Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/28/eaba1972/DC1

Supplementary Materials for

Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis

Arun C. Habermann, Austin J. Gutierrez, Linh T. Bui, Stephanie L. Yahn, Nichelle I. Winters, Carla L. Calvi, Lance Peter, Mei-I Chung, Chase J. Taylor, Christopher Jetter, Latha Raju, Jamie Roberson, Guixiao Ding, Lori Wood, Jennifer M. S. Sucre, Bradley W. Richmond, Ana P. Serezani, Wyatt J. McDonnell, Simon B. Mallal, Matthew J. Bacchetta, James E. Loyd, Ciara M. Shaver, Lorraine B. Ware, Ross Bremner, Rajat Walia, Timothy S. Blackwell, Nicholas E. Banovich*, Jonathan A. Kropski*

*Corresponding author. Email: jon.kropski@vumc.org (J.A.K.); nbanovich@tgen.org (N.E.B.)

Published 8 July 2020, *Sci. Adv.* **6**, eaba1972 (2020)
DOI: 10.1126/sciadv.aba1972

The PDF file includes:

Figs. S1 to S22
Table S1
Legend for table S2
Tables S3 and S4
Legend for table S5
Table S6
Legends for tables S7 to S10
Table S11

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/28/eaba1972/DC1)

Tables S2, S5, S7 to S10

Supplementary Figures

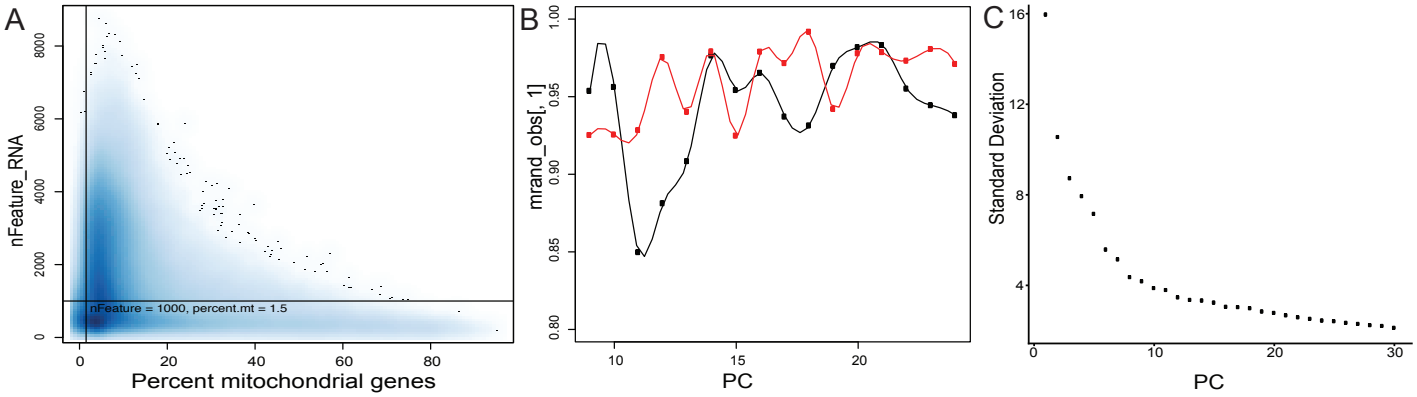


Figure S1. Density plot demonstrating distribution of percentage of mitochondrial genes and nFeature_RNA in all scRNA-seq samples (A). Correlation coefficient values (mrand_obs) between two adjacent UMAP-1 and UMAP-2 (B) and Elbow Plot showing standard deviation between adjacent PC for 30 PCs (C).

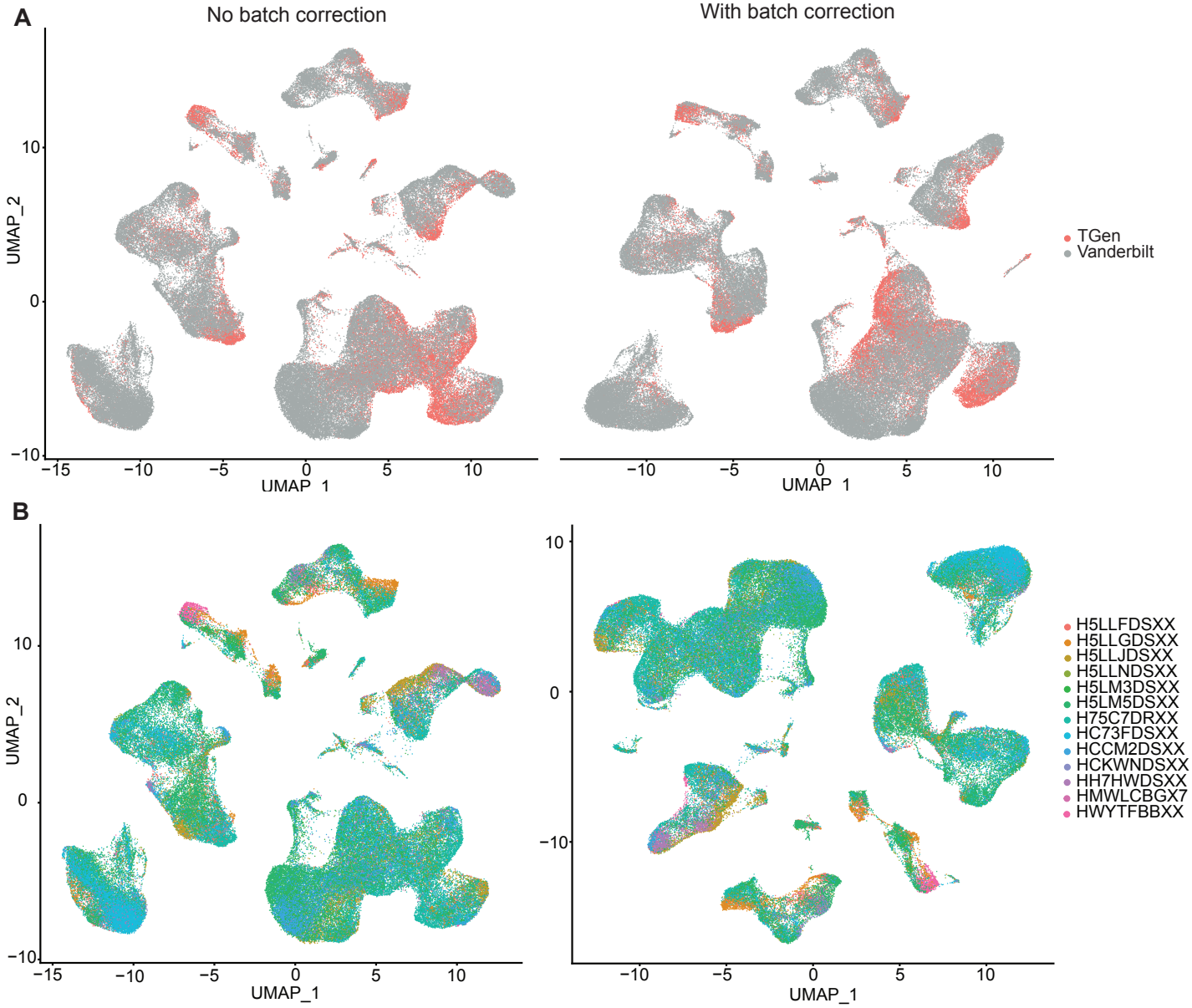


Figure S2. Batch-correction using the batch_var option in SCTtransform for Processing_sites (A) and Flowcell ID (B) demonstrated minimal batch effects on the clustering process. Axis rotation observed in B after batch occurs by change during manifold embedding and is not associated with changes in the underlying data structure.

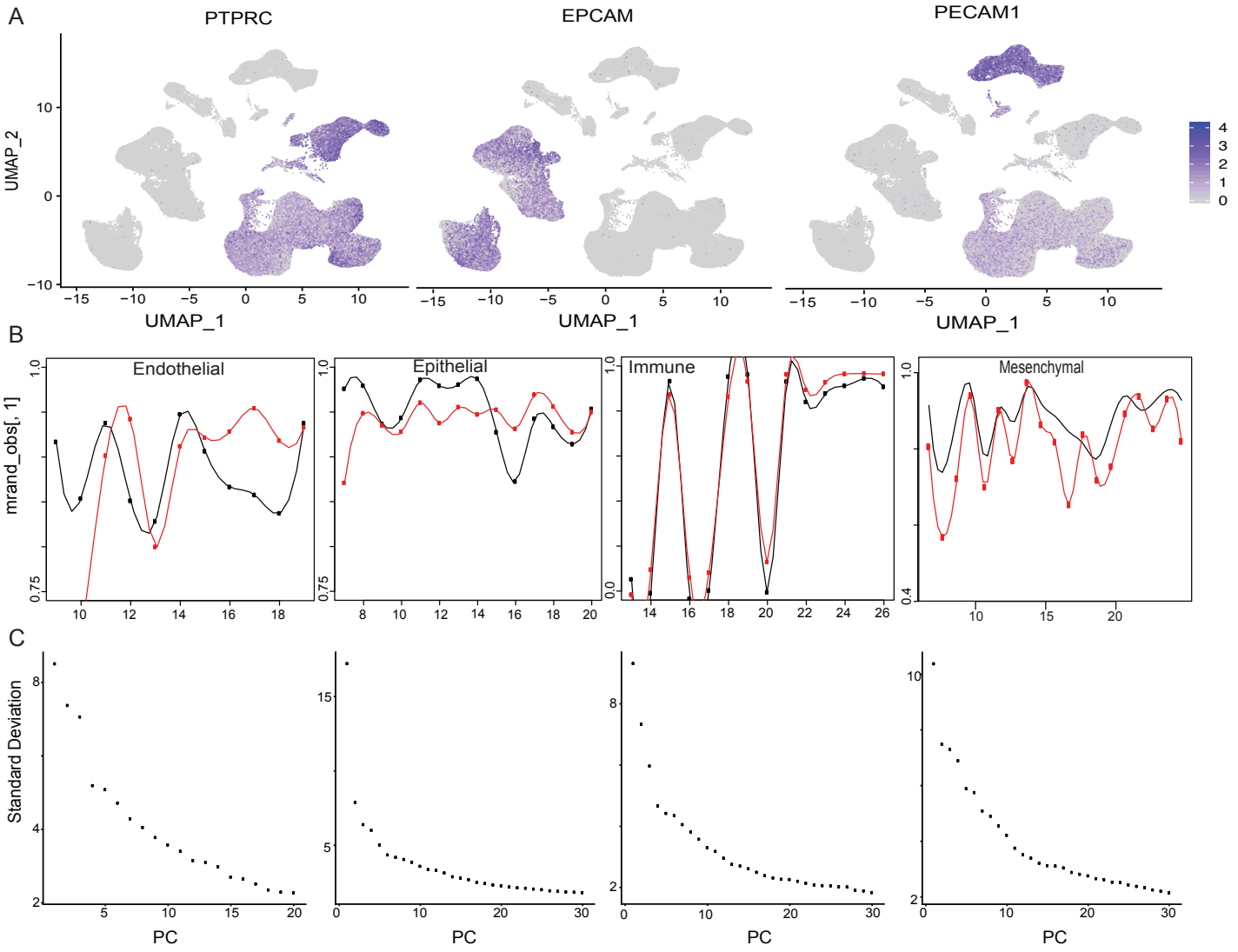


Figure S3. Feature Plots showed expression of cell type specific markers (*PTPRC*, *EPCAM* and *PECAM1*) on Seurat clusters (A). Correlation coefficient values (mrand_obs) between two adjacent UMAP-1 and UMAP-2 (B) and Elbow Plot showing standard deviation between adjacent PC for 30 PCs (C).

