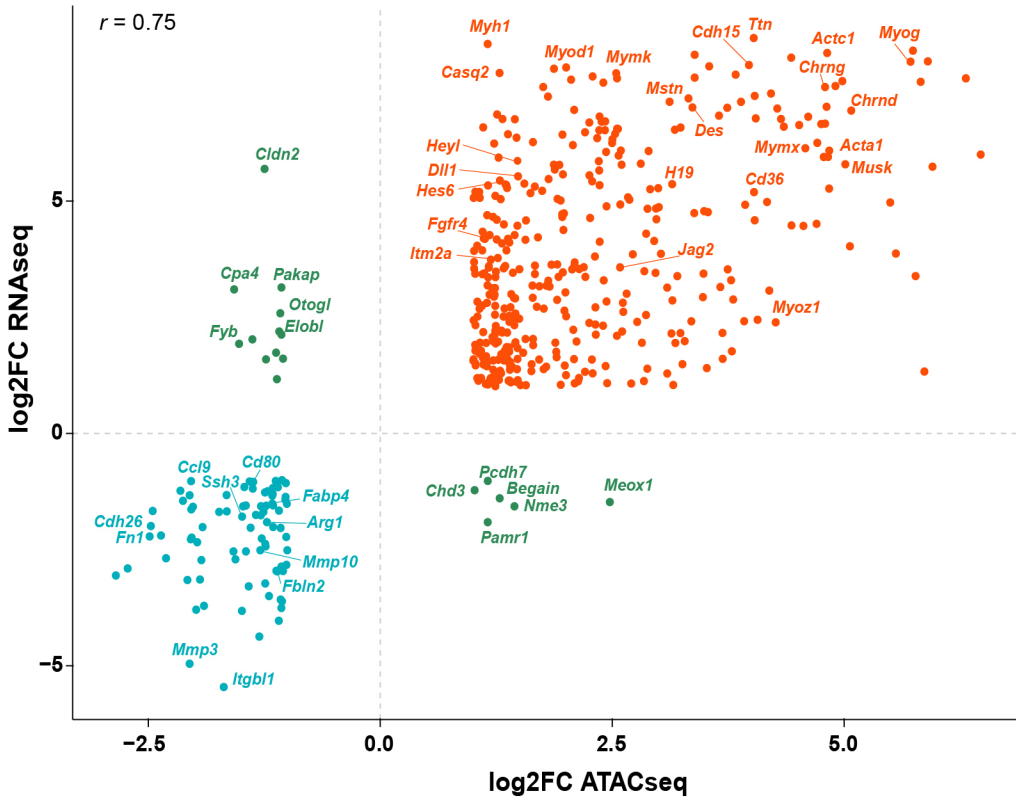


Fig. S5: Integrative gene expression and ATAC-Seq analysis

Scatterplot for overlapping genes showing the correlation between chromatin accessibility in promoter regions (± 1 kb of TSS) and gene expression in day 2 of MyoD+F/R/C condition vs. MEFs. The significant genes (p-value < 0.01) with $|\log_2FC| > 0.5$ are shown. Note rare anti-correlation trend between gene expression and chromatin accessibility for a few markers.

