ONLINE DATA SUPPLEMENT

Single Cell RNA-Sequencing Identifies Diverse Roles of Epithelial Cells in Idiopathic Pulmonary Fibrosis

Yan Xu

Takako Mizuno

Anusha Sridharan

Yina Du

Minzhe Guo

Jie Tang

Kathryn A. Wikenheiser-Brokamp

Anne-Karina T. Perl

Vincent Funari

Jason J. Gokey

Barry R. Stripp

Jeffrey A. Whitsett

SUPPLEMENTAL LEGENDS AND FIGURES

Figure S1: Histologic features of usual interstitial pneumonia in lung biopsies from IPF patients. Lung tissue were stained with H&E for pathological evaluation: A) Normal lung parenchyma consisting of a bronchiole with surrounding alveoli separated by thin alveolar septae in tissue from a normal donor (CC019-14). B) Patchy interstitial fibrosis adjacent to areas of preserved lung parenchyma with thin alveolar septae was characteristic of IPF (IPF001-14). C) Fibroblastic foci at the interface between fibrotic and preserved lung parenchyma (IPF001-14). D) Honeycomb fibrosis with dense collagen scars, loss of lung architecture and cysts containing intraluminal mucin (IPF 185). E) Normal uniformly thin alveolar septae in donor lung (CC019-14). F) Alveolar septae expanded by fibrosis and chronic inflammatory infiltrates in lesions lined by hyperplastic AT2 cells (IPF 180). G) IPF cysts lined by cuboidal and squamous alveolar type epithelium contiguous to mucinous cells showing distinct demarcations between the different cell types (IPF 171). H) Cysts lined by mucinous, flattened alveolar and squamous epithelium (IPF 185). I) Normal bronchiole in biopsy from donor lung (CC019-14). J) Bronchiolar-like epithelial hyperplasia lining cysts in a fibrotic area (IPF 185). K) Cyst in a honeycomb lesion with sharply demarcated ciliated and cuboidal alveolar type epithelial cells (IPF 180). L) Cyst in honeycomb lung with sharply demarcated squamous metaplastic and squamous alveolar-like epithelial cells (IPF 180). Short scale bar=50 μM, Long scale bar=250 μM.

Figure S2. RNAs expression profiles from CD326/HTII-280 FACS sorted IPF and control epithelial cells were analyzed using Ingenuity Pathways Analysis to generate a biological association network. A) Predicted key regulators that were activated in IPF epithelial cells are shown in bold and larger font. Target genes are represented by nodes, and the regulatory relationships between key regulators and their predicted targets are identified through literature

mining using the Ingenuity knowledge base represented by an edge (line). A solid line predicts direct interaction; a dashed line, indirect interaction. Color range indicates the relative induction of genes in IPF condition (the darker the red color, the larger induction of gene expression in IPF vs control). Signaling and transcriptional pathways, including TP53/63, TGFB1, NFKB, CTNNB1, and ETS family members are predicted to serve as important regulatory hubs in the network. B) EZH2 is predicted to be an epigenetic regulator of genes involved in IPF related bio-processes.

Figure S3. Selected cytokines and growth factors induced in IPF Epithelial Cells. A) Heatmap represents predicted induction of genes related to" response to cytokines." B). Heatmap represents the induction of genes responding to growth factors in IPF samples. C). Predicted biological network of cytokines, growth factors and associated bio-processes in IPF. Circled nodes represent growth factors; diamond nodes represent cytokines; pink represents genes induced in IPF; yellow represents bioprocesses activated by growth factors and cytokines. Samples were CD326/HTII-280 FACS sorted epithelial cells from IPF and normal peripheral lung.

Figure S4. Heterogeneity and variation of the distribution of four major epithelial cell types. Pie charts show the % detection of the four major epithelial cell types identified by single cell RNA-Seq of normal and IPF single cells from cells from IPF (n=6) and normal adult (n=3) lungs. Distinct cell types are color-coded as indicated.

Figure S5. Single cell RNA analysis supports the activation of epithelial-mesenchymal-like transition in basal and indeterminate cells. Signaling via *TGFBR1*, *Wnt/β-catenin*, *EGFR*, and *PI3K/AKT*, and transcription factors *ZEB1*, *SNAl2*, *SMAD2/3*, and *ESRP1*, an RNA splicing factor, were selectively induced in C3 basal cells. E-Cadherin (*CDH1*) was suppressed in C3 basal cells. Mesenchymal markers included MMPs, vimentin, N-cadherin (*CHD12*), and fibronectin (*FN1*)

are selectively induced basal (pink) or indeterminate cells (blue), supporting the potential role of basal cells as progenitors of the "indeterminate" cell type. The network analysis was performed using IPA (Ingenuity Pathway Analysis). The regulatory relationships (lines with arrows) connecting gene nodes are identified through literature mining using the Ingenuity knowledge base. A solid line predicts direct interaction; a dashed line, indirect interaction.