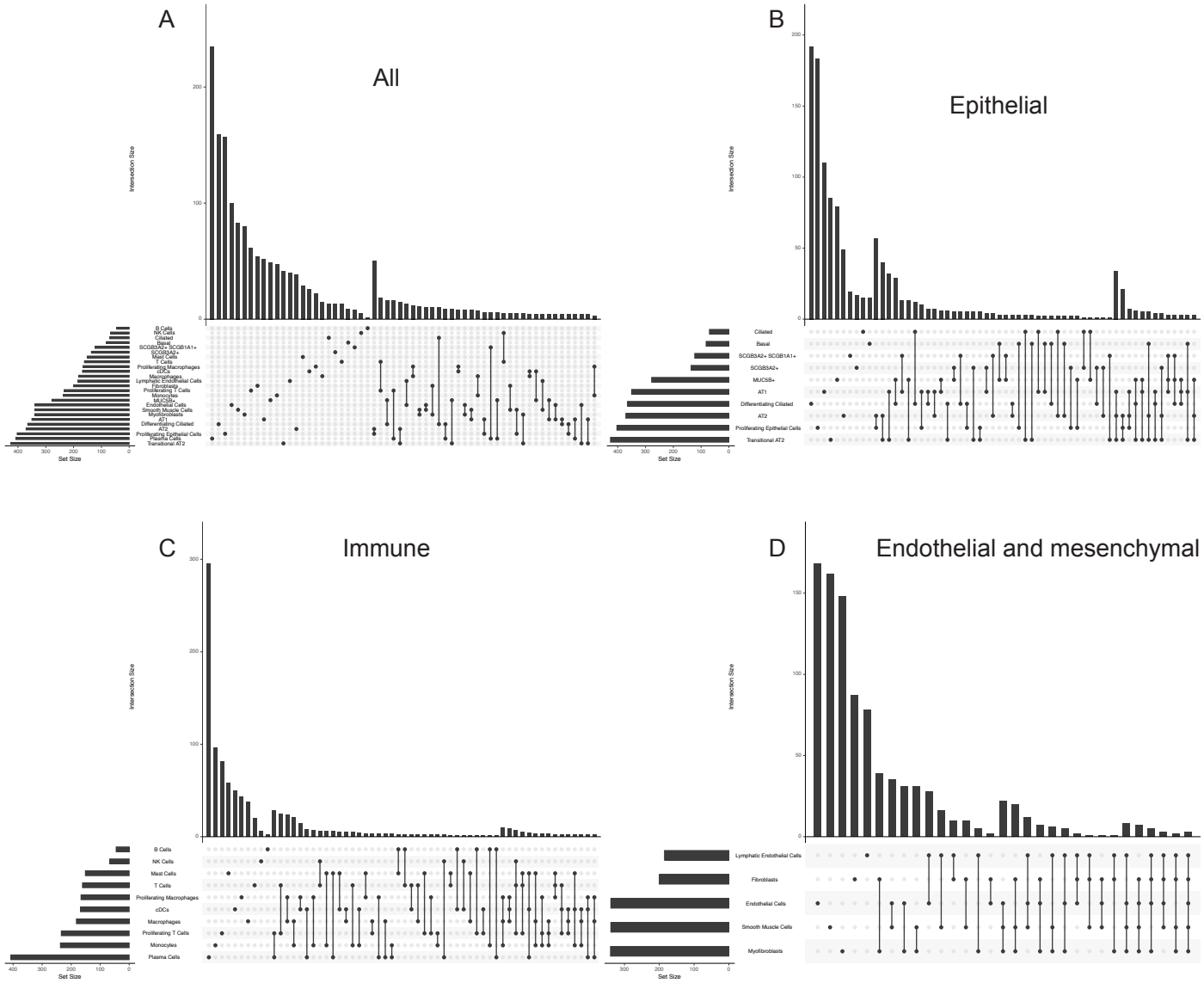


Figure S4. Upset plots of shared DEGs between cell types. Set size indicates the number of genes that are DE in each cell type. The histogram represents the number of genes within each column. Individual dots represent cell type specific expression. Dots connected by lines represent DE genes that are shared between the cell types. These plots represent the top 60 intersections across **A)** all cell types, **B)** epithelial cells, **C)** immune cells, and **D)** endothelial and mesenchymal cells.

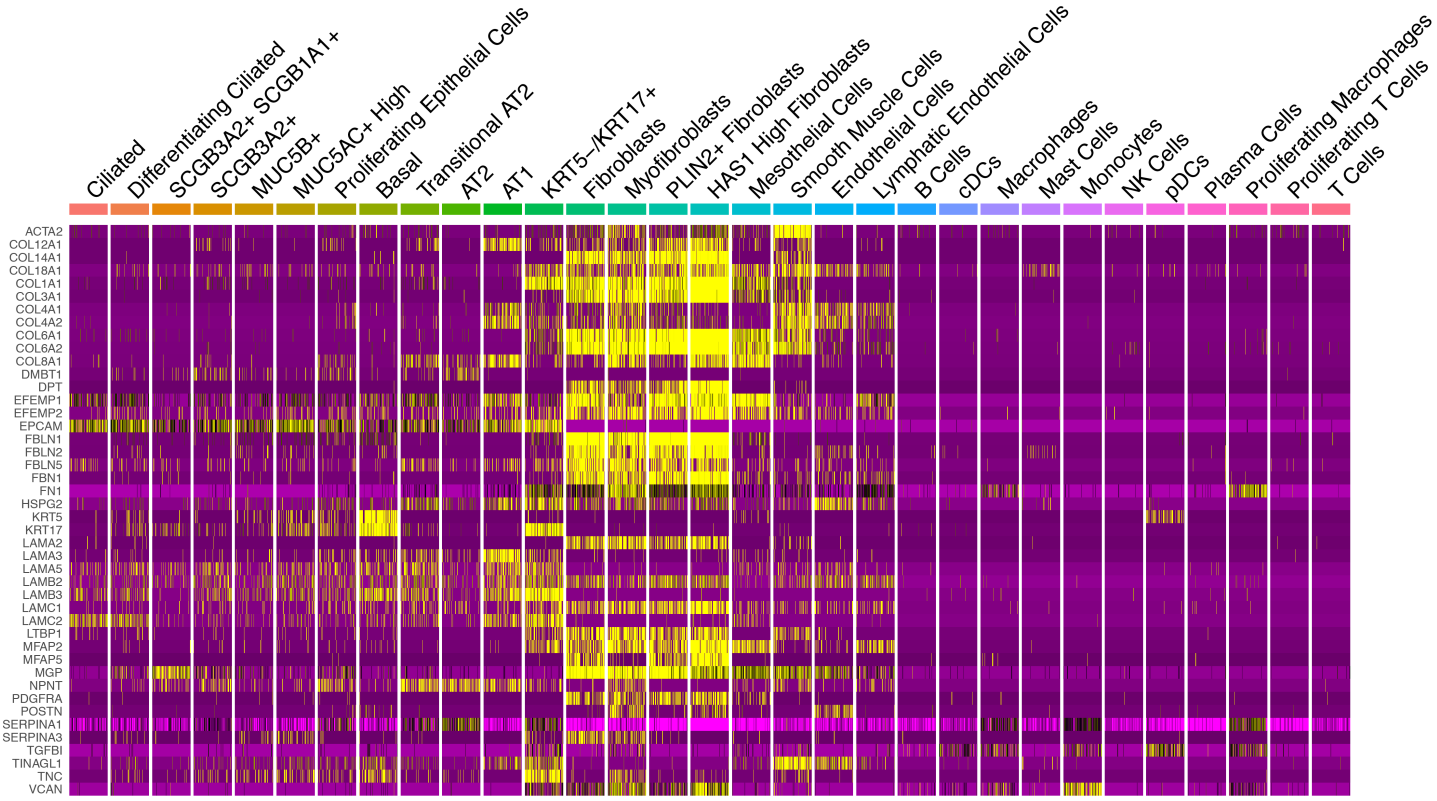


Figure S5. Heatmap demonstrating relative expression of IPF-enriched ECM components including all cell-types (expanded from Figure 1F). Data presented are normalized scaled Z scores, and downsampled to 100 cells per cell-type for visualization.

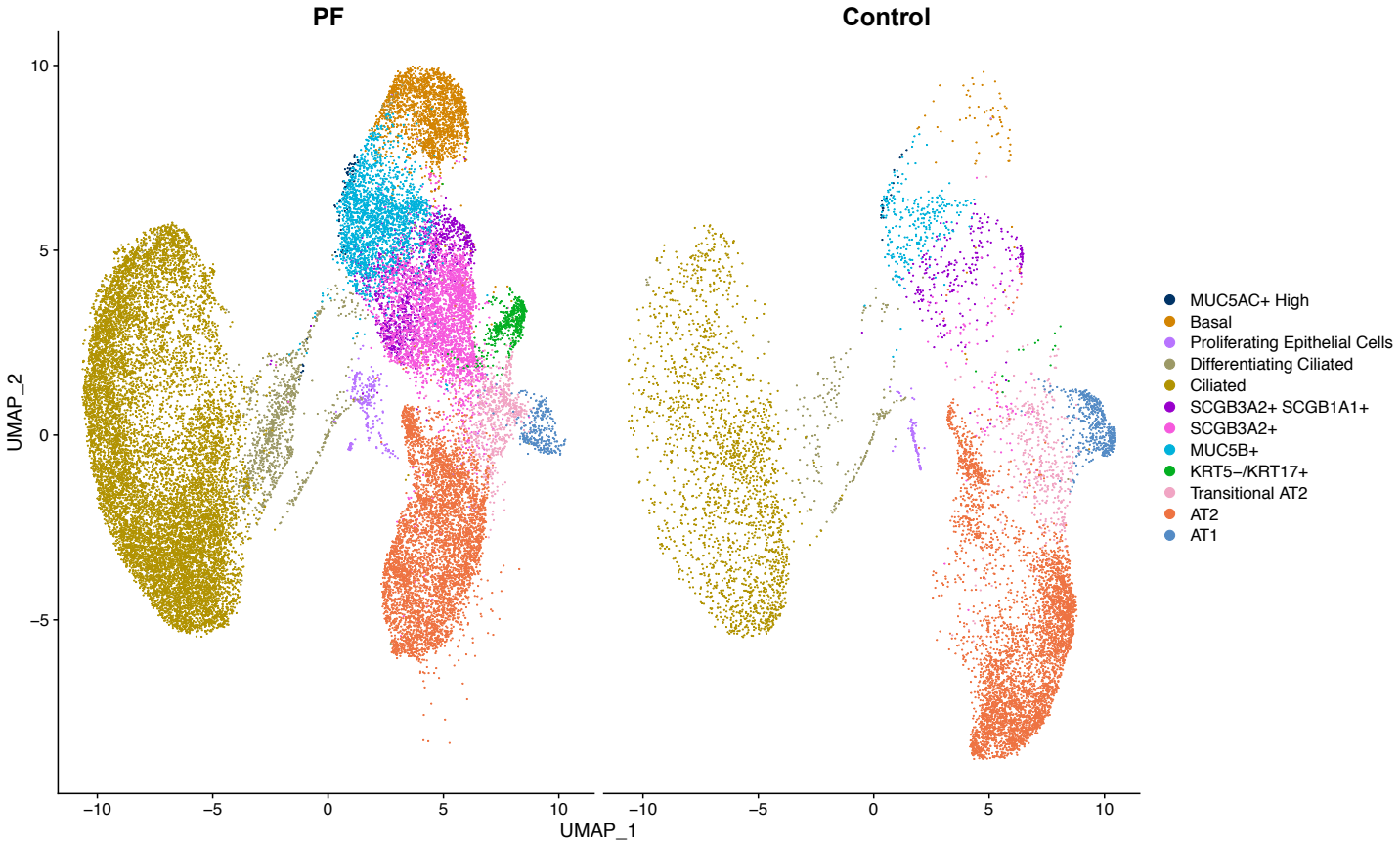


Figure S6. UMAP embedding demonstrating the distribution of epithelial cell types in PF and Control samples.

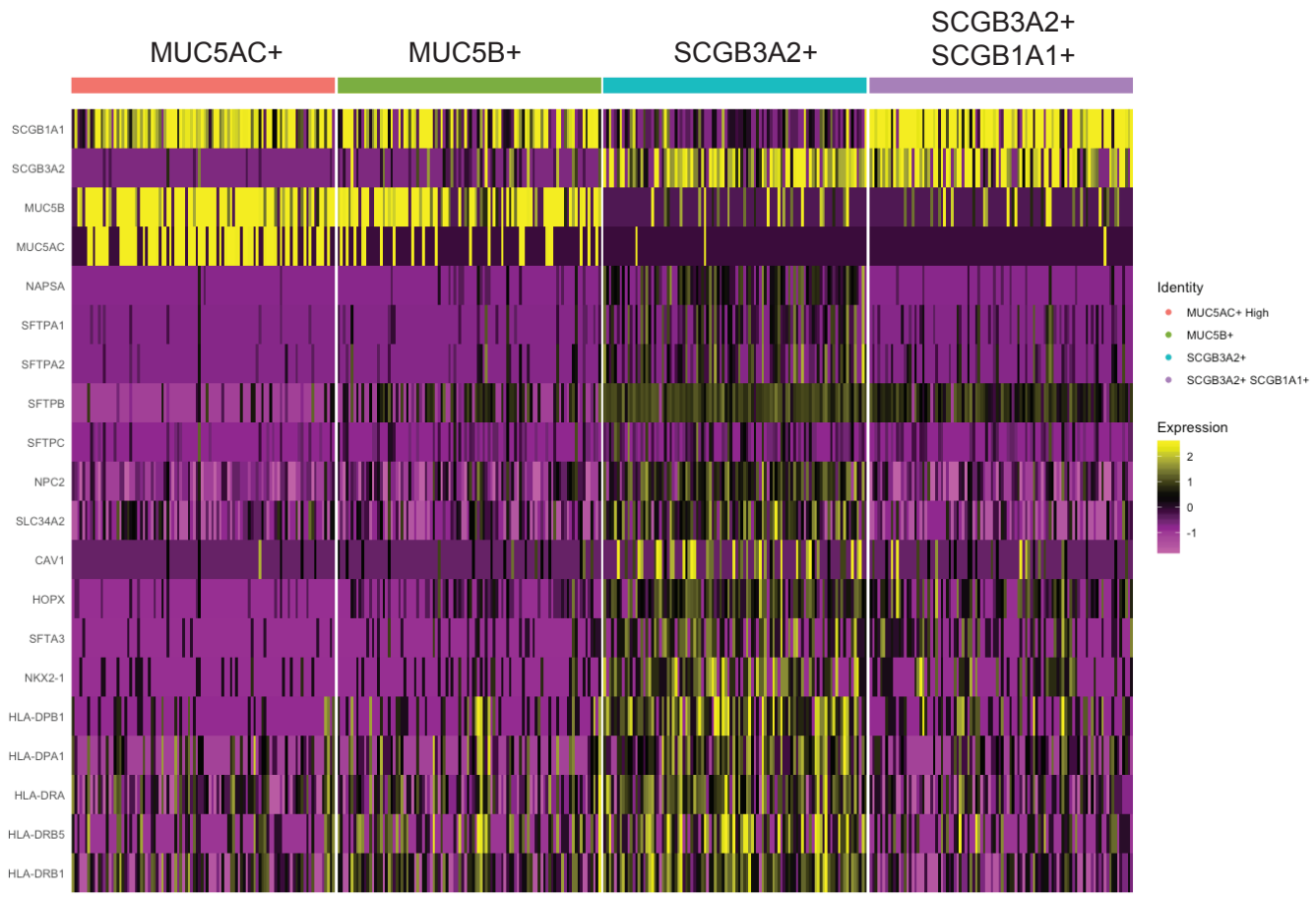


Figure S7. Heatmap of expression levels for selected markers that differ significantly (FDR<0.05, log FC > 0.4) between *SCGB3A2*+ and other secretory cell subtypes. Data presented are normalized scaled Z scores, and downsampled to 50 cells per cell-type for visualization.

