

**Transcription factor FOXF1 identifies compartmentally distinct mesenchymal cells with a
role in lung allograft fibrogenesis**

Russell R. Braeuer¹, Natalie M. Walker¹, Keizo Misumi¹, Serina Mazzoni-Putman¹,
Yoshiro Aoki¹, Ruohan Liao², Ragini Vittal¹, Gabriel Kleer¹, David S. Wheeler¹,
Jonathan Z. Sexton³, Carol F. Farver⁴, Joshua D. Welch^{2,5} and Vibha N. Lama¹

Departments of Internal Medicine¹, Computational Medicine and Bioinformatics²,
Pharmacy³, Pathology⁴, and Computer Science and Engineering⁵

University of Michigan, Ann Arbor, MI, USA

Supplemental Material - Methods and Supplemental Figure Legends

Corresponding Address:

Vibha N. Lama, M.D., M.S.

Division of Pulmonary and Critical Care Medicine, University of Michigan

1500 W Medical Center Drive, Ann Arbor, MI, USA 48109-0360

734-936-5010, vlama@umich.edu

Conflict of interest: The authors declare that no conflict of interest exists

Materials and Methods

Transgenic mice: *C57BL/6J-Tg(Foxf1_tdTomato)* (officially *C57BL/6J-Tg(RP23-55G)VLama*) mice were generated through University of Michigan Transgenic Animal Model Core, wherein a *C57BL/6J* mouse genomic clone from a bacterial artificial chromosome (BAC) library that carries the *Foxf1* gene was identified (clone RP23-55G4) and obtained from the BACPAC resource (Oakland, CA). Standard BAC recombineering methods (1) were applied to generate a modified BAC in which exon 1 of the *Foxf1* gene was replaced with the coding sequence for tdTomato followed by a polyadenylation sequence to enable stable expression of the fluorescent protein in cells that normally express *Foxf1*. The final engineered BAC used for transgenic mouse production contained approximately 116,000 base pairs of DNA upstream from the tdTomato start site. Type I collagen-GFP (*Col1GFP*; stock number 013134, Jackson Laboratory) reporter mice are described previously (2). Double transgenic mice expressing type I collagen and FOXF1 were generated by crossing *Col1GFP* with *Foxf1_tdTomato* mice. *Gli1^{CreERT2/WT};Rosa26^{mTmG/WT}* mice were generated by crossing *tdTomato^{fllox}* (B6.129(Cg)-Gt(ROSA)26Sor^{tm4(ACTB-tdTomato,-EGFP)Luo/J}, stock number 007676 with *Gli1^{CreERT2}* (*Gli1^{tm3(cre/ERT2)Alj}*, stock number 007913 backcrossed 10 generations onto *C57BL6/J* background), purchased from Jackson laboratory and are described previously (3). *Gli1^{CreERT2/WT};Rosa26^{EYFP}* mice were generated by crossing *Gli1^{CreERT2}* with B6.129X1-Gt(ROSA)26Sor^{tm1(EYFP)Cos/J} (stock number 006148). B6.129S4-Pdgfra^{tm11(EGFP)Sor/J}, stock number 007669, were acquired from Jackson Laboratory.

Murine lung transplant model and treatments: The *Gli1^{CreERT2/WT};Rosa26^{mTmG/WT}* B6D2F1/J donor mice were dosed with tamoxifen chow for 14 days at 4-6 weeks of age. Mice were given normal chow for 3 days and then transplanted into *C57BL/6J* recipients. Mice were between 6-8 weeks of age for all studies. For the murine orthotopic left lung transplant model that mimics restrictive allograft syndrome, allogeneic transplants were performed utilizing a B6D2F1/J lungs→*C57BL6/J* strain combination with previously

described surgical techniques (4). All experiments were performed according to the protocols approved by the Institutional Animal Care and Use Committee at the University of Michigan.

Immunohistochemistry: For frozen mouse lung tissue sections, lungs were inflated with 50% OCT in PBS and frozen in cryomolds containing OCT. For FOXF1 detection, 5 μ M sections were fixed in paraformaldehyde (10 minutes), while frozen sections stained for TdTomato (Rockland; cat# 600-401-379) required ice-cold acetone fixation for 20 minutes before staining protocol. Formalin-fixed, paraffin-embedded sections required 10 minutes of citrate-based heat retrieval prior to staining for FOXF1 nuclear expression. All slides were blocked with horse serum, followed by quenching for endogenous peroxidase activity (0.3% H₂O₂ in methanol x 10 minutes), avidin/biotin block, and then incubated with anti-FOXF1 (R&D Systems; Cat#AF4798) in PBS overnight. After washes, anti-goat-biotin and ABC incubations (Vectastain Elite ABC Kit; Cat#PK-6100) were performed, followed by tyramide enhancement as per manufacturer's instructions. Slides were then blocked with goat serum and incubated with goat anti-GFP-FITC conjugated antibodies (Abcam; ab6662) in goat serum-containing PBS. Anti- α -SMA antibody conjugated with eFluor660 was used (Invitrogen; 50-976082). Anti-TTF1 (Santa Cruz; sc1304) and Anti-Aquaporin5 (Abcam; ab78486) with the secondary anti-rabbit-APC allowed for triple staining on selected sections. Human lung sections were cut from paraffin blocks and subjected to staining protocol described for mouse lung sections with anti- α -SMA-Cy3 (Millipore Sigma; Cat#C6198) used for smooth muscle actin staining.

Whole lung digestion: For mesenchymal cell isolation, lungs were harvested and digested with collagenase A (37°C x 30 minutes) in serum-free DMEM. Digests were sheared with an 18-gauge needle, incubated (37°C x 10 minutes), sheared again, and then filtered through a 70 μ M filter. Red blood cells were removed using a red blood cell lysis buffer. The reaction was stopped using 10% FBS containing DMEM, and then filtered through a 70 μ M filter. For epithelial cell isolation, lungs were perfused with elastase (1 ml x 10 minutes), followed by mincing with scissors, and another 10-minute incubation. Lung digest was brought

up to 5 mL with DMEM, followed by stepwise incubations (37°C x 10 minutes) with DNase 1 and then collagenase A. Shearing, filtration, and red blood cell lysis protocols were performed similar to mesenchymal cell isolation. All enzymes were obtained from Sigma.

Flow cytometry analysis and Cell Sorting: For flow cytometry analysis CD45 (BD Biosciences; Cat#557235), CD31 (Invitrogen; Cat#46-0311-82), EpCAM (Invitrogen; Cat#46-5791-82) antibodies were used to remove immune, endothelial, and epithelial cells respectively. DAPI was used as a live/dead cell discriminator. APC-Sca1 (BioLegend; Cat#108126), PE-Cy7-PDGFR α (eBioscience Cat#25-1401-82), APC-CD44 (BioLegend; Cat#103012), PerCP-CD73 (BioLegend; Cat#127214), APC-CD90.2 (BD Pharmingen; Cat#553007), CD34 (BD Pharmingen; Cat#560233), and Alexa Fluor 488-ITGA8 (R&D Systems; Cat#: FAB6194G) antibodies were used for given flow cytometry analysis. All gates are based off of full minus one controls. Flow sorting was performed by removing the DAPI positive and CD45 positive cells followed by gating for Col1GFP and FOXF1_tdTomato fluorescence. Between 1×10^5 and 1×10^6 cells per population were collected and total RNA was immediately purified. For epithelial cell sorting, cells were sorted as previously described (4). Briefly, DAPI, CD45-PacBlue (Invitrogen; 48-0451-82), CD31-PacBlue (Invitrogen; Cat#48-0311-82) positive cells were removed. Cells were then gated along FMO-controls for EpCAM-FITC and Sca1-APC-Cy7.