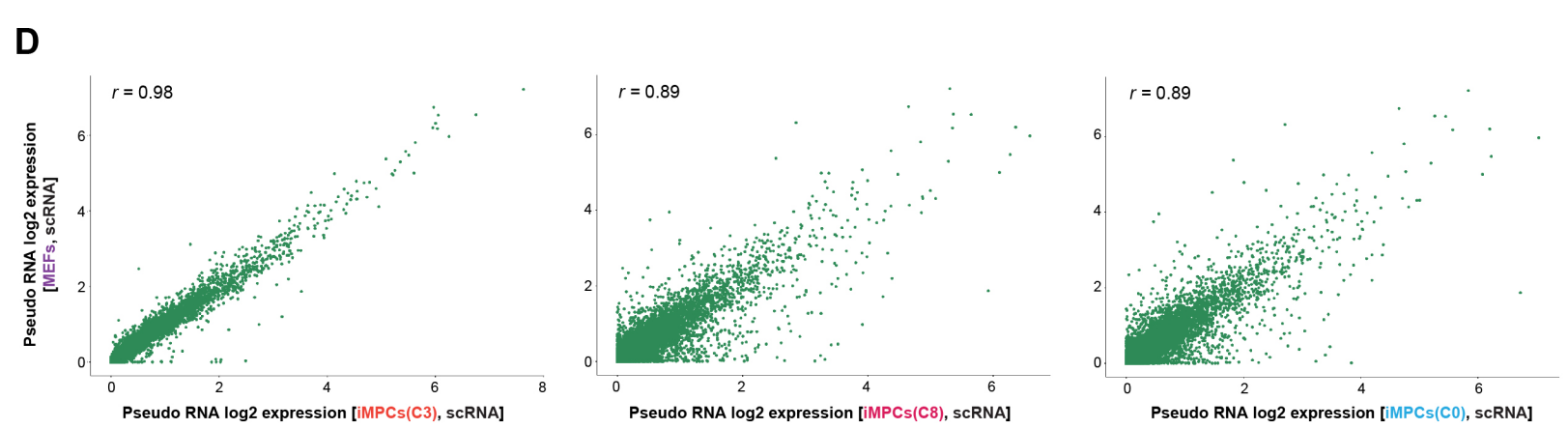
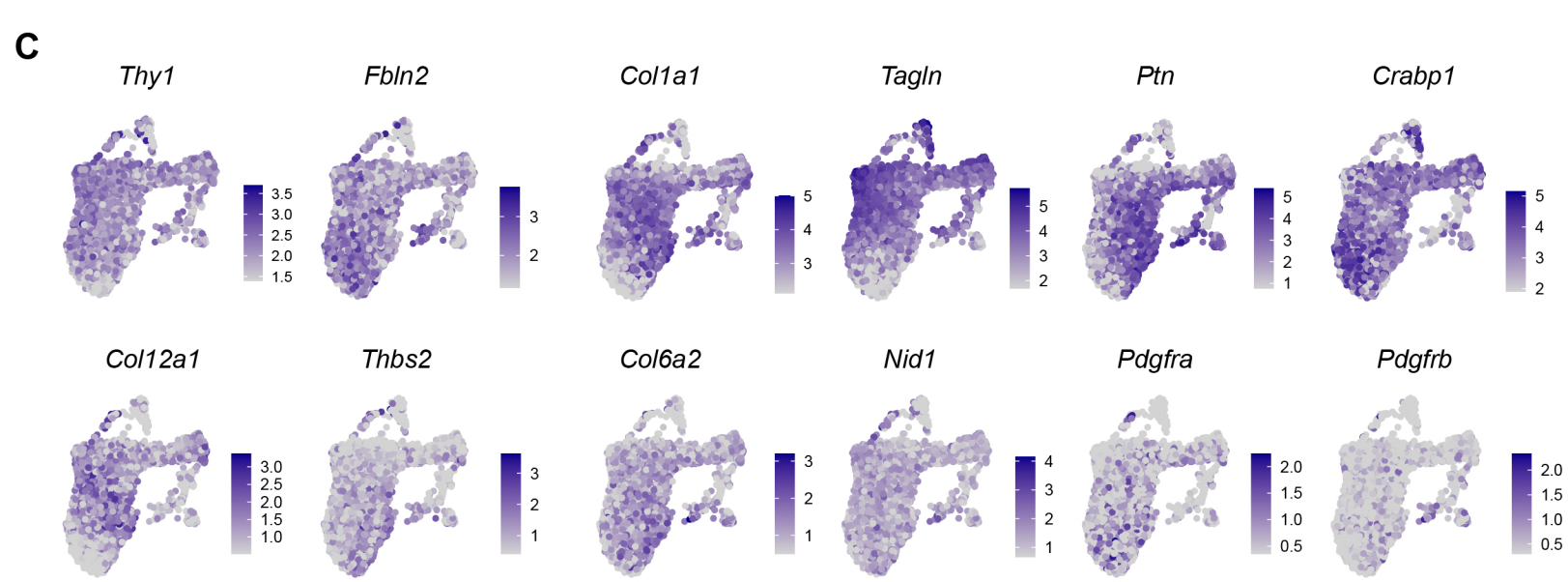
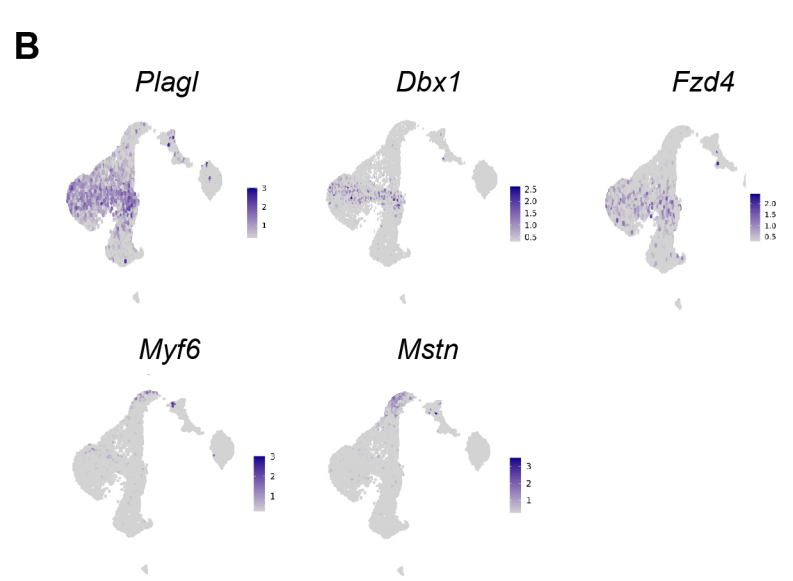
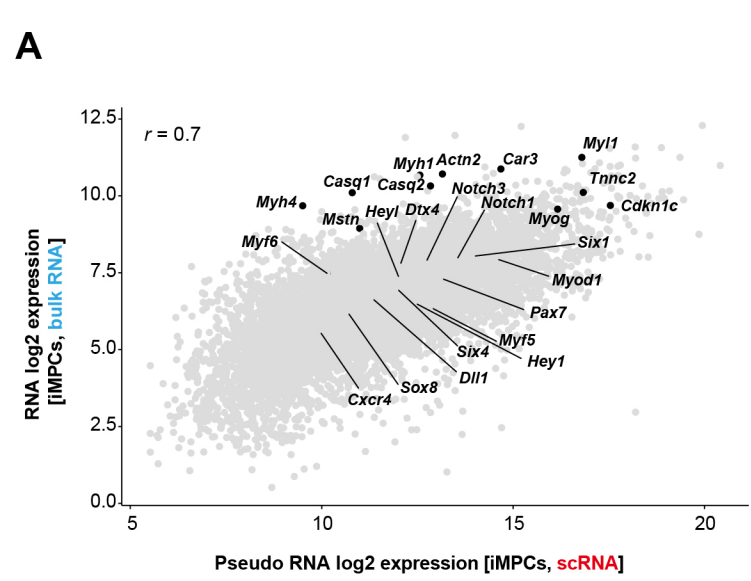