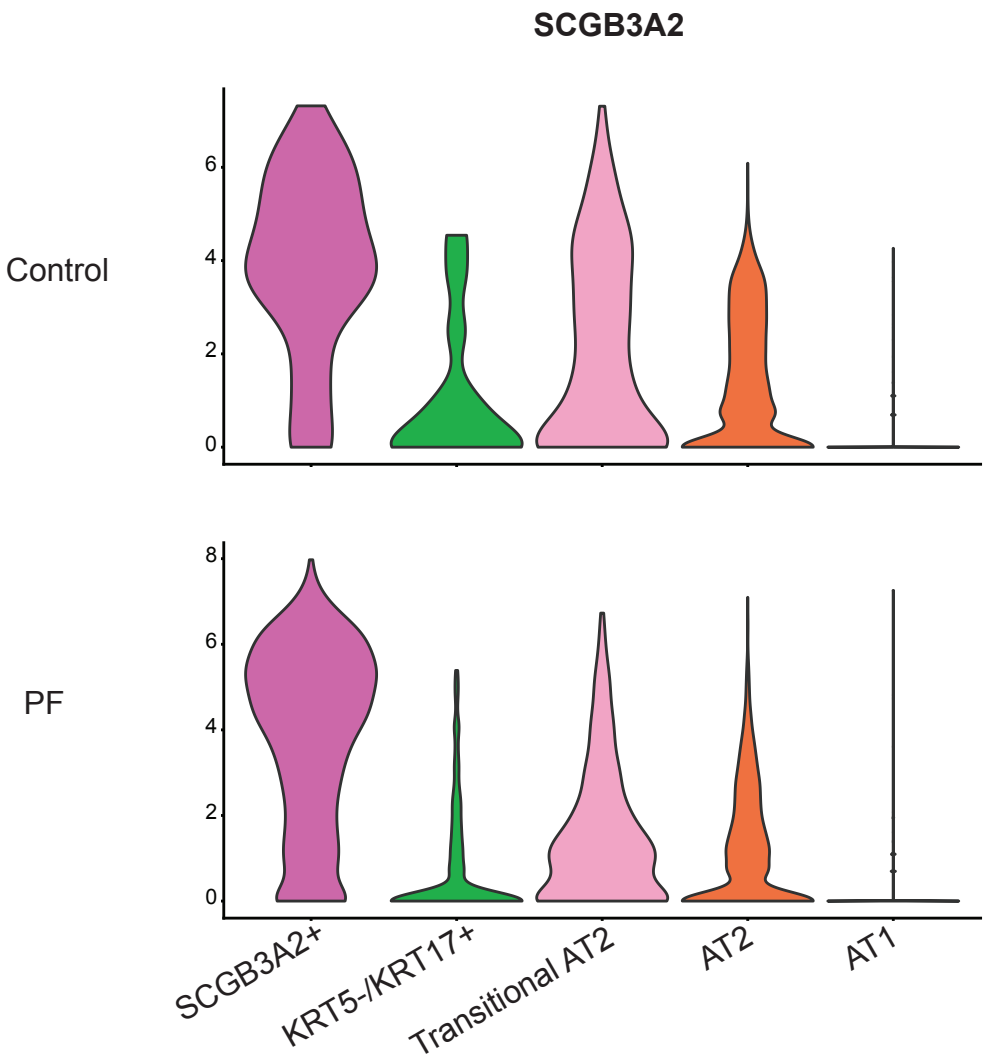


Figure S8. Violin plot of normalized expression for *SCGB3A2* in Transitional AT2, *SCGB3A2*$^{+}$, *KRT5*$^{-}$*/KRT17*$^{+}$, and AT1 cells in control and PF samples.

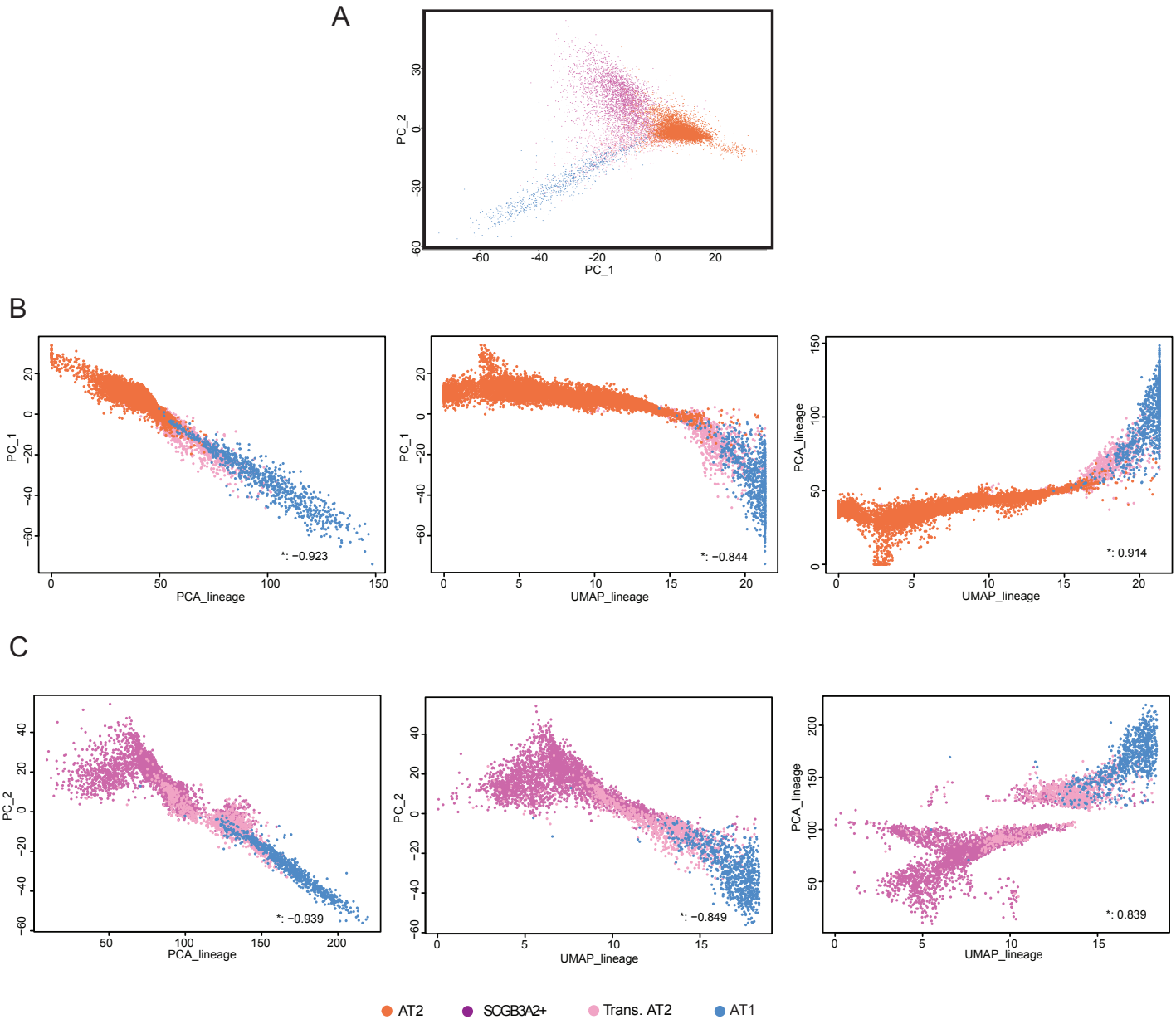


Figure S9. Correlation of cell order using alternative trajectory approaches (Fig. 3A). **A)** PCA plot of 14,462 cells from PF and control samples used to calculate alternative trajectories. Spearman correlation between linear cell ordering, PCA based slingshot, and UMAP based slingshot for the **B)** AT2 to AT1 trajectory and **C)** the *SCGB3A2*$^{+}$ trajectory.

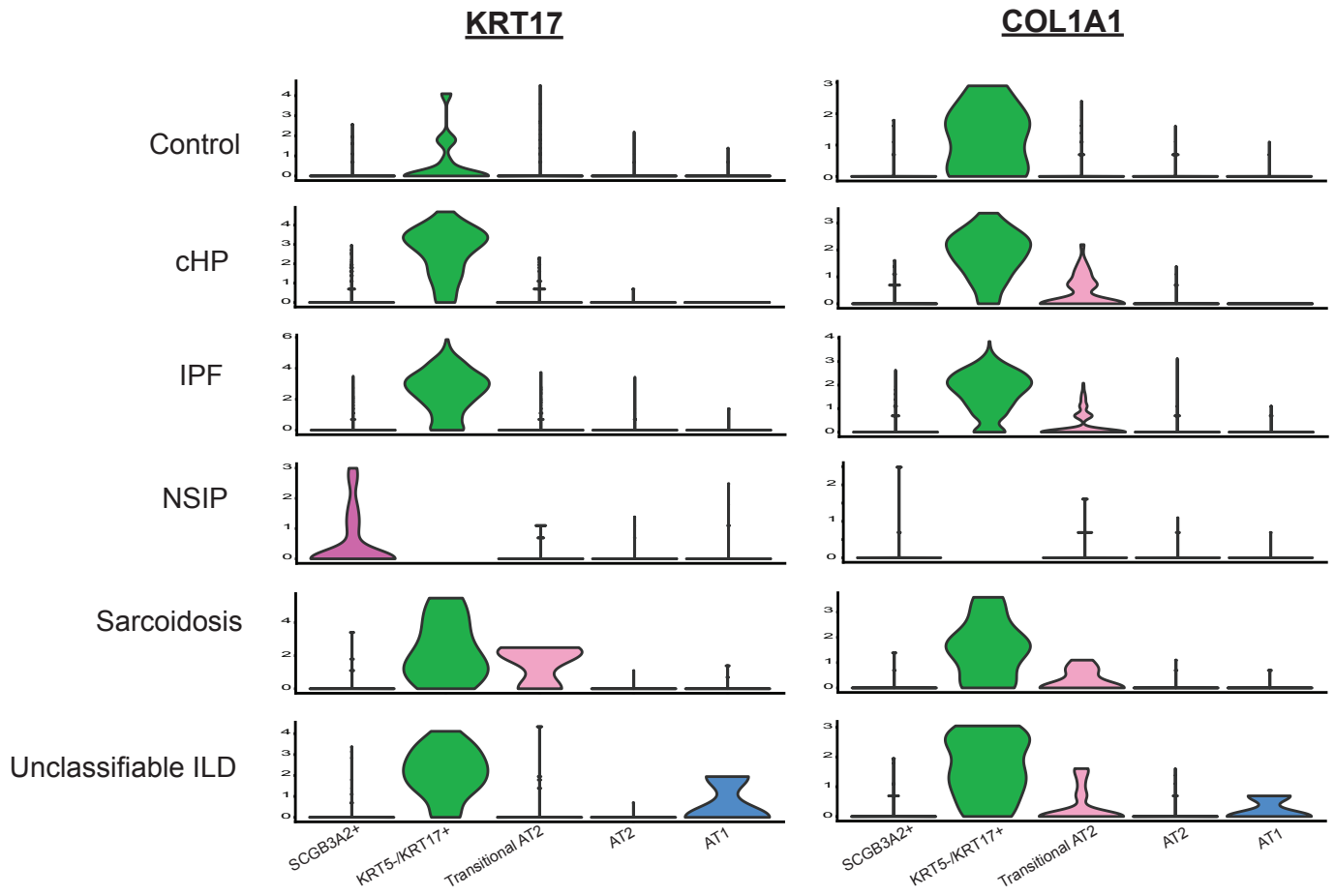


Figure S10. Violin plot of normalized expression for *COL1A1* in Transitional AT2, *SCGB3A2*$^{+}$, *KRT5*$^{-}$*/KRT17*$^{+}$, and AT1 cells comparing different histopathologic patterns of PF and control samples.