Epithelial Organoid Co-culture: Mice 6-10 weeks of age were used for generation of organoids. Lin⁻Col1GFP⁺PDGFR α ⁺*Foxf1*_tdTomato⁺ MCs were plated and amplified in vitro. *Foxf1*_tdTomato⁺ MCs were subsequently mixed with stem cell grade Matrigel (354277, Corning) at a concentration of 5×10^5 cells per ml. Freshly sorted BASCs from Rosa26^{mTmG/mTmG} mice were mixed with DMEM/F12 (11320-033, Gibco) at a concentration of 6×10^4 cells per ml. FOXF1_tdTomato⁺ MCs and BASC solutions were mixed at a ratio of 1:1. As a control, FOXF1_tdTomato⁺ MCs were mixed with DMEM/F12 without BASCs. The mixtures were added to 24 well transwell inserts with 0.4 μ m pores (3470, Corning). 500 μ l of DMEM/F12 containing 10% FBS, 1x transferrin/selenium/insulin (51500-056, Gibco), Penicillin/Streptomycin (Gibco),

10 μ M ROCK inhibitor Y-27632 dihydrochloride (A3008, Apex Bio), and 10 μ M TGF β inhibitor SB431542 (S1067, Selleckchem) was added to the bottom wells. The following day, 100 μ l of media cocktail was added on top of the solidified Matrigel. Media was replenished every other day. The ROCK and TGF β inhibitors were removed from the media after 4 days. To inhibit Shh signaling in select organoids, the Smoothed inhibitor LDE225 (Chemietek) was included in the media at a concentration of 0.75 μ M at day 0 and was maintained and replenished every other day. At day 8, organoids were digested with cell recovery solution (354253, Corning) in 4° Celsius for 1 hour, followed by trypsin digestion for 10 minutes (Gibco). Organoid suspension was pipetted up and down until organoids were mostly single cells. Digested epithelial/mesenchymal organoids were sorted for tdTomato⁺ BASCs and Col1GFP⁺Foxf1_tdTomato⁺ MCs. RNA was isolated from sorted cells and used for gene expression analysis.

Nuclear FOXF1 intensity analysis: Lungs from ColGFP mice stained for GFP/FOXF1/CCSP were used to determine nuclear intensity of FOXF1 in Collagen1 α 1 expressing mesenchymal cells. 40x images were taken in a field of view that contained both peri-bronchial and alveolar GFP⁺ MCs. Extended Depth of Focus (EDF) tool was used in NIS Elements to bring the entire field of view in focus and relative fluorescent units (RFU) were put on scale for each fluorescent channel. The Subtract Background tool in ImageJ software was used to remove background autofluorescence. DAPI⁺ nuclei were selected in GFP⁺ cells with ROI Manager in ImageJ. Nuclear intensity was measured and cells were classified according to peribronchiolar or alveolar region. FOXF1 nuclear intensity of each GFP⁺ MC was then standardized to the average FOXF1 nuclear intensity of alveolar MCs for each image. After each image was standardized to the alveolar compartment, FOXF1 nuclear intensity between the two groups were compiled together from 7 different images. Images were taken in distal and proximal regions of the lung. A cutoff of 1.75 fold over average FOXF1 nuclear intensity of alveolar MCs was used to describe cells with high vs low FOXF1 expression.

Gene Expression Analysis: After cDNA synthesis, gene expression analysis with TaqMan gene expression master mix (4369016, Applied Biosystems) was performed with TaqMan assays (Applied Biosystems) for mouse *Foxf1* (Mm00487497_m1), *Col1a1* (Mm00801666_g1), *Gli1* (Mm00494654_m1), *Shh* (Mm00436528_m1), *Atx* (Mm100516572_m1), and *Actb/β* (Mm02619580_g1). Power Syber Green PCR mastermix (4367659, Applied Biosystems) was used for quantification of mouse *Itga8* (Forward: CCGAAGGCCAAGGTTACTG, Reverse: AACTTCCAGGTCCTCCCACT) with *Gapdh* as a control (Forward: TGTCAGCAATGCATCCTGCA, Reverse: CCGTTCAGCTCTGGGATGAC).

Transient expression of *Foxf1* in vitro: Mouse *Foxf1* was overexpressed in LR-MSCs by transfection of pShuttle A-CMV-*mFoxf1* utilizing Lipofectamine transfection reagent in Opti-MEM I reduced serum medium. The *mFoxf1* overexpression plasmids were a generous gift by Dr. Vladimir V. Kalinichenko, MD, PhD, Cincinnati Children's Hospital Medical Center, OH. Plasmid transfection was performed on mouse lung mesenchymal cells with *Foxf1* gene or empty vector control. Protein lysates were separated on 4–12% gradient BisTris gels prior to immunoblot analysis. Plasmid transfection efficiency was determined by blotting for FOXF1 (AF4798, R&D) with GAPDH used as a loading control.

Western Blot: Freshly sorted cells were plated and passaged to P1. Cells were then harvested with RIPA buffer containing protease inhibitors. Blots were incubated with antibodies against FOXF1 and GAPDH in 5% milk/TBS-Tween, and probed for these proteins as previously described (5).

RNA-seq Analysis: RNA sequencing and Biostatistical Analysis: Three different subsets were sorted such that cell populations expressed both type 1 collagen and Foxf1 (DBL), only type 1 collagen (ColGFP), or only Foxf1 (TD Only). RNA was isolated from cell population subsets and submitted to the University of Michigan DNA Sequencing Core where ribosome ribosomal RNA depletion was performed prior to RNA-Sequencing utilizing an Illumina Hi-Seq 4000 platform. Bioinformatic analyses was provided by the University of Michigan Bioinformatics Core in the form of differential expression at gene and isoform

level, GO annotation of differentially expressed genes, functional enrichment analysis of set of differentially expressed genes and diagnostic plots. Briefly, the reads files from the Sequencing Core's storage were downloaded and concatenated into a single fastq file for each sample. The quality of the raw reads data for each sample was checked using FastQC [1] (version v0.11.3) to identify features of the data that may indicate quality problems (e.g. low quality scores, over-represented sequences, inappropriate GC content). The Tuxedo Suite software package for alignment, differential expression analysis, and post-analysis diagnostics was used (6-8). Briefly, the Bioinformatics Core aligned reads to the reference mRNA transcriptome (mm10) using TopHat2 (version 2.0.13) and Bowtie2 (version 2.2.1.). Default parameter settings for alignment, with the exception of: "--b2-very-sensitive" telling the software to spend extra time searching for valid alignments were used. FastQC for a second round of quality control (post-alignment) to ensure that only high-quality data would be input to expression quantitation and differential expression analysis and Cufflinks/CuffDiff (version 2.2.1) was utilized for expression quantitation, normalization, and differential expression analysis, using mm10.fa as the reference genome sequence. For this analysis, parameter settings: "--multi-read-correct" to adjust expression calculations for reads that map in more than one locus, as well as "--compatible-hits-norm" and "--upper-quartile-norm" for normalization of expression values were used. We generated diagnostic plots using the CummeRbund R package. Locally developed scripts to format and annotate the differential expression data output from CuffDiff were used. Briefly, the Bioinformatics Core identified genes and transcripts as being differentially expressed based on three criteria: test status = "OK", $FDR \leq 0.05$, and fold change $\geq \pm 1.5$. Genes and isoforms were annotated with NCBI Entrez Gene IDs and text descriptions.