Fig. S6: scRNA-Seq analysis of iMPCs and MEFs

(A) Scatter plot displaying gene expression correlation profile between scRNA-Seq and bulk RNA-Seq datasets for an iMPC clone. Pseudo bulk mRNA expression was computed from scRNA-Seq data. **(B)** UMAP projection showing all cells colored by the indicated myogenic markers based on scRNA-Seq data of a stable iMPC clone. **(C)** UMAP projection showing all cells colored by the indicated fibroblast markers based on scRNA-Seq data of MEFs. **(D)** Scatter plot displaying pseudo bulk RNA expression correlation profile between scRNA-Seq data of MEFs and each of the indicated connective tissue-like cell clusters present in iMPCs including C3 (left), C8 (middle) and C0 (right). The cluster identities correspond to the UMAP projection shown in Fig. 5A.

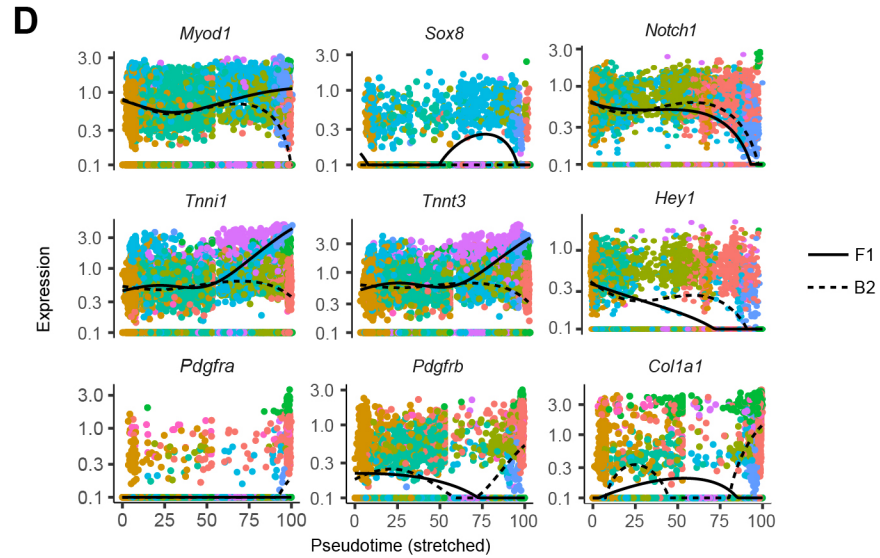
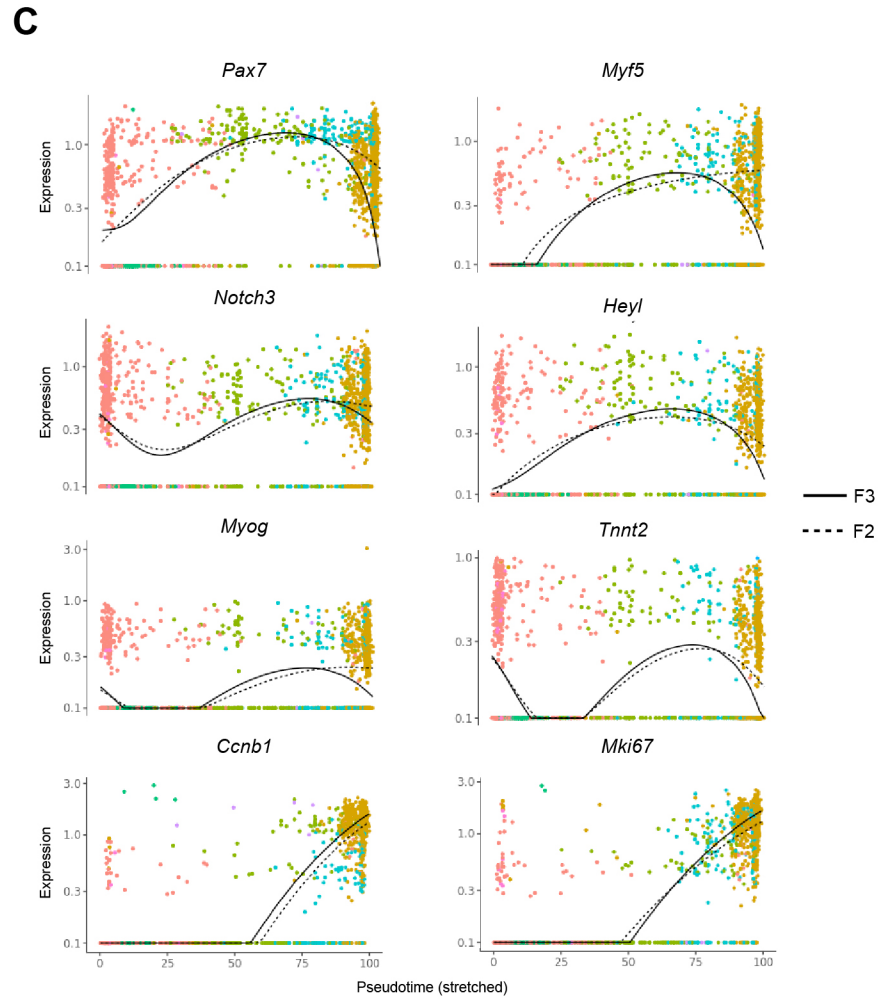
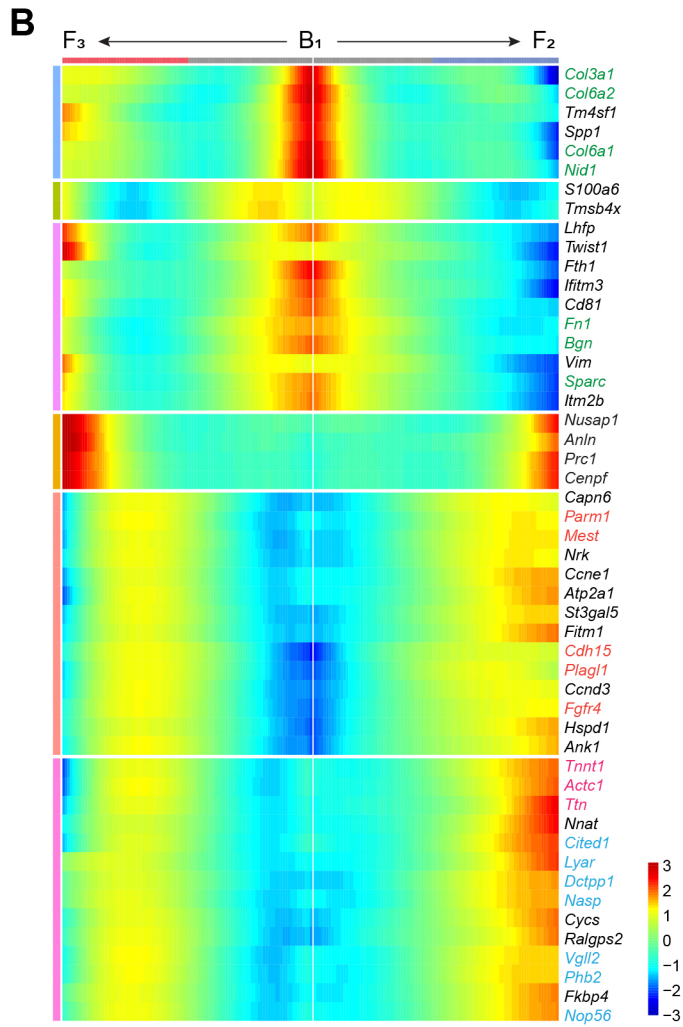
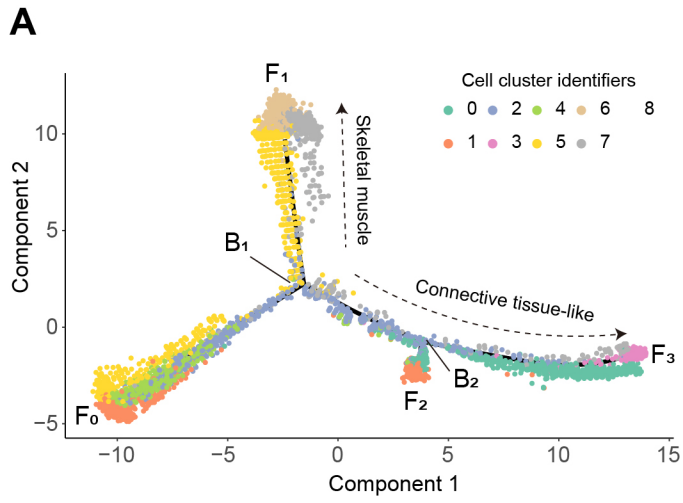