Figure S11. Representative high-resolution CT (HRCT) image from the asymptomatic subject from whom the transbronchial biopsy image in Fig. 3F is depicted. HRCT demonstrates minimal reticulation in the right lower lobe.

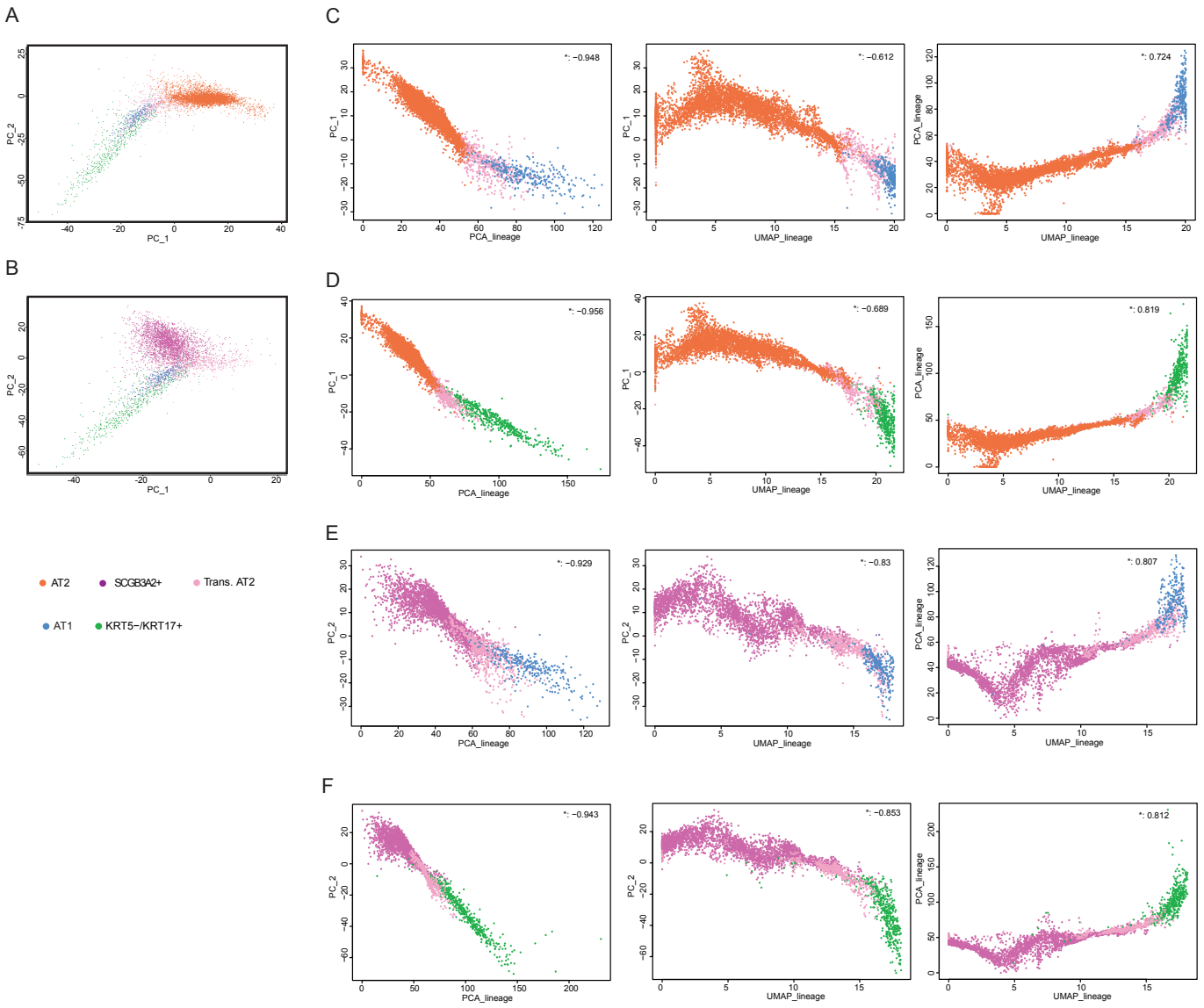


Figure S12. Correlation of cell order using alternative trajectory approaches (Fig. 5A-B). PCA plot of **A)** 6,406 AT2, Transitional AT2, AT1, and *KRT5*$^{-}$/*KRT17*$^{+}$ cells and **B)** 4,611 *SCGB3A2*$^{+}$, Transitional AT2, AT1 and KRT5-/KRT17+ cells from PF samples used to calculate alternative trajectories. The PCA plots were generated in the same manifold and then plot for the trajectory analysis. Spearman correlation between linear cell ordering, PCA based slingshot, and UMAP based slingshot for **C)** the AT2 to AT1 trajectory, **D)** the AT2 to *KRT5*$^{-}$/*KRT17*$^{+}$ trajectory, **E)** the *SCGB3A2*$^{+}$ to AT1 trajectory and **F)** the *SCGB3A2*$^{+}$ to *KRT5*$^{-}$/*KRT17*$^{+}$ trajectory.

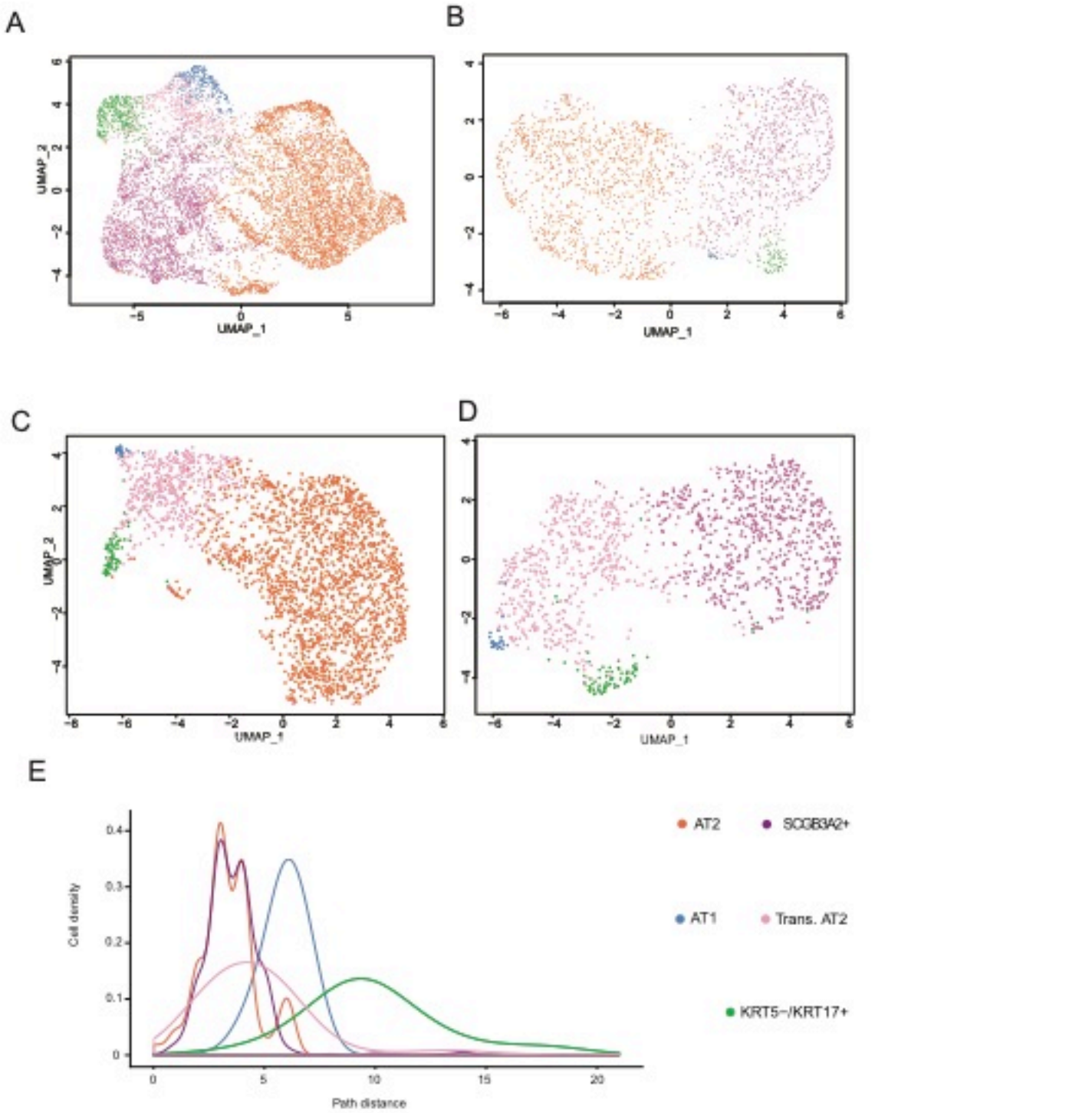


Figure S13. Relationship between cell types maintained at site and individual level. A UMAP manifold of AT2, *SCGB3A2*$^{+}$, Transitional AT2, AT1, and *KRT5*$^{-}$/*KRT17*$^{+}$ cell from PF samples processed at **A)** Vanderbilt or **B)** TGen. A UMAP manifold of Transitional AT2, AT1, and *KRT5*$^{-}$/*KRT17*$^{+}$ and either **C)** AT2 or **D)** *SCGB3A2*$^{+}$ cells from a single individual (VUILD59). The observed relationship between cell types in the UMAP manifold is consistent with the jointly analyzed data (Fig. 5A-B) regardless if the data are analyzed separately by site or even within a single individual. **E)** A complementary approach that does not rely on dimensionally reduction (TooManyCells) also found a consistent pattern. The X axis represents the distance from the start of the trajectory. Here AT2 and *SCGB3A2*$^{+}$ cells are largely overlapping near the start of the trajectory, the Transitional AT2 population splits between the AT2/*SCGB3A2*$^{+}$ and AT1 cell distribution extending out into the *KRT5*$^{-}$/*KRT17*$^{+}$ distribution.

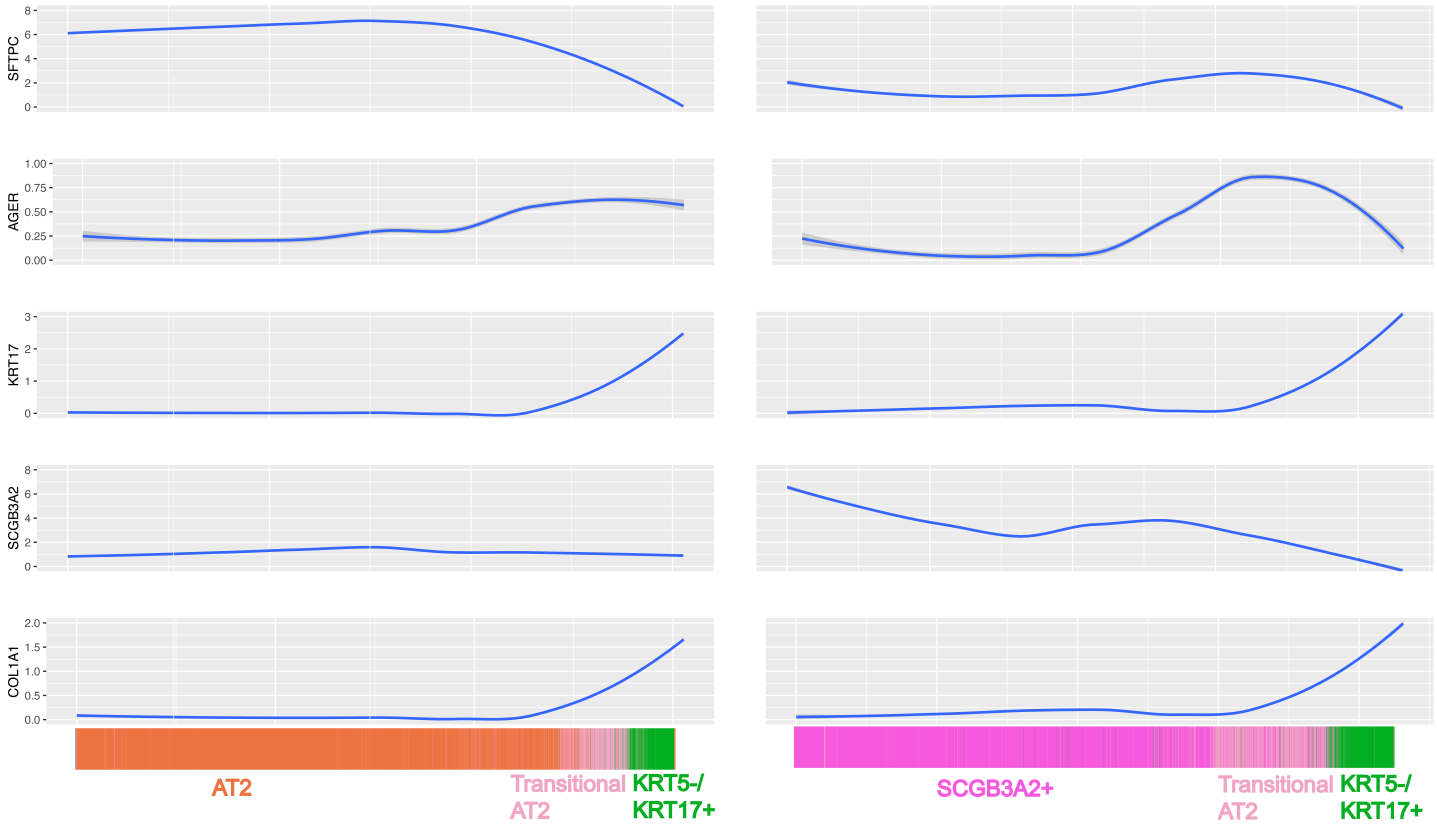


Figure S14. Smoothed expression of key marker genes across the pseudotime trajectories for *KRT5-/KRT17+* in PF cells (same pseudotime trajectories as in Fig. 5) .