Single-cell RNAseq Analysis: Lungs from *Gli1*^{CreERT2/WT};Rosa26^{mTmG/WT} B6D2F1/J mice were harvested. Single cell suspension was sorted for CD45⁻CD31⁻ cells by flow cytometry, checked to ensure >80% viability and submitted to the University of Michigan Advanced Genomics core for processing and sequencing. Single-cell RNA-seq libraries were prepared using the 10X Genomics Chromium Next GEM Single Cell 3' Kit v3.1 (part number 1000268) following the manufacturer's protocol. The libraries were

run on an Agilent TapeStation 4200 (part number G2991BA) for library quality control before sequencing on a NovaSeq6000 with the following run configuration: Read 1 - 150 cycles; i7 index read - 8 cycles; Read 2 - 150 cycles. The total number of cells recovered was 11,118. We performed dimensionality reduction and clustering of single-cell RNA-seq data using our previously published LIGER algorithm (9, 10). Briefly, we performed integrative nonnegative matrix factorization with using the *optimizeALS* function with $k = 20$ metagene factors. We then identified clusters using the *louvainCluster* function with default parameters. We used the Wilcoxon rank-sum test to find differentially expressed genes. We then visualized the cells in two dimensions using Uniform Manifold Approximation and Projection (UMAP) using the *runUMAP* function in the *liger* package with default parameters. RNA velocity analysis was performed using the *scvelo* package (11). We fit the dynamical model using default parameters and projected the velocity vectors onto the UMAP plot from the LIGER analysis. For the correlation analysis in Fig. 3H, we used the MAGIC python package (12), to smooth gene expression values among neighboring cells and account for the high rate of missing values in single-cell gene expression measurements. We then calculated Pearson correlation of the smoothed values and computed the p-value of the correlation from a student t distribution.

High content analysis to phenotype $\text{Lin}^- \text{PDGFR}\alpha^+ \text{Itga8}^{+/-}$ mesenchymal cells. Lungs were harvested from three B6D2F1/J mice, sorted by flow cytometry for $\text{Lin}^- \text{PDGFR}\alpha^+$ mesenchymal cells that were either $\text{Itga8}^{+/-}$ and plated for migration, proliferation and morphological analysis.

Cell Migration Assay. For migration, $\text{Itga8}^{+/-} \text{PDGFR}\alpha^+ \text{Lin}^-$ mesenchymal cells were plated at 10,000 cells per well in an Oris™ Cell Migration Assay plate in 8 wells, and allowed to adhere and grow until confluent. When confluent, the stoppers were removed to expose the denuded zone for cell migration and proliferation and were allowed to proliferate or migrate for 48 hours. Cells were fixed in 4% paraformaldehyde and stained with Hoechst-33342 and HCS CellMask Orange (Invitrogen cat#H1399, H32713) at the recommended concentration. The plates were imaged on a Yokogawa Cell Voyager 8000 with a 4x

objective lens and 405 nm/561 nm excitation lines and 445/40 nm and 600/37 nm emission filters. Images were masked using the area determined by a stopper left seated until the end of the experiment and cells inside the denuded zone were identified using CellProfiler(13). Eight replicates were imaged for each condition and a student's t test was performed to determine statistical significance.

Cell Proliferation Assay and Morphologic Cell Profiling. Itga8^{+/+} PDGFR α ⁺ Lin⁻ mesenchymal cells were plated at 5,000 cells per well in a tissue culture-treated optical bottom 96-well plate (NUNC cat#165305). Cells were allowed to adhere and were imaged every two hours for 48 hours in a Yokogawa Cell Voyager V8000 with bright field imaging using a 10x/0.4NA objective lens. Cells were segmented using Cellpose (14), and the number of cells per field was plotted versus time. The log of cell counts versus time was fit to a linear model using Python Scikit-Learn and doubling times were determined from the linear model. At the end of live cell imaging, cells were fixed with 4% paraformaldehyde and stained with Hoechst-33342 (10 μ g/mL, ThermoFisher cat#H1399) and HCS CellMask Orange (1:5,000 dilution, ThermoFisher: cat#H32713). Cells were imaged in a Yokogawa Cell Voyager 8000 with a 20X/1.0NA water immersion objective lens. Image analysis was performed in CellProfiler and single-cell statistical analysis was performed in Scikit-Learn.

Doubling Time Quantitation. Live cell bright field imaging captures images every 2 hours. An average of 4 fields are counted per well per condition. Log transform counts are used to perform a linear fit model. Doubling time is calculated from the slope and confidence interval of the data points are shown as a band around the fit line. The first 12 hours of lag time were not included in the analyses since there was a lag time at the beginning of the run because of no proliferation.

Statistics: Two-tailed student t-Test was performed for experiments where two groups are compared. For more than two groups, one-way ANOVA was used with Bonferroni's multiple comparison test. Biological replicates are represented as points in each graph. Error bars represent standard deviation. Statistical significance is designated as: *p<0.05, **p<0.01, ***p<0.001.

Author contributions:

Designing of research studies: VNL, RRB, NMW; Experiments and data acquisition: RRB, NMW, KM, SM-P, YA, GK, DSW, JZS; Data analysis: VNL, JDW, RRB, NMW, RL, RV, DSW, JZS; Pathology review: CFF, VNL; Manuscript preparation: VNL, RRB, NMW, RV, JDW, JZS.

Acknowledgements:

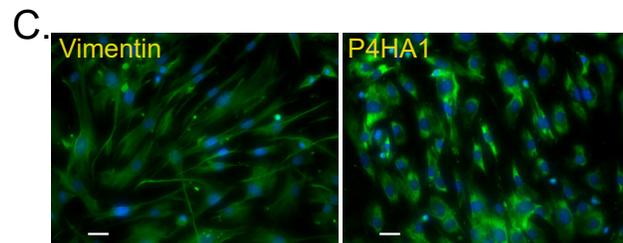
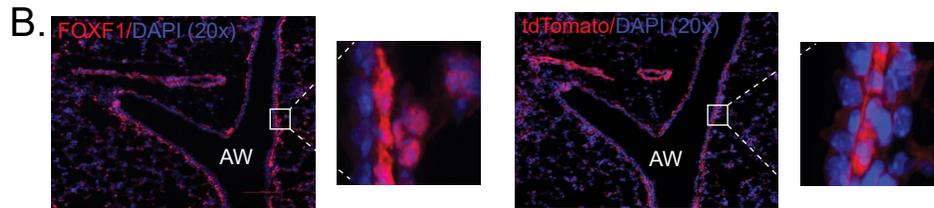
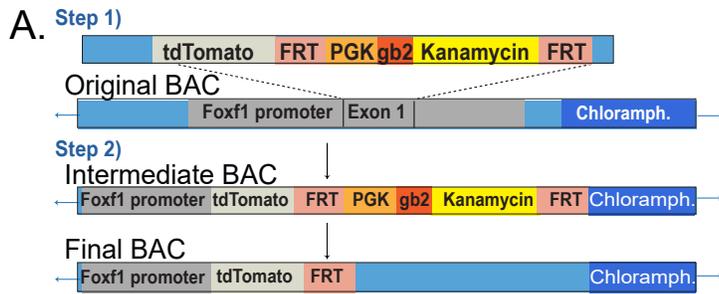
We would like to thank the experienced personnel at the Flow Cytometry Core, Research Histology Core (NIH P30 CA04659229), Microscopy Core, Advanced Genomics Core, Bioinformatics Core, and Transgenic Animal Model Core at the University of Michigan for their contributions to this study.