Fig. S7: Pseudotime trajectory analysis of an iMPC clone

(A) Minimum spanning tree showing ordered cells based on semi-supervised single cell trajectory analysis reconstructed by Monocle2 and colored by cell cluster identifiers. **(B)** Heatmap based on scRNA-Seq data of iMPCs for the top 50 DEGs regulated at branch point B_2 in Fig. 5F. The individual gene expression level initiating at point B_1 and proceeding towards F_3 is shown on the left, whereas for F_2 it is shown on the right. Color-code gradient denotes normalized gene expression level (Z-score) for each gene across all cells. **(C)** Plots showing the expression kinetics of the indicated genes regulated at B_2 as a function of pseudotime and proceeding towards F_2 or F_3 . Dots indicate cells colored by cell identifiers. **(D)** Plots showing the gene expression dynamics of the indicated genes regulated at B_1 as a function of pseudotime and proceeding towards F_1 or B_2 . The dots indicate cells colored by cell identifiers.

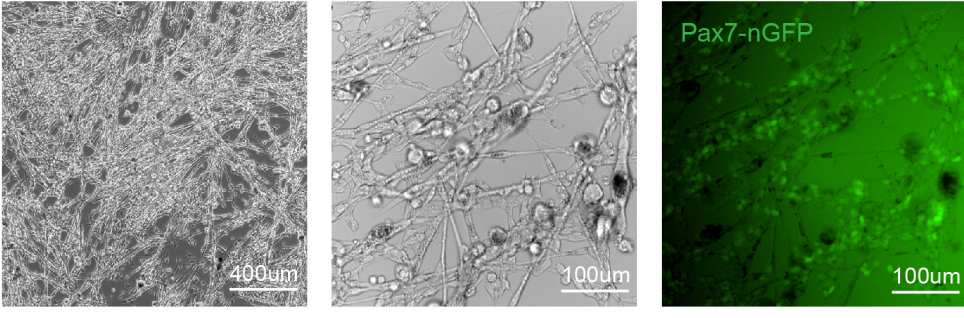
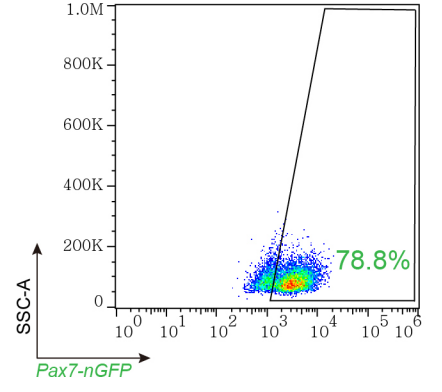
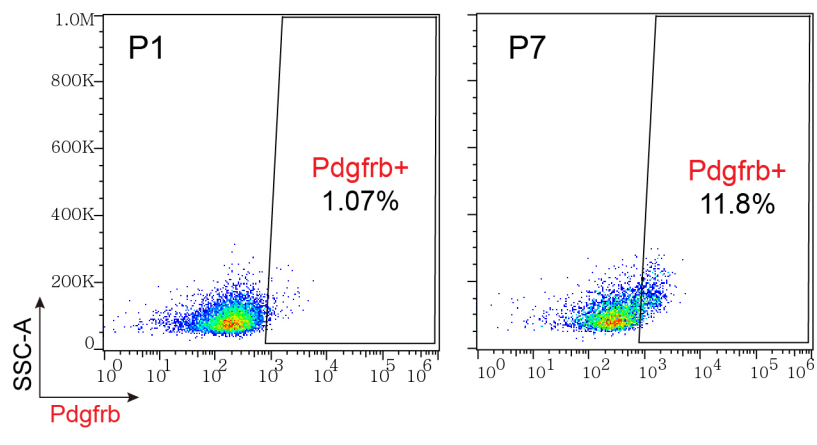
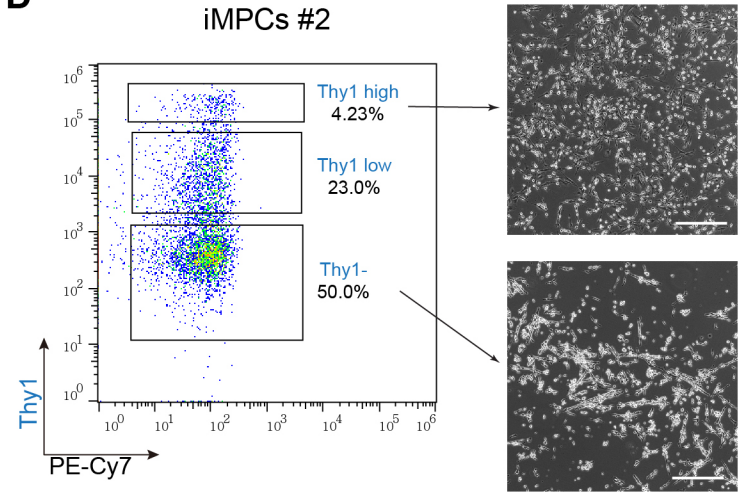
A**B****C****D**