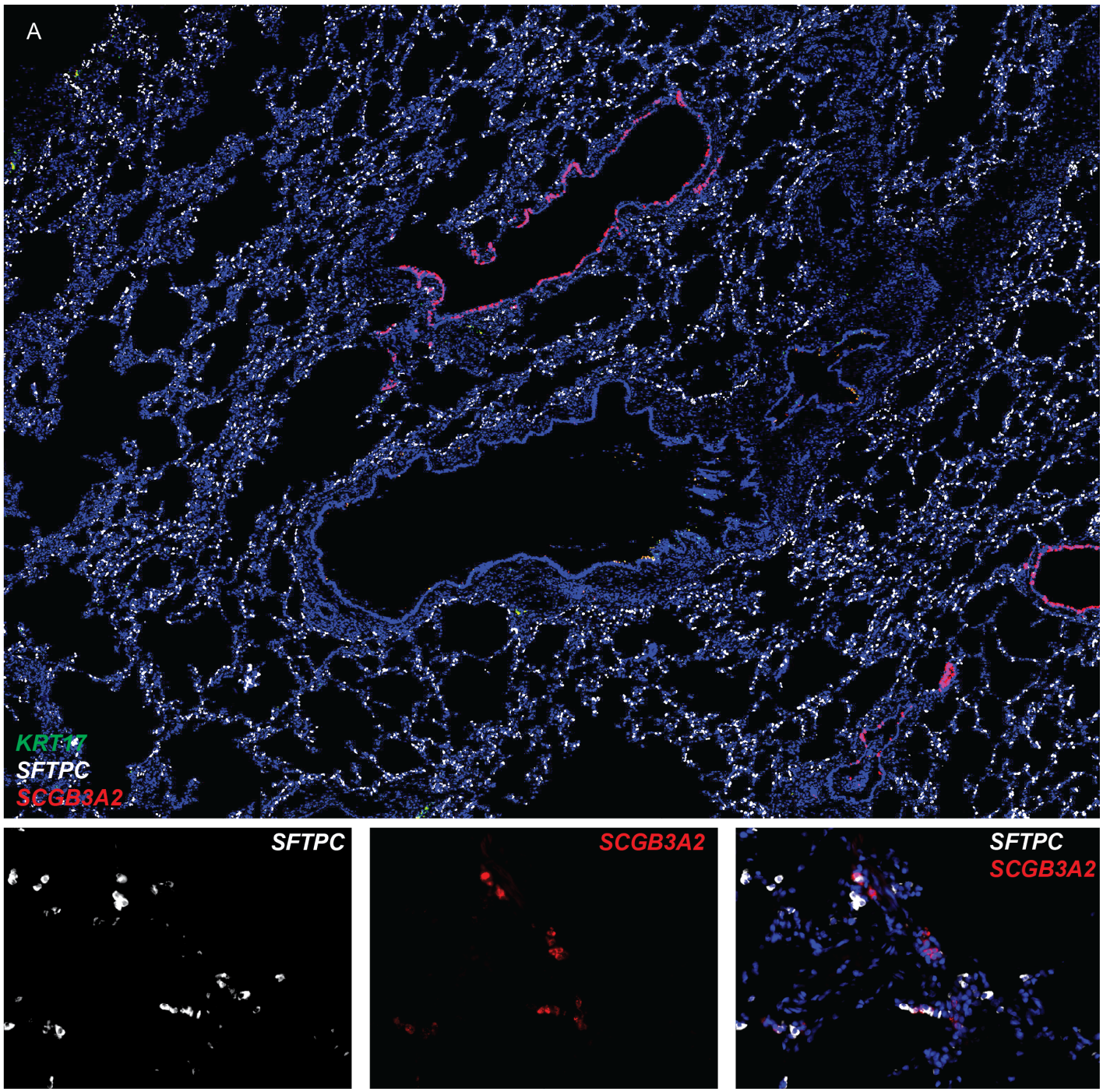


Figure S15. A) Multiplexed RNA-ISH demonstrating distribution of SCGB3A2+, SFTPC+ and KRT17+ cells in control lung tissue. Image is composed of stitched images captured at 100X objective. B) Multiplexed RNA-ISH demonstrating rare *SCGB3A2*$^{+}$/*SFTPC*$^{+}$ dual-positive cells in control lungs (original magnification 400x).

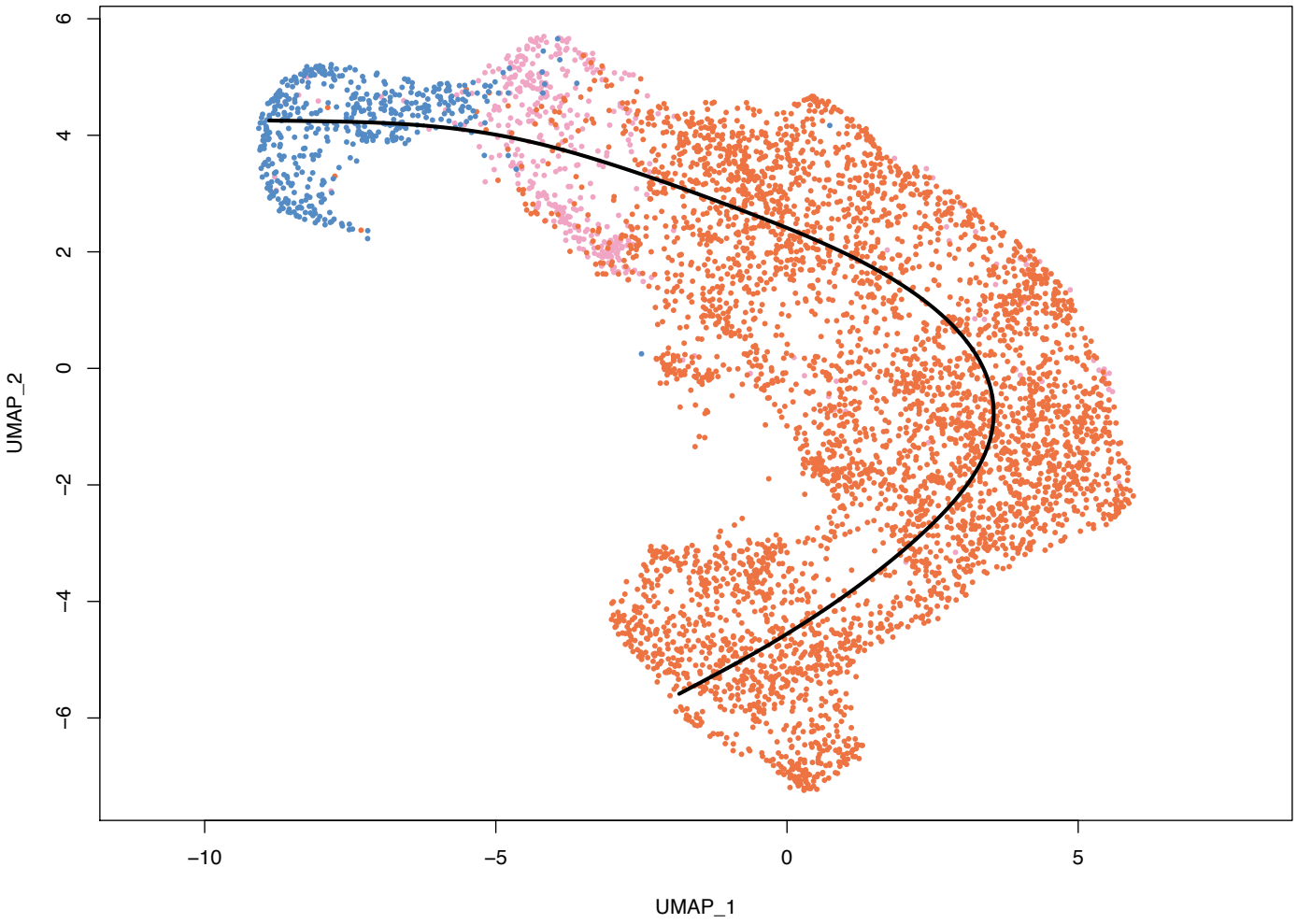


Figure S16. Slingshot based pseudotime of AT2 to AT1 cells using only cells from control samples. This trajectory is used in Fig. 5F.