References:

1. Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, et al. A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics*. 2001;73(1):56-65.
2. Yata Y, Scanga A, Gillan A, Yang L, Reif S, Breindl M, et al. DNase I-hypersensitive sites enhance alpha1(I) collagen gene expression in hepatic stellate cells. *Hepatology*. 2003;37(2):267-76.
3. Moiseenko A, Kheirollahi V, Chao CM, Ahmadvand N, Quantius J, Wilhelm J, et al. Origin and characterization of alpha smooth muscle actin-positive cells during murine lung development. *Stem Cells*. 2017;35(6):1566-78.
4. Mimura T, Walker N, Aoki Y, Manning CM, Murdock BJ, Myers JL, et al. Local origin of mesenchymal cells in a murine orthotopic lung transplantation model of bronchiolitis obliterans. *Am J Pathol*. 2015;185(6):1564-74.
5. Cao P, Walker NM, Braeuer RR, Mazzoni-Putman S, Aoki Y, Misumi K, et al. Loss of FOXF1 expression promotes human lung-resident mesenchymal stromal cell migration via ATX/LPA/LPA1 signaling axis. *Sci Rep*. 2020;10(1):21231.
6. Langmead B, Trapnell C, Pop M, and Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
7. Trapnell C, Pachter L, and Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105-11.
8. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, and Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;31(1):46-53.
9. Liu J, Gao C, Sodicoff J, Kozareva V, Macosko EZ, and Welch JD. Jointly defining cell types from multiple single-cell datasets using LIGER. *Nat Protoc*. 2020;15(11):3632-62.
10. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, and Macosko EZ. Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell*. 2019;177(7):1873-87 e17.
11. Bergen V, Lange M, Peidli S, Wolf FA, and Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol*. 2020;38(12):1408-14.
12. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell*. 2018;174(3):716-29 e27.

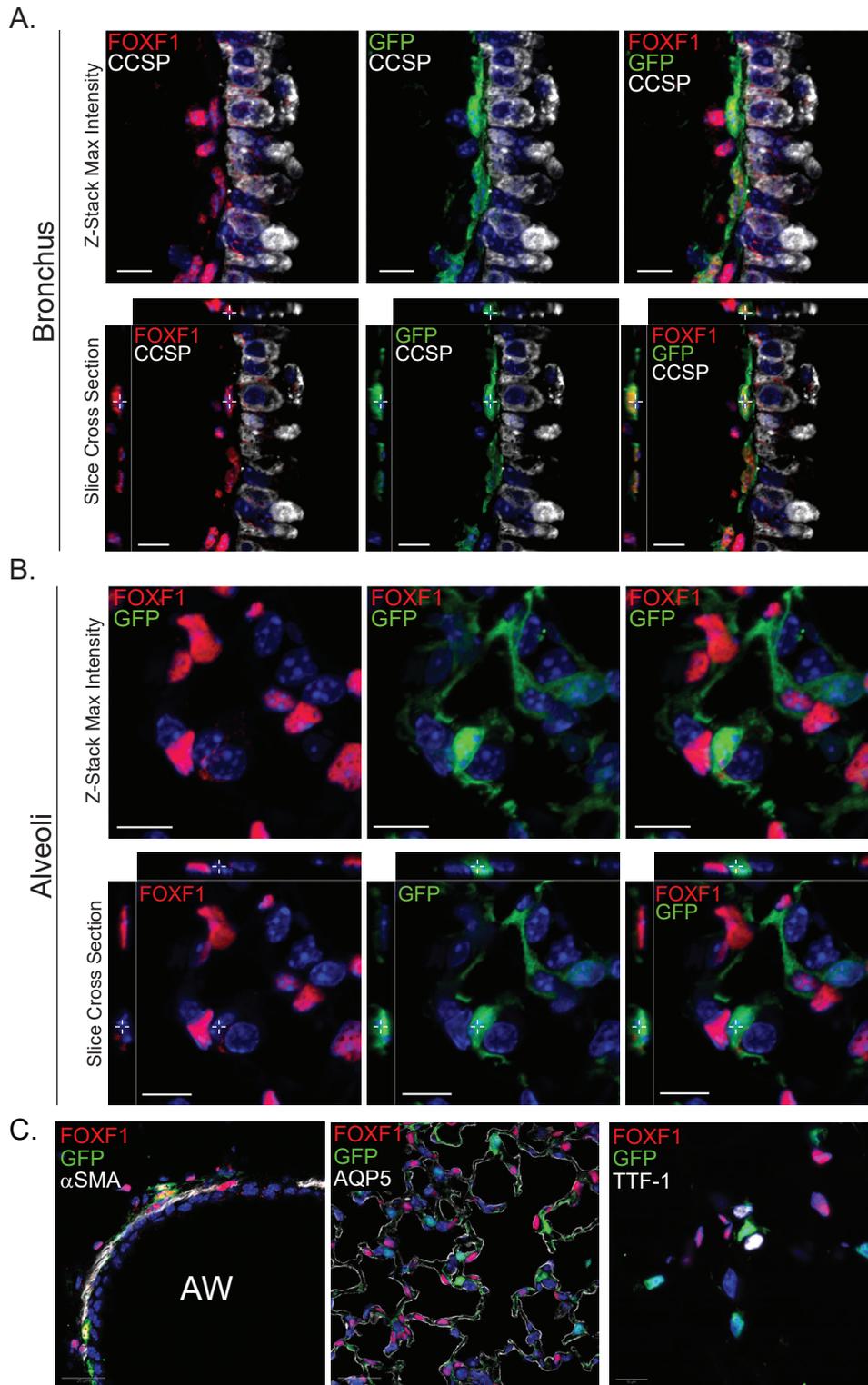
13. McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, et al. CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* 2018;16(7):e2005970.
14. Stringer C, Wang T, Michaelos M, and Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods.* 2021;18(1):100-6.