Figure S6A. Immunofluorescence confocal microscopy identifies atypical epithelial cell co-expressing E-cadherin and Vimentin in IPF. Vimentin staining, restricted to non-epithelial cells in normal tissues, was detected in many pan-cytokeratin and E-cadherin positive epithelial cells in IPF, indicating a partial epithelial to mesenchymal transition. Images were obtained at 10X magnification [scale bar=200 μM]. Insets in yellow and blue boxes are at 60X magnification.

Figure S6B. Phosphohistone H-3 staining of control and IPF lung tissue. Cell proliferation was estimated by immunofluorescence staining for Phosphohistone H-3 (PHH3). Few cells stained in either IPF or control tissue. Images represent n=3 of control and IPF lung. Shown are examples of images co-stained with ABCA3 (AT2 cell marker, white), p63 (basal cell marker, green), and PHH3 (red). Yellow indicates co-localization. Most PHH3 stained cells in IPF were non-epithelial. Images were obtained at 10X magnification [scale bar=200 μM]. Insets in yellow boxes are at 20X magnification with a Nyquist zoom of 2.59.

Figure S7. Immunofluorescence staining of control and IPF tissue for SP-B (surfactant protein, red), MUC5B (green), and tubulin (TUBA4A). MUC5B and TUBA4A are exclusively expressed in goblet or ciliated cells in conducting airways, respectively. SP-B and MUC5B were co-expressed in IPF epithelial cells. Yellow indicates overlapping expression. . Images were obtained at 10X magnification [scale bar=200 μM]. Insets in yellow boxes are at 40X

magnification. Insets in pink boxes were obtained at 100X magnification with high-resolution confocal imaging by oversampling in x, y, z and deconvolved by Landweber 3D.

Figure S8A-D: Representative gating strategy for isolating EpCAM^{+} epithelial cells from both control and IPF samples for single-cell C1 capture and RNA-sequencing experiments. Lungs cells were stained with 7AAD for viability (not shown); A and B) CD45/31-PE-Cy7 for hematopoetic and endothelial cell exclusion and C and D) CD326-APC for epithelial cells.

Figure S8E and F: Relative abundance of basal versus AT2 cells in human distal lung tissue. Representative FACS analysis of CD271-PE and HTII-280-FITC for basal and AT2 cells, respectively, within CD326+ cells.

Figure S8G: Quantification of relative cell abundance in control and IPF tissue sampled for epithelia cell isolation. Representative cell abundance of CD271+ basal cells compared to HTII-280+ AT2 cells in control (n=7) and IPF (n=7) and determined by FACS. Data are presented as mean ± SD.

Table S1. List of 40 RNAs most increased in IPF compared to control epithelial cells

Gene	p-value	Fold Change (IPF vs. Control)
KRT13	1.76E-03	5.15E+02
PCDH7	3.33E-05	4.48E+02
PXDN	1.48E-03	3.53E+02
IGFBP3	4.79E-03	2.88E+02
KCNN4	1.10E-04	2.82E+02
MMP13	2.14E-05	2.82E+02
ANXA8L2	1.50E-03	2.81E+02
ANXA8L1	1.53E-03	2.64E+02
TP63	3.89E-03	2.50E+02
ANXA8	2.34E-03	2.34E+02
GJB3	8.60E-03	2.10E+02
GPR87	4.91E-03	2.06E+02
PKP1	1.41E-03	1.91E+02
LGR4	2.57E-03	1.86E+02
CXCL6	6.02E-03	1.84E+02
SERPINB5	5.60E-03	1.64E+02
UGT1A6	1.46E-03	1.54E+02
IL36G	5.63E-03	1.33E+02
ALDH1A3	3.24E-03	1.30E+02
KRT6A	1.70E-02	1.28E+02
TIMP3	1.02E-03	1.23E+02
ITGB4	1.34E-03	1.10E+02
CXCR7	1.43E-03	9.97E+01
TIAM1	4.45E-03	9.57E+01
MUC5B	4.50E-02	9.51E+01
CARD11	9.42E-04	9.50E+01
CYP24A1	1.97E-04	8.92E+01
SNAI2	6.27E-03	8.79E+01
IGFBP5	3.35E-03	8.78E+01
FAM83A	3.64E-03	8.69E+01
CX3CL1	2.53E-02	8.49E+01
KRTAP2-3	4.26E-03	8.08E+01
PROM1	7.22E-03	7.96E+01
PLAU	8.17E-04	7.73E+01
BPIFB1	3.37E-02	7.56E+01
DUSP4	5.40E-03	7.43E+01
SRPX2	6.36E-04	7.39E+01
PAX9	7.68E-03	7.27E+01
SULT2B1	1.83E-02	7.27E+01
GALNT14	3.46E-03	7.18E+01

Note: EPCAM (CD326) and HTII-280 sorted epithelial cells from control and IPF donors were isolated from peripheral lung tissue by FACS. RNA-seq analysis identified differentially expressed genes in IPF vs. control samples.