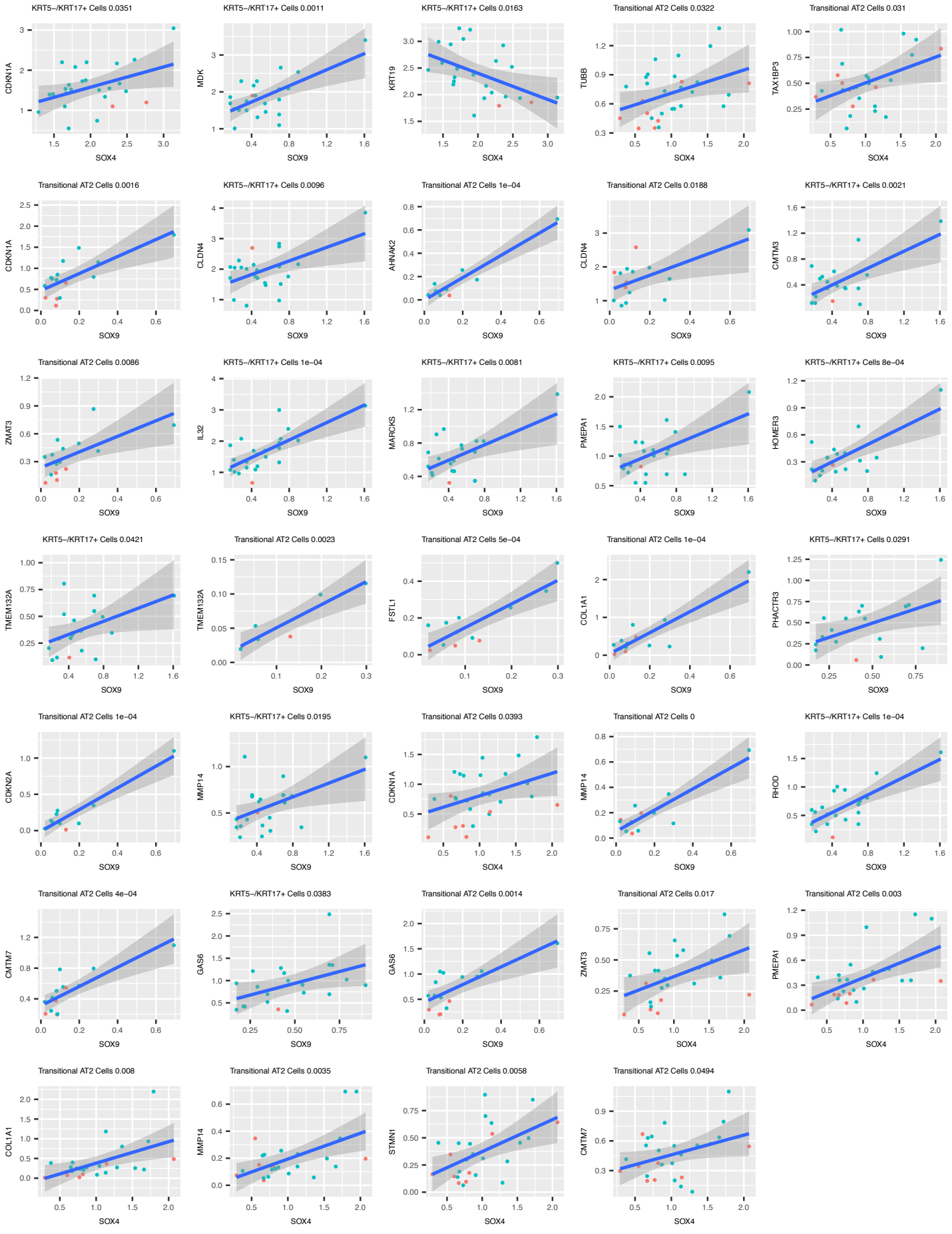

KRT5−/KRT17+ Cells 0.0351
CDKN1A
SOX4
KRT5−/KRT17+ Cells 0.0011
MDK
SOX9
KRT5−/KRT17+ Cells 0.0163
KRT19
SOX4
Transitional AT2 Cells 0.0322
TUBB
SOX4
Transitional AT2 Cells 0.031
TAX1BP3
SOX4
Transitional AT2 Cells 0.0016
CDKN1A
SOX9
KRT5−/KRT17+ Cells 0.0096
CLDN4
SOX9
Transitional AT2 Cells 1e−04
AHNAK2
SOX9
Transitional AT2 Cells 0.0188
CLDN4
SOX9
KRT5−/KRT17+ Cells 0.0021
CMTM3
SOX9
Transitional AT2 Cells 0.0086
ZMAT3
SOX9
KRT5−/KRT17+ Cells 1e−04
IL32
SOX9
KRT5−/KRT17+ Cells 0.0081
MARCKS
SOX9
KRT5−/KRT17+ Cells 0.0095
PMEPA1
SOX9
KRT5−/KRT17+ Cells 8e−04
HOMER3
SOX9
KRT5−/KRT17+ Cells 0.0421
TMEM132A
SOX9
Transitional AT2 Cells 0.0023
TMEM132A
SOX9
Transitional AT2 Cells 5e−04
FSTL1
SOX9
Transitional AT2 Cells 1e−04
COL1A1
SOX9
KRT5−/KRT17+ Cells 0.0291
PHACTR3
SOX9
Transitional AT2 Cells 1e−04
CDKN2A
SOX9
KRT5−/KRT17+ Cells 0.0195
MMP14
SOX9
Transitional AT2 Cells 0.0393
CDKN1A
SOX4
Transitional AT2 Cells 0
MMP14
SOX9
KRT5−/KRT17+ Cells 1e−04
RHOD
SOX9
Transitional AT2 Cells 4e−04
CMTM7
SOX9
KRT5−/KRT17+ Cells 0.0383
GAS6
SOX9
Transitional AT2 Cells 0.0014
GAS6
SOX9
Transitional AT2 Cells 0.017
ZMAT3
SOX4
Transitional AT2 Cells 0.003
PMEPA1
SOX4
Transitional AT2 Cells 0.008
COL1A1
SOX4
Transitional AT2 Cells 0.0035
MMP14
SOX4
Transitional AT2 Cells 0.0058
STMN1
SOX4
Transitional AT2 Cells 0.0494
CMTM7
SOX4

Figure S17. Significant correlation between the transcription factors SOX4 and SOX9 and genes containing motifs for those transcription factors in the Transitional AT2 and KRT5-/KRT17+ cells across individuals. The number above each plot represents the *P* value calculated using a linear regression.

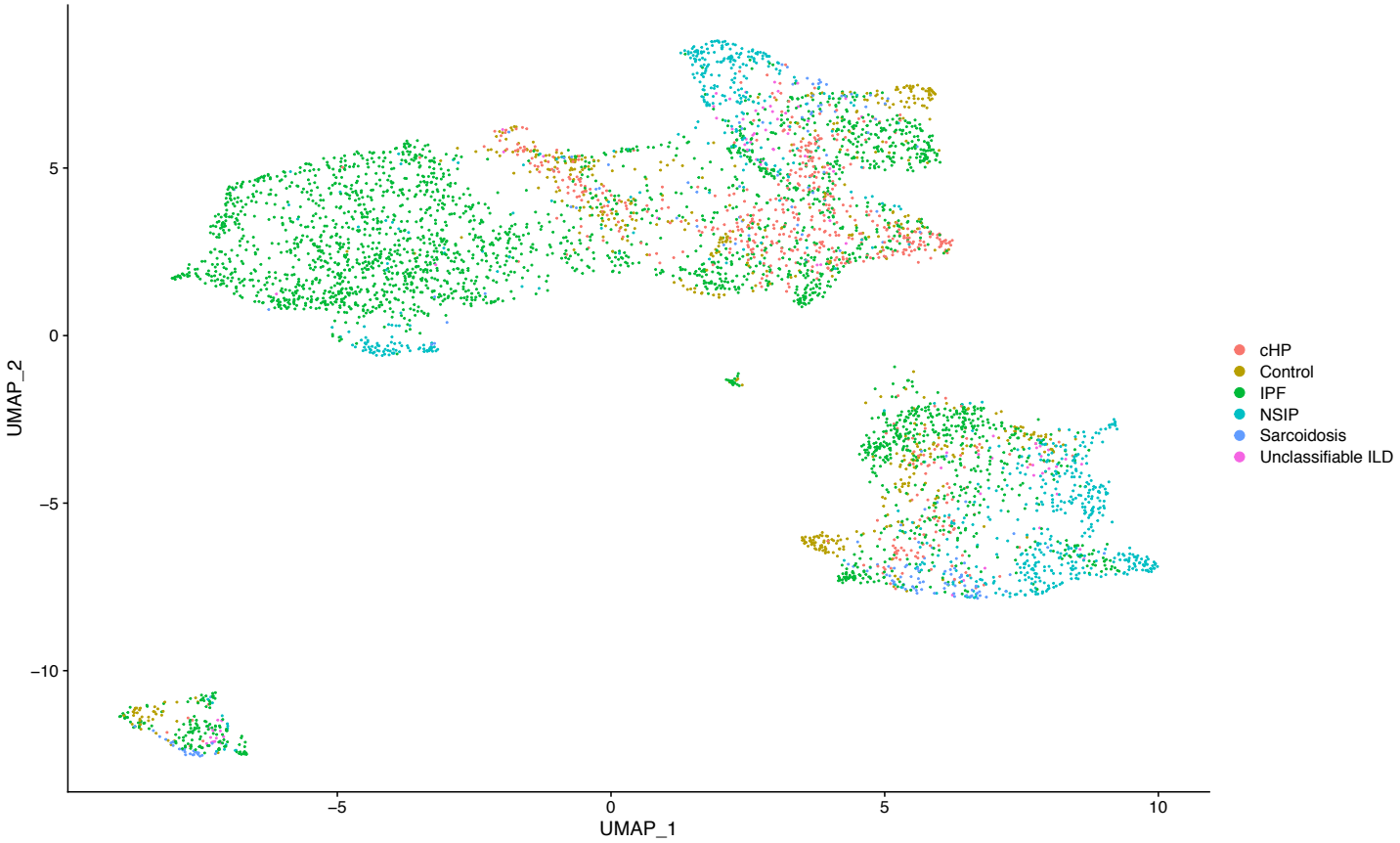


Figure S18. Mesenchymal cell types annotated by histopathologic pattern of PF.

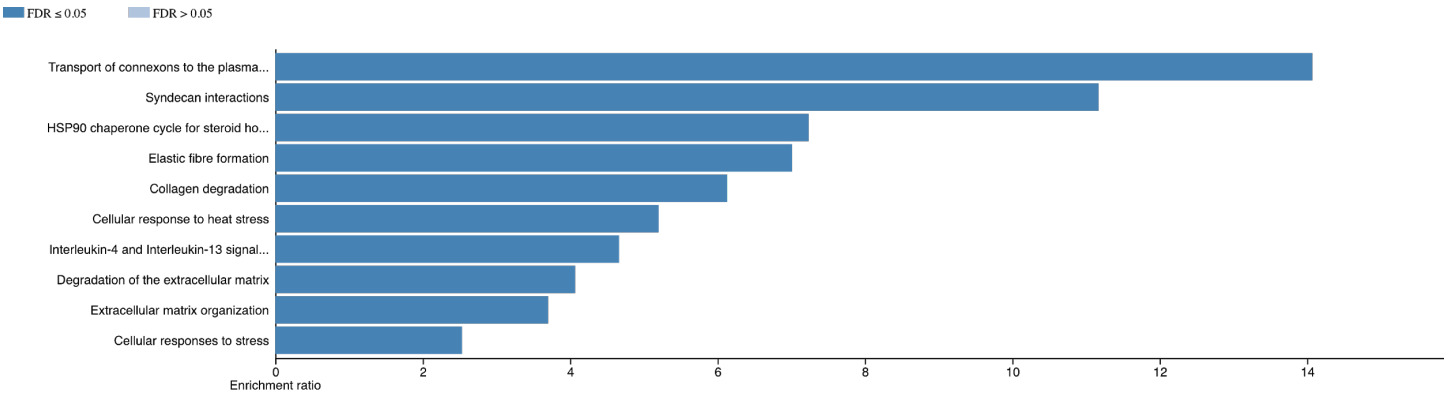
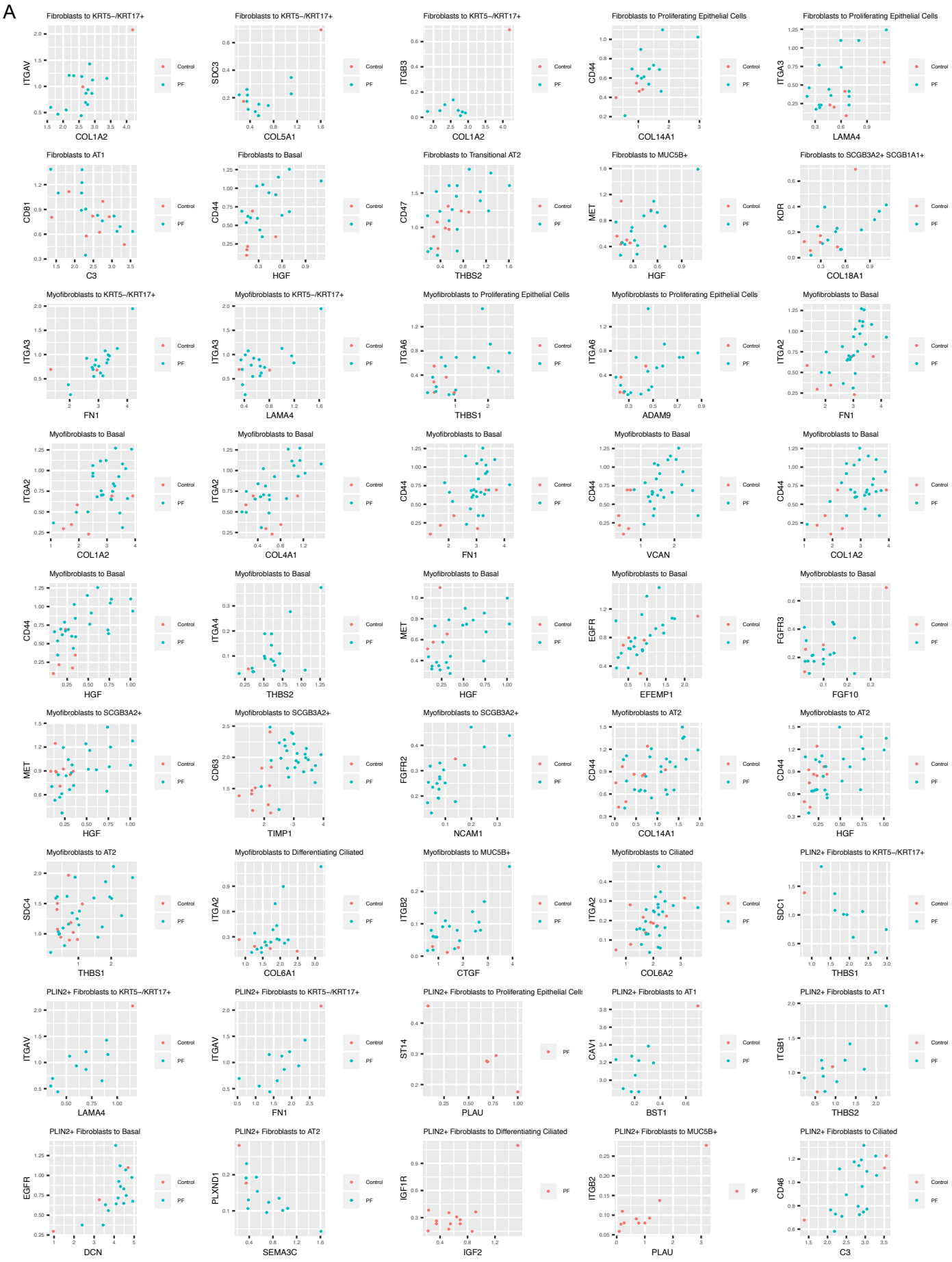


Figure S19. Top pathways enriched from 444 genes significantly upregulated in *HAS1*^{hi} fibroblasts compared to other mesenchymal cells.