SUPPLEMENTAL FIGURE 1



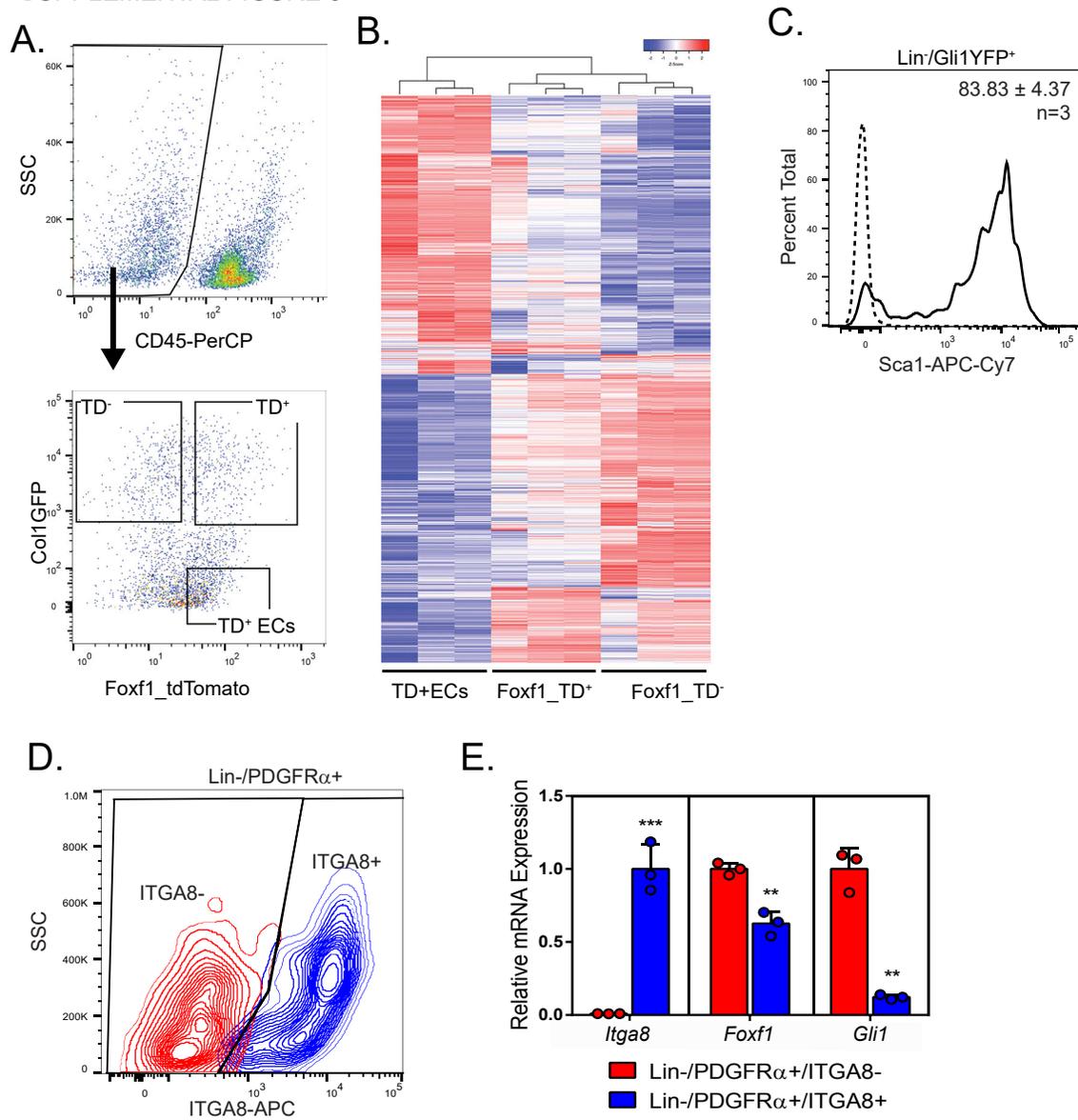
Supplemental Figure 1: (A) Schemata showing the generation of C57BL/6J-Tg(*Foxf1_tdTomato*) (*Foxf1_tdTomato*) mice utilizing a modified bacterial artificial chromosome (BAC) in which exon 1 of the *Foxf1* gene was replaced with the coding sequence of tdTomato. (B) Representative immunofluorescence images of FOXF1 and tdTomato expression in contiguous lung tissue sections from *Foxf1_tdTomato* mice. n=3. Scale bar:100 μ m. (C) Vimentin and P4HA1 (prolyl-4-hydroxylase) staining in sorted *Foxf1_tdTomato*⁺ cells in vitro. n=3. Scale bar:100 μ m.

SUPPLEMENTAL FIGURE 2



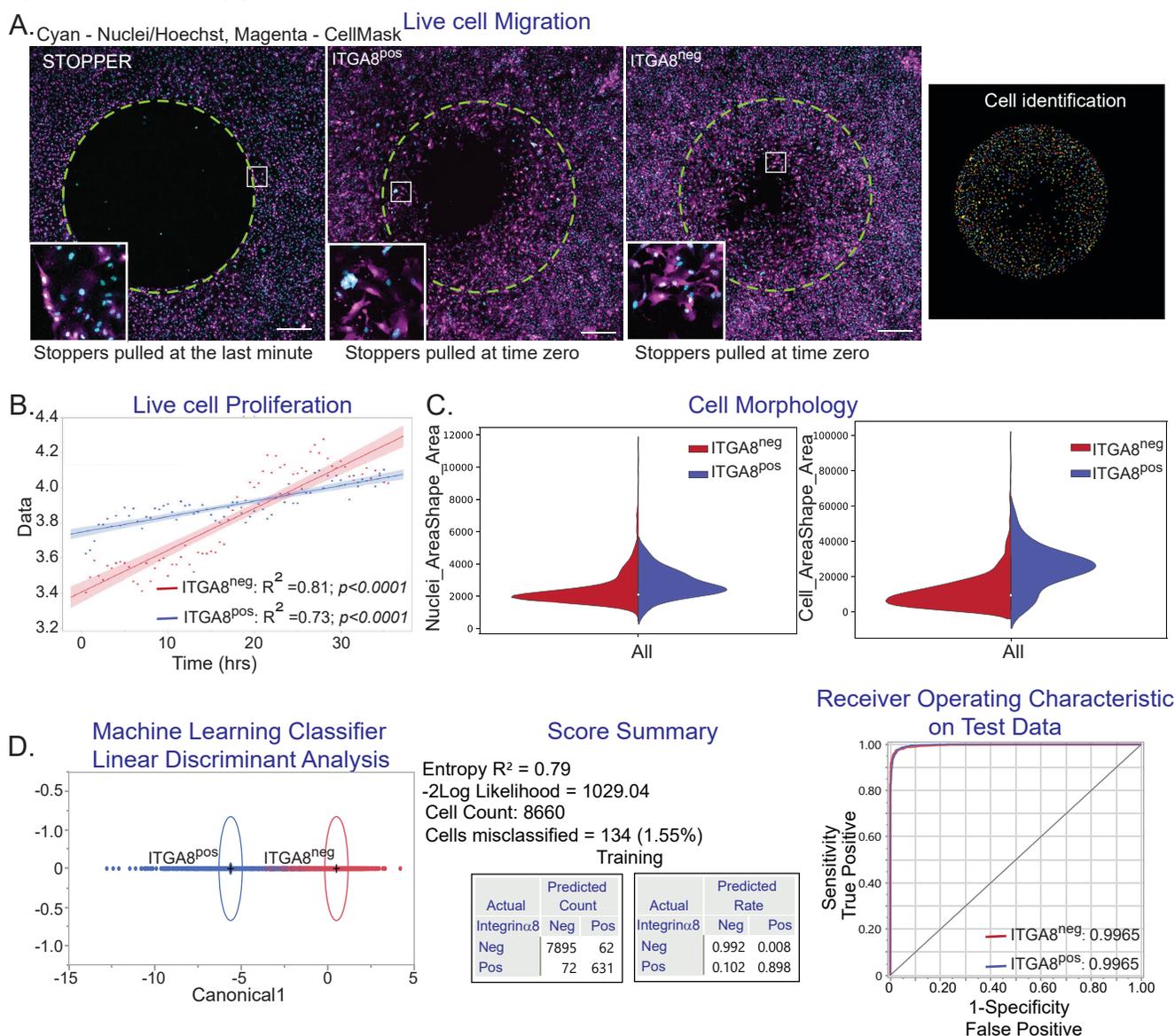
Supplemental Figure 2: FOXF1 expression status in peri-bronchial and alveolar Col1GFP⁺ MCs. A) Zoomed in Z-stack crop of bronchus from Figure 2B-inset 2. X/Y/Z planes are shown to confirm FOXF1 presence in Col1GFP⁺ peribronchial MCs. White crosshairs show location of z-slice. Yellow indicates co-staining of Col1GFP and FOXF1. 10 μ m scale bar is shown. B) Zoomed in Z-stack crop of alveoli from Figure 2B- inset 4. X/Y/Z planes are shown to confirm FOXF1 absence in alveolar Col1GFP⁺ MCs. White crosshairs show location of z-slice. Collagen 1 expressing MCs are seen lying adjacent to FOXF1 expressing endothelial cells. 10 μ m scale bar is shown. C) Confocal imaging of FOXF1 and Col1GFP with the smooth muscle marker α SMA, type 1 pneumocyte marker AQP5, and type 2 pneumocyte marker TTF-1. n=3. 10 μ m scale bar is shown.