Fig. S8: Formation of connective tissue-like cells in FACS-purified iMPCs

(A) Representative images of an iMPC clone which was derived from Pax7-nGFP⁺ cells which were FACS-purified from an iMPC clone at passage 1. Note that Pax7-nGFP⁺ cells can give rise to the entire heterogeneous iMPC culture. **(B)** Flow cytometry analysis for Pax7-GFP expression in an iMPC clone established from FACS-purified Pax7-nGFP⁺ cells. **(C)** Flow cytometry analysis for Pdgfrb⁺ cells at the respective passages in a stable iMPC clone established from FACS-purified Pax7-nGFP⁺ cells. **(D)** Left - flow cytometry analysis for Thy1 expression in a stable iMPC clone. Right - representative bright-field images of cell cultures derived from FACS-purified Thy1 positive cells (top) and Thy1 negative cells (bottom) two days after FACS-purification. Scale bar, 400 μ m.

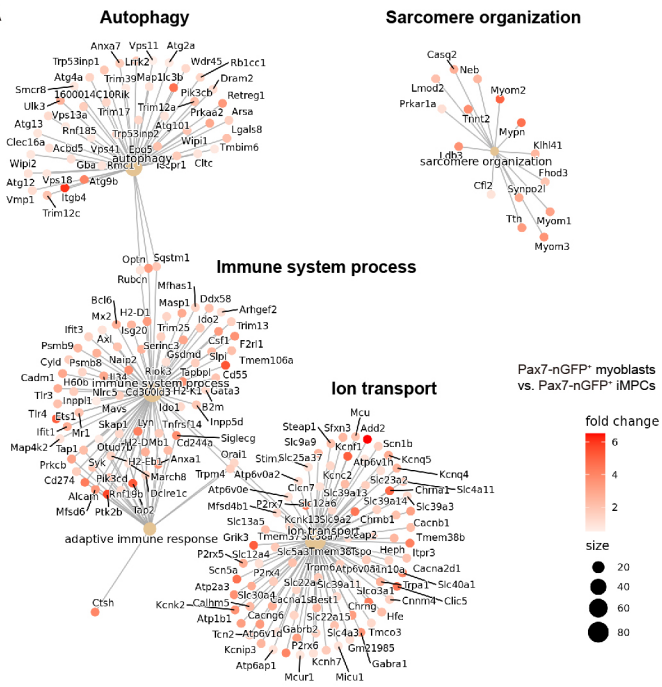
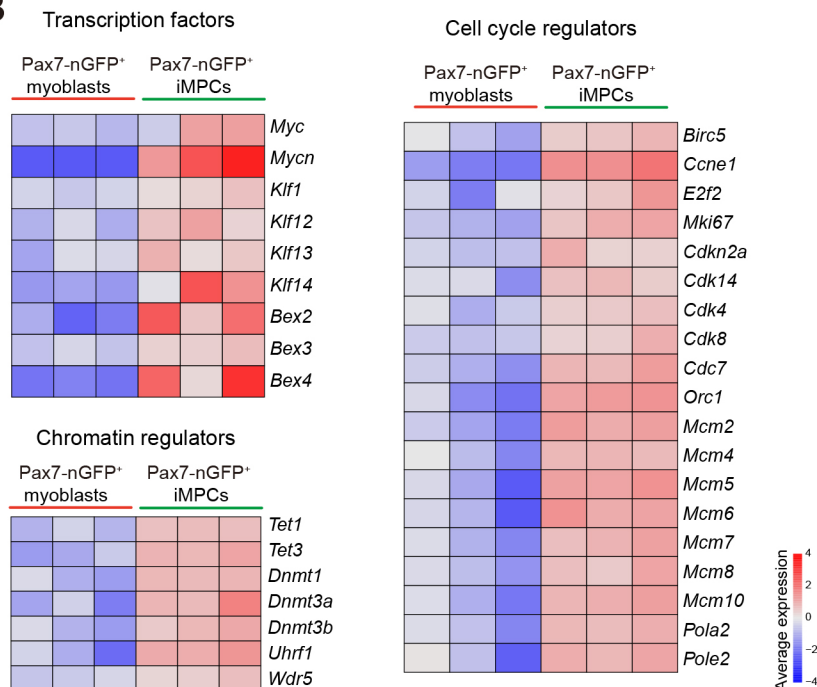
A**B**