A
Fibroblasts to KRT5−/KRT17+
ITGAV
COL1A2
Fibroblasts to KRT5−/KRT17+
SDC3
COL5A1
Fibroblasts to KRT5−/KRT17+
ITGB3
COL1A2
Fibroblasts to Proliferating Epithelial Cells
CD44
COL14A1
Fibroblasts to Proliferating Epithelial Cells
ITGA3
LAMA4
Fibroblasts to AT1
CD81
C3
Fibroblasts to Basal
CD44
HGF
Fibroblasts to Transitional AT2
CD47
THBS2
Fibroblasts to MUC5B+
MET
HGF
Fibroblasts to SCGB3A2+ SCGB1A1+
KDR
COL18A1
Myofibroblasts to KRT5−/KRT17+
ITGA3
FN1
Myofibroblasts to KRT5−/KRT17+
ITGA3
LAMA4
Myofibroblasts to Proliferating Epithelial Cells
ITGA6
THBS1
Myofibroblasts to Proliferating Epithelial Cells
ITGA6
ADAM9
Myofibroblasts to Basal
ITGA2
FN1
Myofibroblasts to Basal
ITGA2
COL1A2
Myofibroblasts to Basal
ITGA2
COL4A1
Myofibroblasts to Basal
CD44
FN1
Myofibroblasts to Basal
CD44
VCAN
Myofibroblasts to Basal
CD44
COL1A2
Myofibroblasts to Basal
CD44
HGF
Myofibroblasts to Basal
ITGA4
THBS2
Myofibroblasts to Basal
MET
HGF
Myofibroblasts to Basal
EGFR
EFEMP1
Myofibroblasts to Basal
FGFR3
FGF10
Myofibroblasts to SCGB3A2+
MET
HGF
Myofibroblasts to SCGB3A2+
CD63
TIMP1
Myofibroblasts to SCGB3A2+
FGFR2
NCAM1
Myofibroblasts to AT2
CD44
COL14A1
Myofibroblasts to AT2
CD44
HGF
Myofibroblasts to AT2
SDC4
THBS1
Myofibroblasts to Differentiating Ciliated
ITGA2
COL6A1
Myofibroblasts to MUC5B+
ITGB2
CTGF
Myofibroblasts to Ciliated
ITGA2
COL6A2
PLIN2+ Fibroblasts to KRT5−/KRT17+
SDC1
THBS1
PLIN2+ Fibroblasts to KRT5−/KRT17+
ITGAV
LAMA4
PLIN2+ Fibroblasts to KRT5−/KRT17+
ITGAV
FN1
PLIN2+ Fibroblasts to Proliferating Epithelial Cells
ST14
PLAU
PLIN2+ Fibroblasts to AT1
CAV1
BST1
PLIN2+ Fibroblasts to AT1
ITGB1
THBS2
PLIN2+ Fibroblasts to Basal
EGFR
DCN
PLIN2+ Fibroblasts to AT2
PLXND1
SEMA3C
PLIN2+ Fibroblasts to Differentiating Ciliated
IGF1R
IGF2
PLIN2+ Fibroblasts to MUC5B+
ITGB2
PLAU
PLIN2+ Fibroblasts to Ciliated
CD46
C3
Control
PF

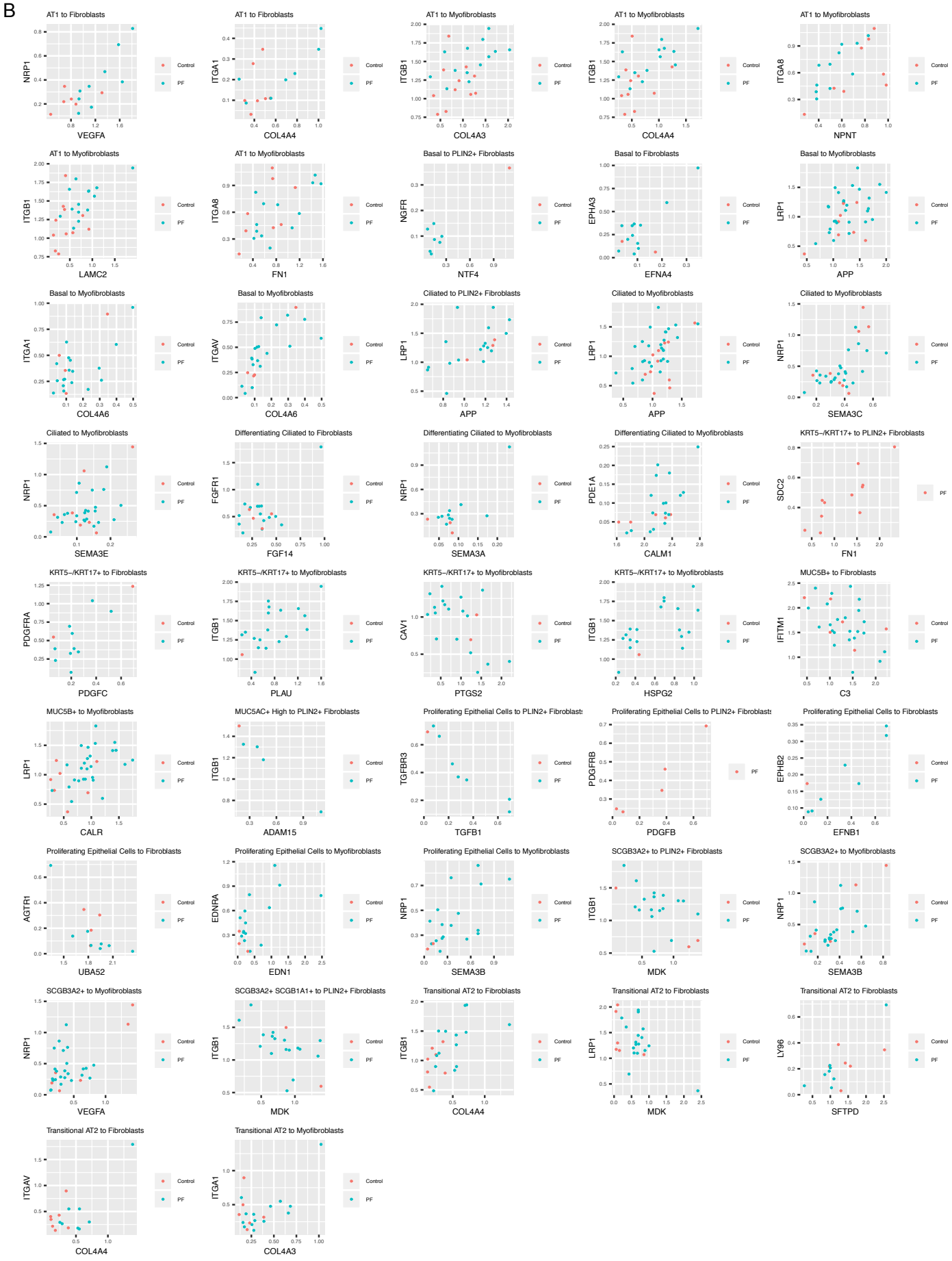

B
AT1 to Fibroblasts
NRP1
VEGFA
AT1 to Fibroblasts
ITGA1
COL4A4
AT1 to Myofibroblasts
ITGB1
COL4A3
AT1 to Myofibroblasts
ITGB1
COL4A4
AT1 to Myofibroblasts
ITGA8
NPNT
AT1 to Myofibroblasts
ITGB1
LAMC2
AT1 to Myofibroblasts
ITGA8
FN1
Basal to PLIN2+ Fibroblasts
NGFR
NTF4
Basal to Fibroblasts
EPHA3
EFNA4
Basal to Myofibroblasts
LRP1
APP
Basal to Myofibroblasts
ITGA1
COL4A6
Basal to Myofibroblasts
ITGAV
COL4A6
Ciliated to PLIN2+ Fibroblasts
LRP1
APP
Ciliated to Myofibroblasts
LRP1
APP
Ciliated to Myofibroblasts
NRP1
SEMA3C
Ciliated to Myofibroblasts
NRP1
SEMA3E
Differentiating Ciliated to Fibroblasts
FGFR1
FGF14
Differentiating Ciliated to Myofibroblasts
NRP1
SEMA3A
Differentiating Ciliated to Myofibroblasts
PDE1A
CALM1
KRT5−/KRT17+ to PLIN2+ Fibroblasts
SDC2
FN1
KRT5−/KRT17+ to Fibroblasts
PDGFRA
PDGFC
KRT5−/KRT17+ to Myofibroblasts
ITGB1
PLAU
KRT5−/KRT17+ to Myofibroblasts
CAV1
PTGS2
KRT5−/KRT17+ to Myofibroblasts
ITGB1
HSPG2
MUC5B+ to Fibroblasts
IFITM1
C3
MUC5B+ to Myofibroblasts
LRP1
CALR
MUC5AC+ High to PLIN2+ Fibroblasts
ITGB1
ADAM15
Proliferating Epithelial Cells to PLIN2+ Fibroblasts
TGFBR3
TGFB1
Proliferating Epithelial Cells to PLIN2+ Fibroblasts
PDGFRB
PDGFB
Proliferating Epithelial Cells to Fibroblasts
EPHB2
EFNB1
Proliferating Epithelial Cells to Fibroblasts
AGTR1
UBA52
Proliferating Epithelial Cells to Myofibroblasts
EDNRA
EDN1
Proliferating Epithelial Cells to Myofibroblasts
NRP1
SEMA3B
SCGB3A2+ to PLIN2+ Fibroblasts
ITGB1
MDK
SCGB3A2+ to Myofibroblasts
NRP1
SEMA3B
SCGB3A2+ to Myofibroblasts
NRP1
VEGFA
SCGB3A2+ SCGB1A1+ to PLIN2+ Fibroblasts
ITGB1
MDK
Transitional AT2 to Fibroblasts
ITGB1
COL4A4
Transitional AT2 to Fibroblasts
LRP1
MDK
Transitional AT2 to Fibroblasts
LY96
SFTPD
Transitional AT2 to Fibroblasts
ITGAV
COL4A4
Transitional AT2 to Myofibroblasts
ITGA1
COL4A3
Control
PF

Figure S20. Significantly correlated ligand-receptor pair gene expression within individuals for mesenchymal ligands to epithelial receptors (A) and epithelial ligands to mesenchymal receptors (B).

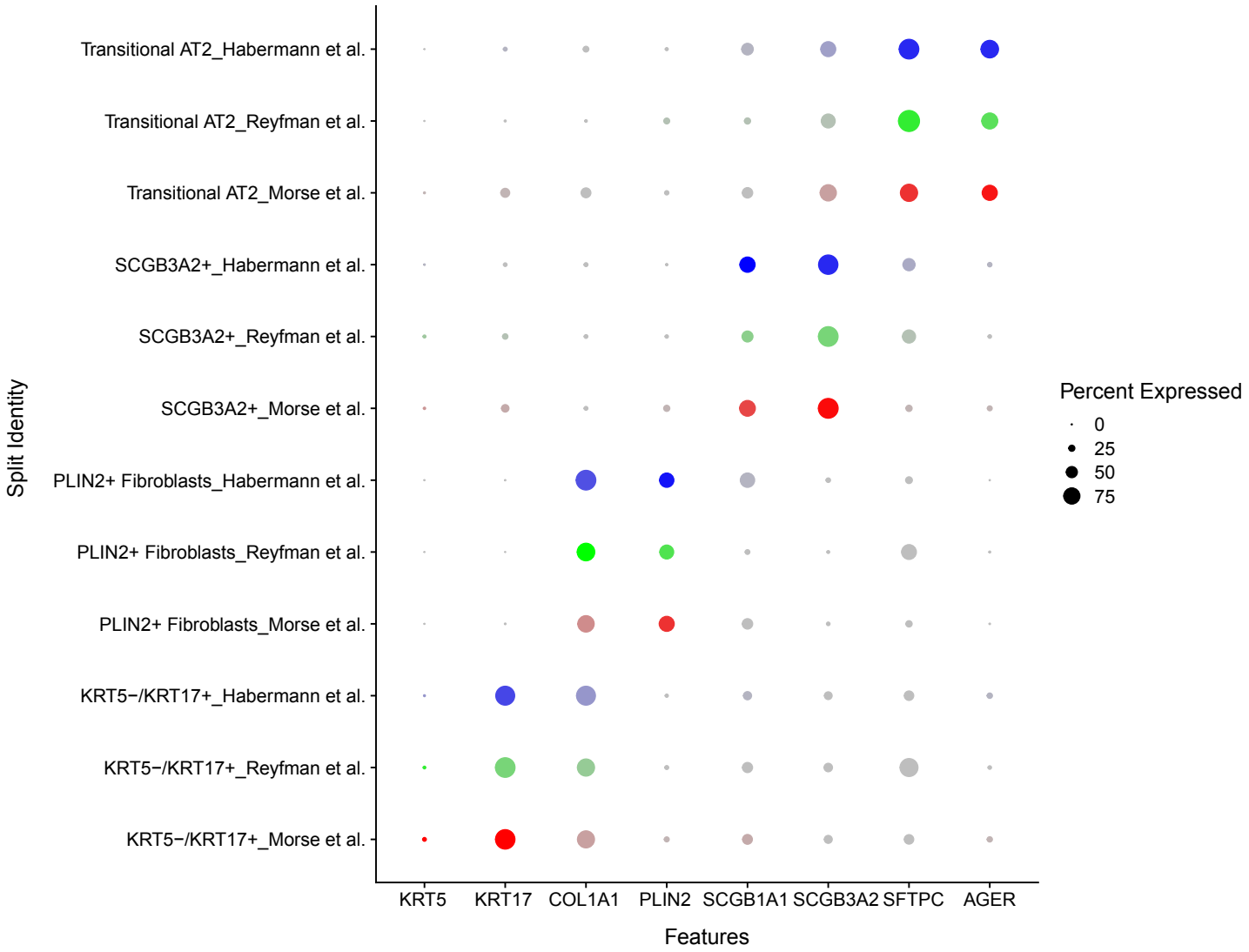


Figure S21. Expression levels of key marker genes used for the four novel cell types described in this study (Habermann *et al.*, blue) compared to recently published dataset (Morse *et al.*, 2019 (red) and Reyfman *et al.*, 2019 (green)). Color intensity indicates relative gene expression level.