SUPPLEMENTAL FIGURE 3



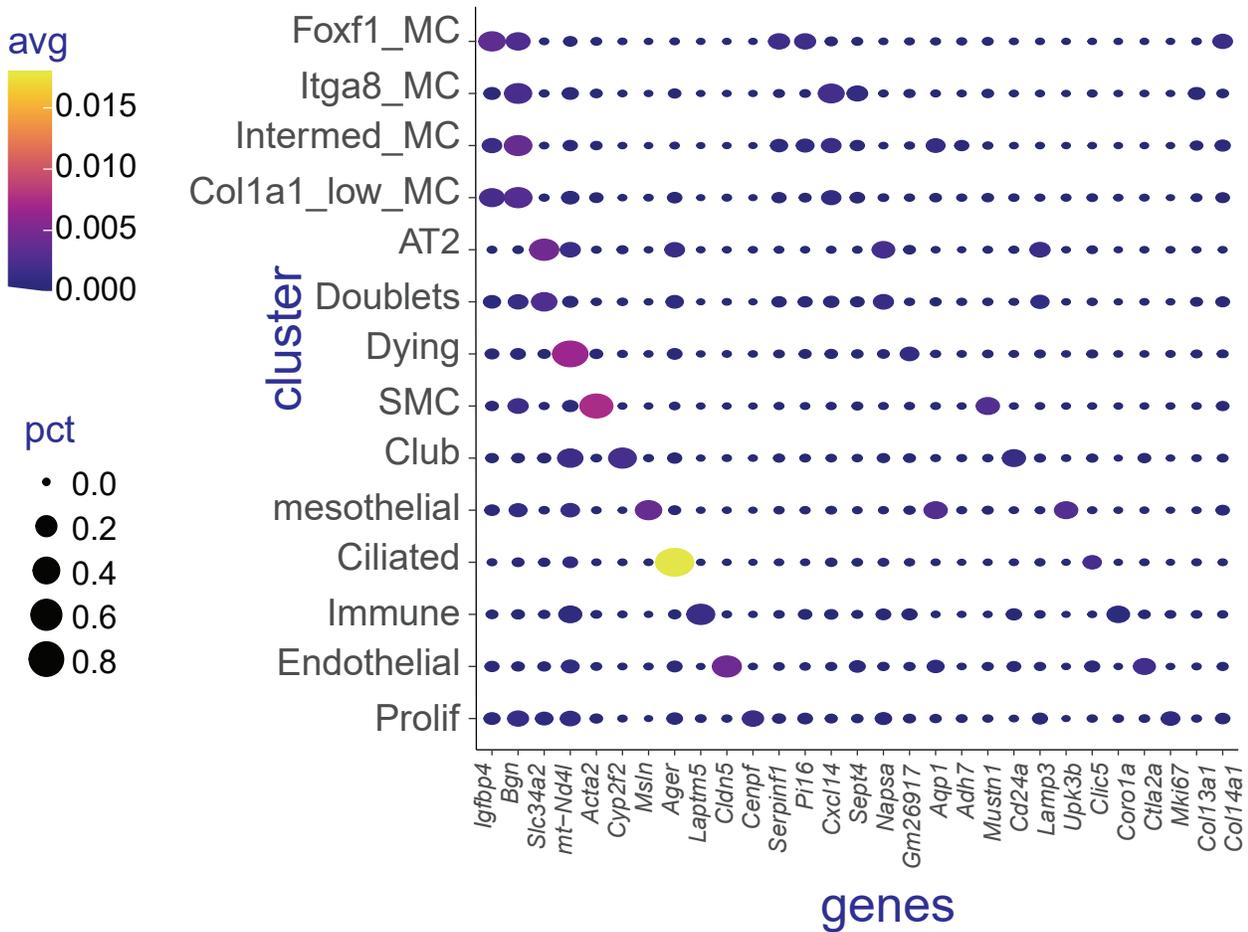
Supplemental Figure 3: (A-B) Transcriptomic analysis of Foxf1(TD⁺) and Foxf1(TD⁻) Col1⁺ MCs isolated from Foxf1_tdTomato;Col1GFP mice. (A) Gating strategy for sorted populations sent for RNA sequencing is shown. (B) RNA seq was performed on CD45⁻ cells that were either Col1GFP⁺Foxf1_tdTomato⁻ (TD⁻) or Col1GFP⁺Foxf1_tdTomato⁺ (TD⁺) MCs. Col1GFP⁻Foxf1_tdTomato⁺ endothelial cells (TD⁺ ECs) were used for comparison of mesenchymal genes. Heatmap showing cluster analysis of expressed genes in TD⁺ ECs, Foxf1_TD⁺ MCs, and Foxf1_TD⁻ MCs is shown. n=3. (C) **Sca1 expression in Gli1⁺ MCs. Lin⁻Gli1YFP⁺ cells sorted from adult lungs of Gli1^{CreERT2}-YFP mice were analyzed for Sca1 expression by flow cytometry. Sca1 expression overlaps with Gli1 expression in Lin⁻ cells. n=3. (D-E) **Integrin alpha 8 (ITGA8) as a differential marker of MC populations.** Lin⁻PDGFRα⁺ITGA8⁺ (blue) or Lin⁻PDGFRα⁺ITGA8⁻ (red) mesenchymal cells were sorted by flow cytometry (D) and then subjected to gene expression analysis by real-time PCR for *Itga8*, *Foxf1*, and *Gli1* (E). n=3. Values: mean±SD. **p<0.01, ***p<0.001; 2-tailed unpaired t test.**