Fig. S9: Molecular analysis of Pax7⁺ iMPCs and Pax7⁺ primary myoblasts

(A) Over Representation Analysis (ORA) of the gene sets that were significantly enriched ($\log_2FC > 0.5$, p-value < 0.01) in Pax7-nGFP⁺ myoblasts in comparison to Pax7-nGFP⁺ iMPCs. Only significant GO biological processes (adj. p-value < 0.05) are shown. The size indicates the number of genes involved in the biological process and the color-coding scale represents the fold change between the two conditions. **(B)** Heatmaps showing average expression of selected markers associated with the indicated gene families in Pax7-nGFP⁺ iMPCs in comparison to Pax7-nGFP⁺ myoblasts. Gradient of high to low expression of each gene relative to the average expression in each comparison is indicated by red to blue. N = 3 cell lines per each group.

A ● MyoD ■ MyoD+F/R/C ▲ iMPCs ▼ Pax7-nGFP⁺ myoblasts

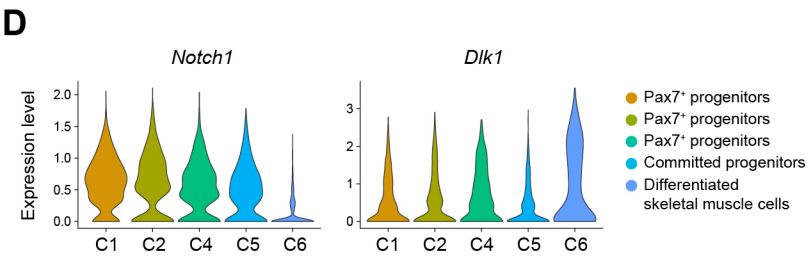
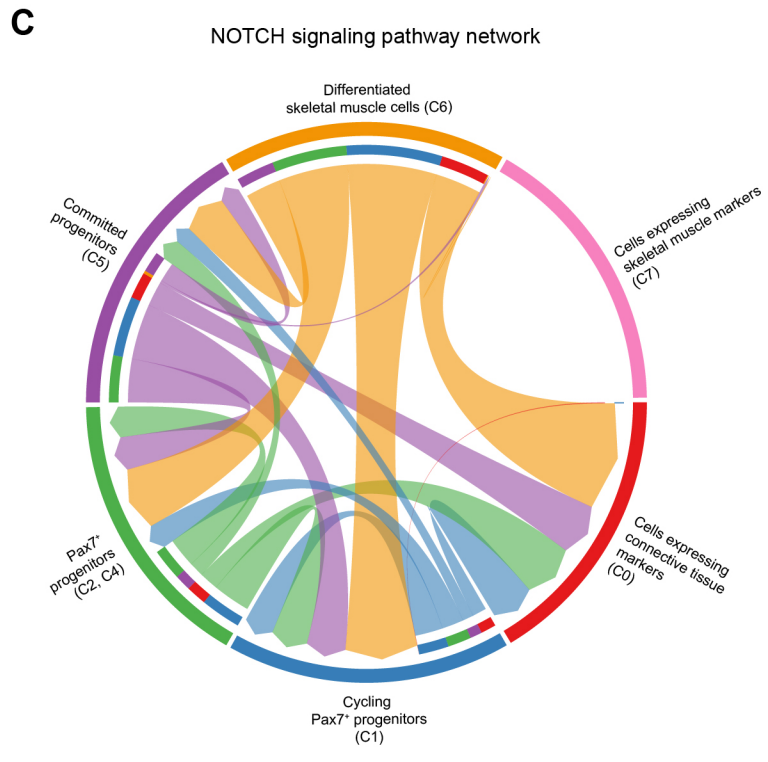
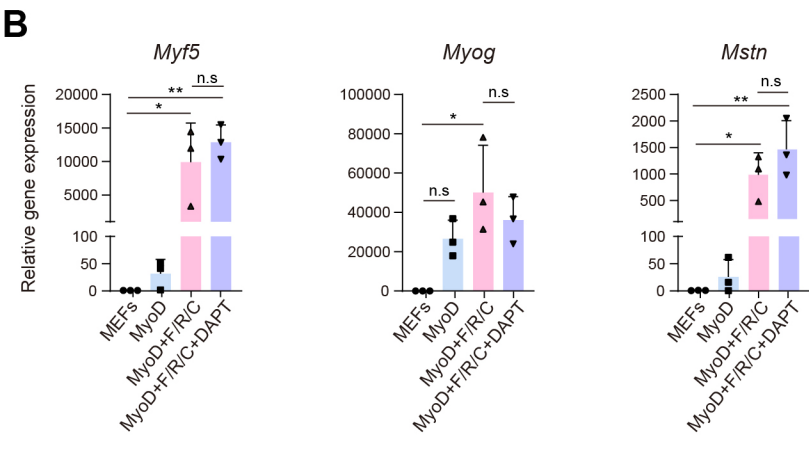
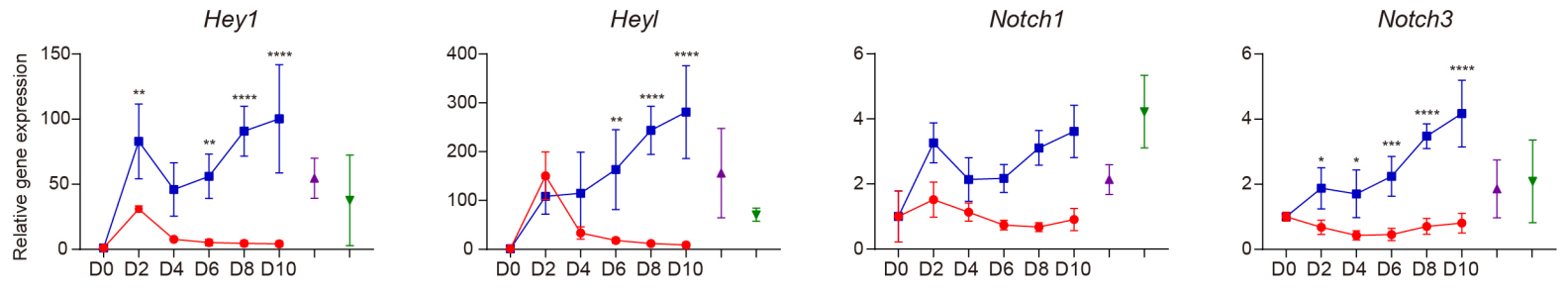