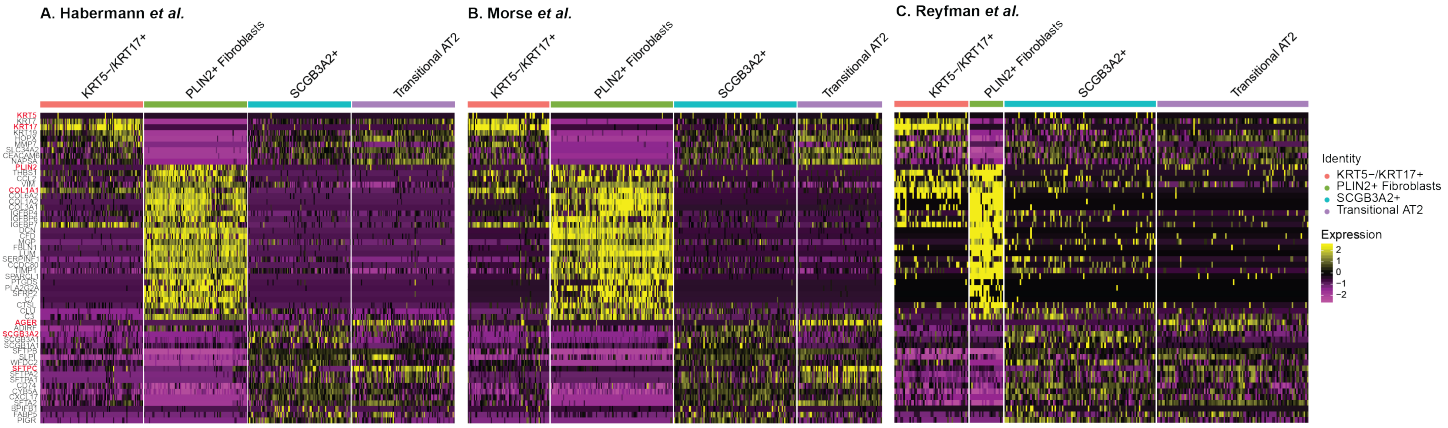


Figure S22. A heatmap showing expression of top differentially expressed genes (see methods) in each cell type in this study (Habermann *et al*.) and two recently published datasets (Morse *et al.* and Reyfman *et al.*). Key marker genes for each cell type are highlighted in red.

	Control	PF
	n=10	n=20
Age	37 [17 - 59]	61.5 [46 -72]
Sex (male)	7 (70%)	13 (65%)
Tobacco	8 (80%)	11 (55%)
Ancestry		
European	7 (70%)	14 (70%)
African	1 (10%)	4 (20%)
Hispanic		2 (10%)
Other/Unknown	2 (20%)	1 (5%)
Diagnosis		
IPF		12 (60%)
NSIP		3 (15%)
cHP		2 (10%)
Sarcoidosis		2 (10%)
Unclassifiable ILD		1 (5%)

Table S1. Demographics of lung donors used for scRNA-seq.

Cell Type	Positive Gene Markers	Negative Gene Markers
Epithelial Cells	EPCAM	
AT1	AGER, PDPN	
AT2	SFTPC, ABCA3, SFTPCD	
Transitional AT2	SFTPC (low), AGER (low)	
Basal	KRT5, KRT17	COL1A1
KRT5-/KRT17+	KRT17, COL1A1	KRT5
MUC5B+	MUC5B, SCGB1A1	
MUC5AC+ High	MUC5AC	
SCGB3A2+	SCGB3A2	SCGB1A1
SCGB3A2+ SCGB1A1+	MGP, SCGB1A1, SCGB3A2	
Ciliated	FOXJ1, TMEM190, CAPS, HYDIN	
Differentiating Ciliated	FOXJ1, SFTPB	
Proliferating Epithelial Cells	MKI67, CDK1	
Immune Cells	PTPRC	
T Cells	CD3E, FOXP3, IL7R, CD8A, CCL5	
NK Cells	NCR1, KLRB1, NKG7 (high)	CD3E
Macrophages	LYZ, MARCO, FCGR1A, C1QA, APOC1	
Monocytes	S100A12, FCN1, S100A9, LYZ, CD14	
cDCs	FCER1A, CD1C, CLEC9A	
pDCs	LILRA4, CLEC4C, JCHAIN	
Plasma Cells	JCHAIN, IGHG1, IGLL5	
B Cells	MS4A1, CD19, CD79A	
Mast Cells	CPA3, KIT	
Proliferating T Cells	MKI67, CDK1, CD3E	
Proliferating Macrophages	MKI67, CDK1, LYZ	
Endothelial Cells	PECAM1	PTPRC
Endothelial Cells	VWF, PECAM1	
Lymphatic Endothelial Cells	CCL21	
Mesenchymal Cells		EPCAM, PTPRC, PECAM1
Smooth Muscle Cells	ACTA2 (high), MYH11, PDGFRB (high)	LUM, PDGFRA
Mesothelial Cells	WT1, UPK3B	LUM
Myofibroblasts	LUM, PDGFRA, ACTA2, MYLK	
HAS1 High Fibroblasts	LUM, PDGFRA, HAS1, TWIST	PLIN2
Fibroblasts	LUM, PDGFRA	
PLIN2+ Fibroblasts	LUM, PDGFRA, PLIN2	

Table S3. Marker genes used for cell-type annotation.

	IPF	Control	NSIP	cHP	Unclassifiable ILD	Sarcoidosis	Total Cells
MUC5AC+ High	75	20	7	1	-	-	103
Basal	1,786	65	61	107	37	2	2,058
Proliferating Epithelial Cells	222	79	9	20	4	7	341
Differentiating Ciliated	1,057	226	8	54	5	12	1,362
Ciliated	11,857	2,119	169	424	26	76	14,671
SCGB3A2+ SCGB1A1+	854	222	58	34	27	215	1,410
SCGB3A2+	2,824	97	20	145	108	26	3,220
MUC5B+	1,940	281	42	112	55	3	2,433
KRT5-/KRT17+	374	13	1	62	26	9	485
Transitional AT2	581	443	21	84	22	9	1,160
AT2	3,225	4,393	299	382	498	514	9,311
AT1	96	472	158	15	3	27	771
Lymphatic Endothelial Cells	584	395	80	68	4	72	1,203
Endothelial Cells	3,496	2,761	1,205	694	290	797	9,243
Proliferating Macrophages	813	200	188	293	53	16	1,563
Proliferating T Cells	132	301	11	22	2	22	490
NK Cells	957	712	383	74	7	189	2,322
pDCs	109	17	27	19	3	13	188
cDCs	1,061	151	373	224	154	169	2,132
Mast Cells	360	254	10	6	36	35	701
Plasma Cells	265	238	45	45	2	9	604
B Cells	550	140	284	13	4	6	997
Monocytes	2,113	2,854	354	257	674	513	6,765
T Cells	2,829	2,622	726	304	46	176	6,703
Macrophages	16,507	12,004	3,299	3,492	1,827	1,799	38,928
PLIN2+ Fibroblasts	1,167	6	98	6	3	7	1,287
Fibroblasts	152	140	19	102	3	11	427
HAS1 High Fibroblasts	179	-	-	-	-	-	179
Myofibroblasts	813	196	237	367	45	54	1,712
Mesothelial Cells	137	35	11	7	15	28	233
Smooth Muscle Cells	567	188	435	102	27	75	1,394
Total Cells	57,682	31,644	8,638	7,535	4,006	4,891	114,396

Table S4. Distribution of cell types (as a proportion of all cells) in PF compared to control lungs.

Cell Type	Control	Processing Site	Flowcell_ID	Processing Site + Flowcell_ID
MUC5AC+ High	555	555	312	316
Basal	235	235	235	235
Proliferating Epithelial Cells	720	719	720	717
Differentiating Ciliated	390	390	390	390
Ciliated	70	70	70	70
SCGB3A2+ SCGB1A1+	154	154	154	154
SCGB3A2+	266	266	266	266
MUC5B+	296	296	296	296
KRT5-/KRT17+	1083	1083	1083	1083
Transitional AT2	433	433	433	433
AT2	371	371	371	371
AT1	371	370	371	371
Lymphatic Endothelial Cells	195	195	195	195
Endothelial Cells	339	339	339	339
Proliferating Macrophages	202	202	202	202
Proliferating T Cells	311	311	311	311
NK Cells	67	67	67	67
pDCs	684	684	684	684
cDCs	230	230	230	229
Mast Cells	190	190	190	190
Plasma Cells	512	512	512	509
B Cells	113	113	113	113
Monocytes	237	237	237	237
T Cells	161	161	161	161
Macrophages	182	182	182	182
PLIN2+ Fibroblasts	2034	2034	2033	2034
Fibroblasts	349	349	349	348
HAS1 High Fibroblasts	0	0	0	0
Myofibroblasts	383	383	383	383
Mesothelial Cells	693	693	693	693
Smooth Muscle Cells	404	404	404	404

Table S6. DE analysis comparison with and without latent_var correction for batch effects. Numbers indicate number of differential expressed genes of each cell type against all other cell types, with logFC $\geq$ 0.25 and FDR q_value < 0.01.

Cell Type	Habermann *et al.*	Morse *et al.*	Reyfman *et al.*
AT1	0.67	0.47	1.31
AT2	8.14	3.08	35.88
B Cells	0.87	0.43	0.48
Basal	1.80	2.47	1.26
cDCs	1.86	1.42	1.51
Ciliated	12.82	4.07	2.16
Differentiating Ciliated	1.19	0.98	0.81
Endothelial Cells	8.08	10.54	1.64
Fibroblasts	0.37	1.04	0.06
KRT5-/KRT17+	0.42	0.18	0.07
Lymphatic Endothelial Cells	1.05	1.55	0.50
Macrophages	34.03	51.19	42.16
Mast Cells	0.61	1.03	0.32
Mesothelial Cells	0.20	0.13	0.17
Monocytes	5.91	2.87	2.62
MUC5AC+ High	0.09	0.02	0.00
MUC5B+	2.13	1.75	0.53
Myofibroblasts	1.50	5.61	0.35
NK Cells	2.03	0.38	0.12
pDCs	0.16	0.11	0.11
Plasma Cells	0.53	0.31	1.02
PLIN2+ Fibroblasts	1.13	0.72	0.03
Proliferating Epithelial Cells	0.30	0.19	0.38
Proliferating Macrophages	1.37	1.45	1.33
Proliferating T Cells	0.43	0.15	0.17
SCGB3A2+	2.81	2.74	1.56
SCGB3A2+ SCGB1A1+	1.23	0.95	0.57
Smooth Muscle Cells	1.22	1.45	0.27
T Cells	5.86	2.55	1.12
Transitional AT2	1.01	0.19	1.49

Table S11. Comparison of cell proportion of individual cell type identified in this study (Habermann *et al.*), to two recently published studies (Morse *et al.* and Reyfman *et al.)*

Table S2. Sequencing QC metrics and demographic information. This table contains QC metrics from sequencing (output from 10XGenomisc cell ranger) and demographic information for each sample.

Table S9. SOX and NR1D1 motif distribution on promoter of genes in bin III and IV from the AT2_transition_to KRT5-/KRT17+ heatmap.

Table S10. The top 20% highly co-expressed ligand-receptor pairs for each epithelial-mesenchymal interaction from the top 20% highly expressed genes for each cell type (related to Fig. 4 and Fig. S21).

See zip file

Table S5. Differentially expressed genes.

Mesenchymal: Differential expression analysis between each cell type against all cell types in Mesenchymal population (IPF samples only) and between each diagnosis against control for individual cell types in Mesenchymal population. **KRT5/KRT17_DEG:** Differential expression analysis between KRT5-/KRT17+ cell type against AT1, AT2, Basal, Trans_AT2, Fibroblast, HAS1 High, Myofibroblast and PLIN2+ (Figure S3). **IPF_vs_NonIPF**: Differential expression analysis between IPF and non IPF samples in all cell types (with > 50 cells). **Disease_vs_Control**: Differential expression analysis between disease and control for all cell types (with > 50 cells; Figure 1). **Control_only_CT_vs_all**: Differential expression analysis between each cell type against all cell types in control samples only. **IPF_only**: Differential expression analysis between HAS1 and KRT5-/KRT17+ cell type vs all mesenchymal cells, fibroblast, mesothelial cells, myofibroblast and PLIN2+ for only IPF samples. **CT_vs_all_per_group**: Differential expression analysis between each cell type against cell type sub groups in control samples only.

Table S7. Differentially expressed genes in KRT5-/KRT17+ cells compared to all epithelial cells, basal cells, AT1 and AT2 cells.

Table S8. Variable genes across KRT5-/KRT17+ pseudotime trajectories.