SUPPLEMENTAL FIGURE 4



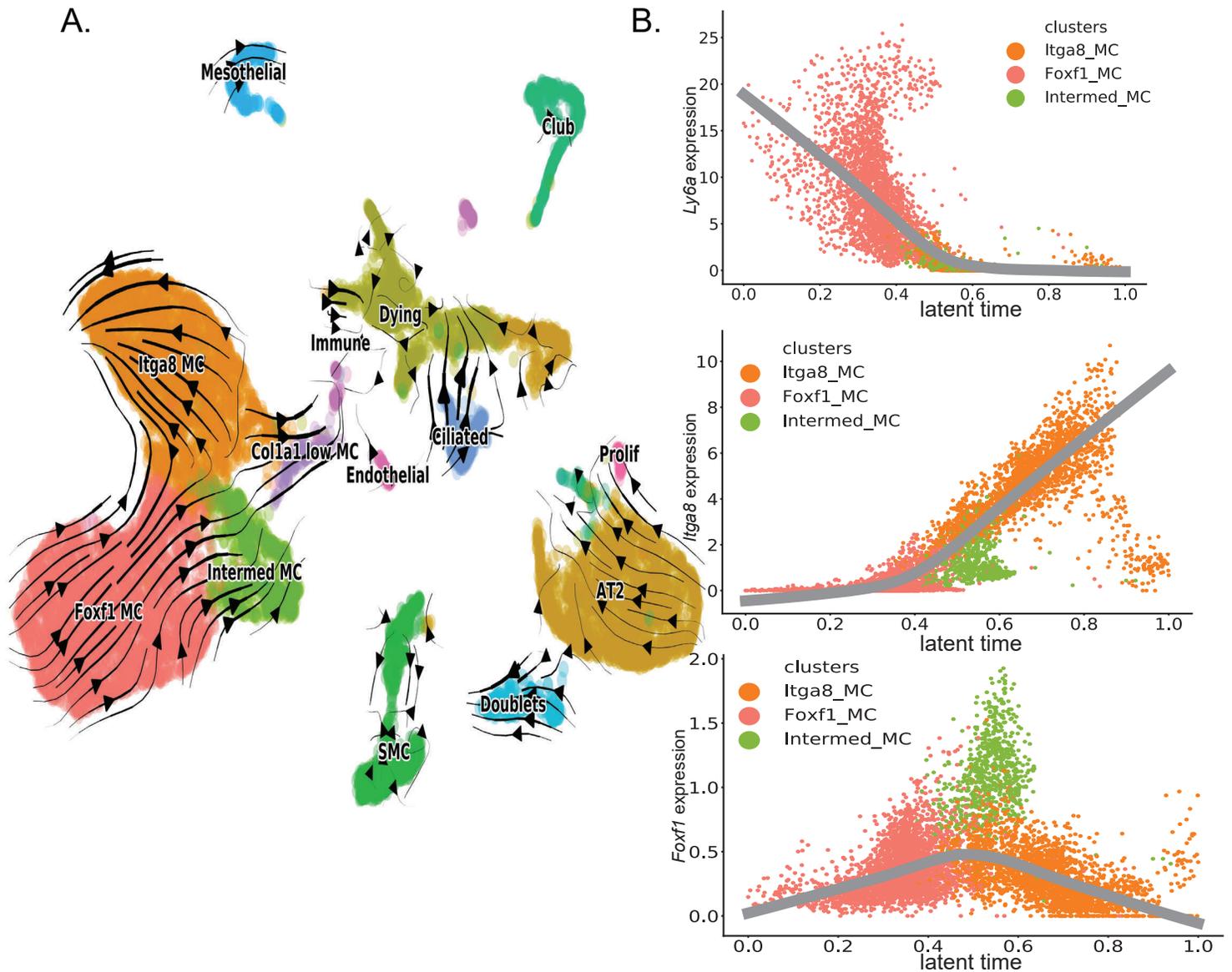
Supplemental Figure 4. High content analyses of live cell migration, proliferation and morphology of Lin⁻ PDGFR α ⁺ ITGA8^{pos/neg} mesenchymal cells. Lungs were harvested from three B6D2F1/J mice and sorted for Lin⁻ PDGFR α ⁺ mesenchymal cells that were either ITGA8^{+/-} and plated for migration, proliferation and morphological analysis. **(A)** Plating density was 10,000 cells per well in an Oris™ Cell Migration Assay plate with control (stopper pulled after fixation to delineate the denuded zone) and ITGA8^{+/-} cells showing invasion into the denuded zone (inside green circle) with an average of 1,404 cells in the ITGA8⁻ condition, and 892 cells in the ITGA8⁺ cells that show less invasion. 8 replicates per condition and $p=0.0027$. Representative images are shown with the nuclei stained with Cyan Hoeschst stain and the cell is identified by magenta. Cell identification process is shown in the representative image where 1,641 objects were accepted with sizes ranging between 9 and 13.3 pixels. Scale bar: 400 μ m. **(B)** Plating density was 5,000 cells per well with cell proliferation assay showing ITGA8^{+/-} cells observed for 48 hours. The timeframe for the linear fit was trimmed uniformly at early and late times to eliminate early lag-phase and growth saturation/confluent conditions. Doubling times for ITGA8^{neg/pos} cells were determined to be 18.2 hours and 34.6 hours, respectively. **(C)** Morphologic cell profiling was performed on both cell populations fixed and stained with Hoechst-33342 and CellMask Orange to investigate differences and distributions of nuclear area and whole cell area are presented as violin plots. **(D)** Linear discriminant analysis was used to classify cells as ITGA8^{+/-} with an accuracy of 99.65% demonstrating that these cells are morphologically distinct. 8,660 cells were classified accurately with 7,895 as ITGA8^{neg} cells and 631 as ITGA8^{pos} cells, and only 134 cells (1.55%) were misclassified.