Fig. S10: Notch signaling pathway involvement during iMPC formation

(A) Expression dynamics based on bulk RNA-Seq data for the indicated Notch pathway associated genes during either MyoD or MyoD+F/R/C lineage conversions. Established iMPCs and Pax7-nGFP⁺ myoblasts served as positive controls. Relative gene expression was calculated by normalizing the RPKM values of each sample to that of parental MEFs (D0 MEFs). N = 3 cell lines per group. Statistical significance was determined by two-way ANOVA between the conditions at each respective time point (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001). **(B)** Relative qRT-PCR gene expression analysis for the indicated myogenic genes at day 10 of MEF reprogramming using the indicated conditions. The data is shown as mean ± S.D. N = 3 cell lines per each group. Statistical significance was determined by one-way ANOVA (*p<0.05, **p<0.01, n.s.=non-significant). **(C)** Chord diagram showing cell-cell interactions for the indicated cell types based on scRNA-Seq of a stable iMPC clone and using the Notch signaling pathway receptor-ligand network as reference. **(D)** Violin plots showing the average expression of *Notch1* and *Dlk1* in the respective cell clusters based on scRNA-Seq data of iMPCs. The clusters correspond to the cell populations shown in Fig. 5A.

Table S1. Sequences of the primers used for qRT-PCR

Gene	Direction	Sequence 5' -> 3'
<i>Pax7</i>	Forward	GACGACGAGGAAGGAGACAA
	Reverse	ACATCTGAGCCCTCATCCAG
<i>Myf5</i>	Forward	AAGGCTCCTGTATCCCCTCAC
	Reverse	TGACCTTCTTCAGGCGTCTAC
<i>Myod1</i>	Forward	CCACTCCGGGACATAGACTTG
	Reverse	AAAAGCGCAGGTCTGGTGAG
<i>Myog</i>	Forward	GAGACATCCCCCTATTTCTACCA
	Reverse	GCTCAGTCCGCTCATAGCC
<i>Myf6</i>	Forward	AGATCGTCGGAAAGCAGC
	Reverse	CCTGGAATGATCCGAAACAC
<i>MyHC</i>	Forward	GCGAATCGAGGCTCAGAACAA
	Reverse	GTAGTTCCGCCTTCGGTCTTG
<i>Pgk</i>	Forward	ATGTCGCTTTCCAACAAGCTG
	Reverse	GCTCCATTGTCCAAGCAGAAT
<i>Thy1</i>	Forward	CCT TAC CCT AGC CAA CTT CAC
	Reverse	AGG ATG TGT TCT GAA CCA GC
<i>Col1a1</i>	Forward	GCA ACA GTC GCT TCA CCT AC
	Reverse	GTG GGA GGG AAC CAG ATT G
<i>Notch1</i>	Forward	CAGGAAAGAGGGCATCAG
	Reverse	AGCGTTAGGCAGAGCAAG
<i>Notch3</i>	Forward	GTCCAGAGGCCAAGAGACTG
	Reverse	CAGAAGGAGGCCAGCATAAG
<i>Hey1</i>	Forward	TGAATCCAGATGACCAGCTACTGT
	Reverse	TACTTTTCAGACTCCGATCGCTTAC
<i>Heyl</i>	Forward	CAGATGCAAGCCCGGAAGAA
	Reverse	ACCAGAGGCATGGAGCATCT
<i>Mstn</i>	Forward	CAGGAGAAGATGGGCTGAATC
	Reverse	GGAGTGCTCATCGCAGTCAA