SUPPLEMENTAL FIGURE 5



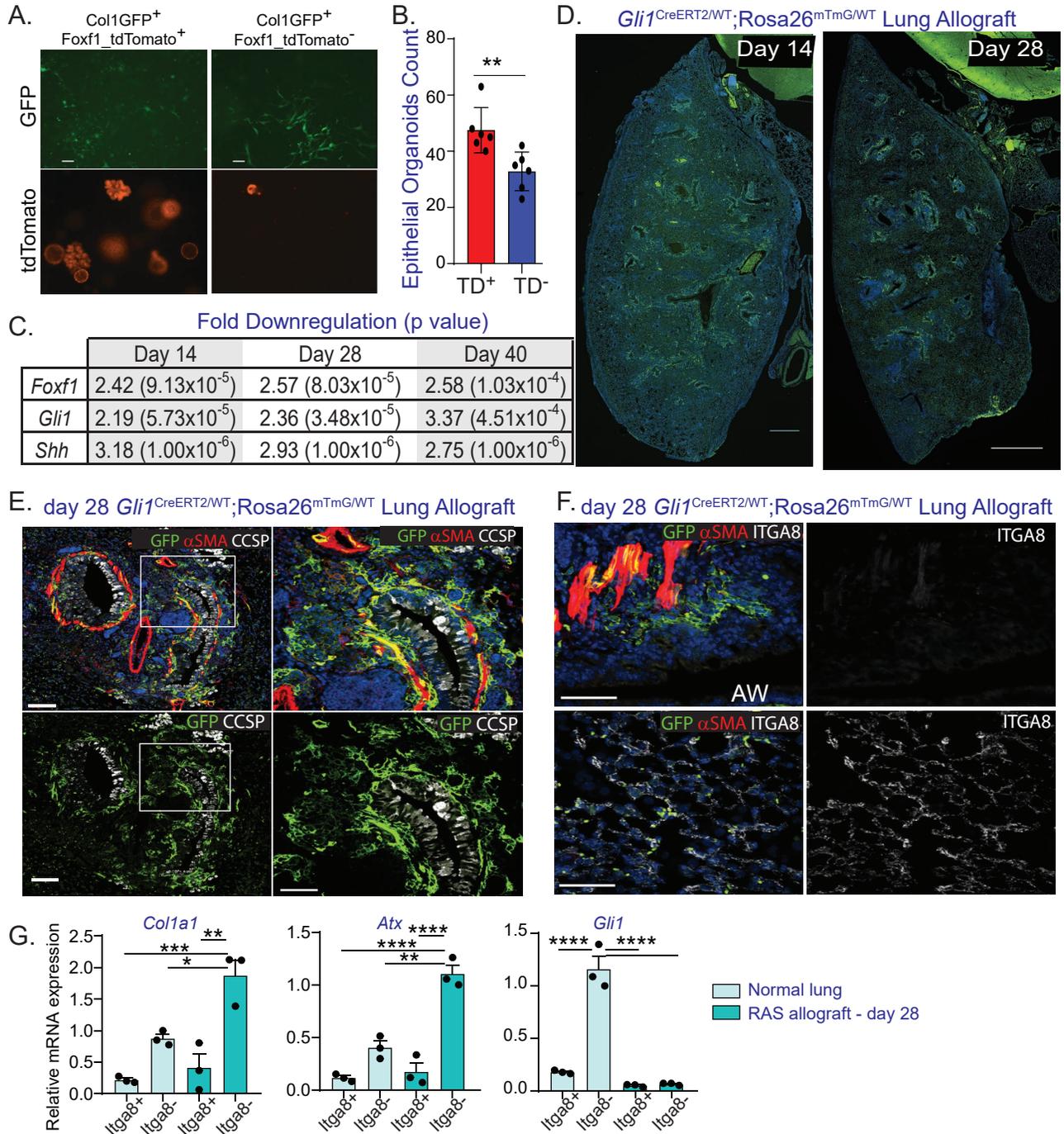
Supplemental Figure 5. Single-cell RNA sequencing (scRNA-seq) on CD45⁺CD31⁻ populations sorted from adult murine lungs. Dot plot representation of 11,118 single-cell RNAseq profiles. Dot plot showing proportion of cells expressing each gene (circle size) and mean expression level (circle color) of key marker genes within each cluster.

SUPPLEMENTAL FIGURE 6



Supplemental Figure 6: RNA velocity analysis of single-cell RNA-seq from lungs of *Gli1^{CreERT2/WT};Rosa26^{mTmG/WT}* mice. RNA velocity analysis uses the ratio between spliced mRNA and unspliced pre-mRNA to estimate a “velocity vector” representing the likely future state of each cell, which can give insights into dynamic transitions among cell populations. **(A)** RNA velocity estimates for 11,118 cells projected on the UMAP coordinates from Fig. 3E. Streamlines indicate the local average velocity evaluated on a regular grid. **(B)** The x-axis indicates latent time value inferred from velocity analysis. The y-axis value is the expression level of each gene within each cell, and the colors indicate cluster membership for each cell. The grey lines show inferred expression trends from polynomial regression fits. These findings are consistent with our findings that the Foxf1_MC population are localized in the proximal airway while the Itga8_MC population are localized in the distal alveolar compartment; this direction is also consistent with the known spatial axis of development of lung.

SUPPLEMENTAL FIGURE 7



Supplemental Figure 7: (A-B) Organoid forming capacity of Foxf1TD⁺ and Foxf1TD⁻ Col1⁺ MCs. Col1GFP⁺ MCs that are Foxf1TD⁺ and Foxf1TD⁻ were sorted from *Foxf1*_tdTomato;Col1a1_GFP mice bi-transgenic mice and co-cultured with freshly isolated bronchioalveolar stem cells (BASCs) from *Rosa26*^{mTmG} mice in Matrigel. Organoid formation at day 8 was quantitated and compared. n=3. Values: mean±SD. **p<0.01; 2-tailed unpaired t test. Scale bar:100 μm. **(C) Differential regulation of *Gli1*, *Foxf1* and *Shh* signaling in CLAD mouse model.** Gene expression of *Foxf1*, *Gli1*, and *Shh* from whole allograft lung digests at day 14, 28, and 40 post lung transplants. Values represent fold change compared to isograft. Statistics: one-way ANOVA; Bonferroni test. **(D-G) Contribution of BVB-MCs to fibrogenesis in a rejecting lung allograft.** *Gli1*^{CreERT2/WT}; *Rosa26*^{mTmG/WT} mice were labeled at 4-6 weeks of age and prior to allograft lung transplantation. **(D)** Whole lung immunofluorescence image of allografts is shown at day 14 and day 28 post-lung transplantation with robust expansion of GFP labeled *Gli1*⁺ cells noted in the peri-bronchial distribution. day 14: n=2; day 28: n=3. Scale bar:1 mm. **(E)** Allografts immunostained for CCSP and αSMA demonstrating expansion and myfibroblast differentiation of *Gli1*⁺ MCs. Scale bar:100 μm and 50 μm (zoomed crop images). **(F)** ITGA8 and αSMA immunofluorescent staining demonstrating the continued expression of ITGA8 in alveolar MCs with no fibrotic expansion or myfibroblast differentiation. Scale bar:100 μm and 50 μm (zoomed crop images). **(G)** PDGFRα⁺ MCs that were either ITGA8⁺ or ITGA8⁻ were sorted from normal C57BL/6 lungs or RAS allografts at day 28. Real-time PCR analysis on the sorted cells for *Col1a1*, *Atx* and *Gli1* is shown. Values: mean±SEM. *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001; one-way ANOVA; Bonferroni test.

Supplemental Table 1.**Biological processes associated with epithelial regulation in Col1GFP⁺/Foxf1_tdTomato⁻ vs. Col1GFP⁺/Foxf1_tdTomato⁺ mesenchymal cells.**

GO:ID	Name	DE genes	Total genes	P value
GO:0060429	epithelium development	254	824	2.31E-16
GO:0002009	morphogenesis of an epithelium	143	445	2.11E-09
GO:0030855	epithelial cell differentiation	130	399	8.74E-09
GO:0010631	epithelial cell migration	79	207	5.00E-08
GO:0090132	epithelium migration	79	208	6.63E-08
GO:0050673	epithelial cell proliferation	107	320	1.63E-07
GO:0060562	epithelial tube morphogenesis	100	307	3.94E-06
GO:0010632	regulation of epithelial cell migration	60	156	1.25E-05
GO:0050678	regulation of epithelial cell proliferation	88	267	2.31E-05
GO:0061138	morphogenesis of a branching epithelium	68	189	2.79E-05

Shown are the top 10 significantly enriched GO terms for Col1GFP⁺/Foxf1_tdTomato⁻ vs.

Col1GFP⁺/Foxf1_tdTomato⁺ mesenchymal cells in the physiological regulation of epithelial cells, and ranked by FDR-adjusted p values. Full database is submitted in the GEO - Accession number:GSE176476.