Table S2. List of 40 RNAs most decreased in IPF compared to control epithelial cells

Gene	p-value	Fold Change (IPF vs. Control)
EDNRB	4.39E-02	-3.31E+01
CHI3L2	2.33E-02	-2.44E+01
C8A	1.69E-02	-1.94E+01
NECAB1	6.89E-03	-1.40E+01
RMST	2.65E-03	-1.20E+01
SLC10A4	2.46E-02	-9.84E+00
MUC15	2.93E-02	-9.58E+00
NAP1L2	1.09E-03	-7.07E+00
ACPP	1.32E-02	-7.06E+00
HSPA6	3.55E-02	-6.86E+00
ASRGL1	4.34E-02	-6.74E+00
FAM46C	4.80E-03	-6.60E+00
TNIP3	3.66E-02	-5.87E+00
GUCY1B3	4.51E-02	-5.79E+00
GPR68	2.93E-02	-5.73E+00
PEG3	9.34E-03	-5.48E+00
CHIAP2	4.62E-02	-5.35E+00
DAAM2	3.97E-02	-5.15E+00
DGKG	2.37E-04	-5.13E+00
ACSS3	3.29E-02	-5.06E+00
FTLP3	6.13E-03	-4.94E+00
EDIL3	4.22E-02	-4.93E+00
ARHGAP31	1.44E-02	-4.84E+00
ZNF878	1.85E-03	-4.81E+00
NRGN	4.98E-02	-4.76E+00
TMOD1	3.06E-02	-4.75E+00
CA8	3.68E-03	-4.72E+00
CES1	3.23E-02	-4.69E+00
FZD7	2.29E-03	-4.67E+00
SLC7A8	4.70E-02	-4.67E+00
BEX1	8.44E-03	-4.56E+00
PRKD1	3.86E-02	-4.49E+00
SDC2	1.77E-02	-4.43E+00
GEM	2.09E-02	-4.31E+00
CDKL2	3.28E-02	-4.29E+00
LRRK2	8.97E-03	-4.16E+00
FAM49A	4.33E-02	-4.14E+00
ACADL	2.52E-02	-4.10E+00
PLAG1	8.50E-03	-4.07E+00
FNIP2	1.93E-02	-4.06E+00

Note: EPCAM (CD326) and HTII-280 positive epithelial cells from control and IPF donors were isolated from peripheral lung tissue by FACS. RNA-seq analysis identified differentially expressed genes in IPF vs. control samples.

Table S3. Expression of vimentin in IPF and normal epithelial single cells.

Statistics	Control Cells (n = 215)	Relative Normal IPF Cells (n = 9)	IPF Cells (n = 316)	C1: AT2 (n=217)	C2: Indeterminate (n=91)	C3: Basal (n=131)	C4: Goblet (n=101)
Average Expression	8.55 ± 1.41	24.94 ± 8.07	15.69 ± 2.45	9.49 ± 1.44	19.22 ± 4.59	18.69 ± 4.62	7.56 ± 2.39
T-test p-value		0.02	0.03		0.01	0.02	0.47
Fold change		2.92	1.84		2.02	1.97	-1.26

Note: Average expression of *VIM* in control, IPF cells clustering with C1 and other IPF cells is shown in the left panel, average expression of VIM in four predicted cell types is shown in the right panel. P values were determined by one-tailed Welch's t-test. Vimentin expression was significantly induced in basal and indeterminate cells in IPF, with highest expression in the relatively "normal" IPF AT2-like cells.

Table S4. Antisera and antibodies used for immunofluorescence microscopy.

	Antibody	*Dilution*	*Source*	*Catalog No.*	*Antigen Retrieval*
1	ABCA3	1:100	Seven Hills Bioreagents	GP-984	Citrate
2	ACTA2	1:400?	Sigma		Citrate
3	Acetylated-TUBA4A	1:3000	Sigma	T7451	-
4	CD326 (323/A3)	1:100	Thermo Scientific Pierce	MA5-12436	Do not
5	CDH1 (24E10)	1:200	Cell Signaling	3195	Tris-EDTA
6	HOPX	1:100	Santa Cruz Biotechnology	SC-30216	Citrate
7	HTII-280	1:200	Gift from Dr. Leland Dobbs	--	Citrate
8	KRT14	1:100	Abcam	ab7800	Citrate
9	KRT8	1:100?	DHSB	TROMA-I	Tris-EDTA
10	MUC5B	1:300	Santa Cruz Biotechnology	SC-20119	Citrate
11	P63	1:100	Biocare Medical	CM163A	Citrate
12	pan-Cytokeratin	1:100	Sigma	C2931	Tris-EDTA
13	SMA	1:2000	Sigma	A5228	-
14	Vimentin	1:50	Santa Cruz Biotechnology	sc-7557	Tris-EDTA

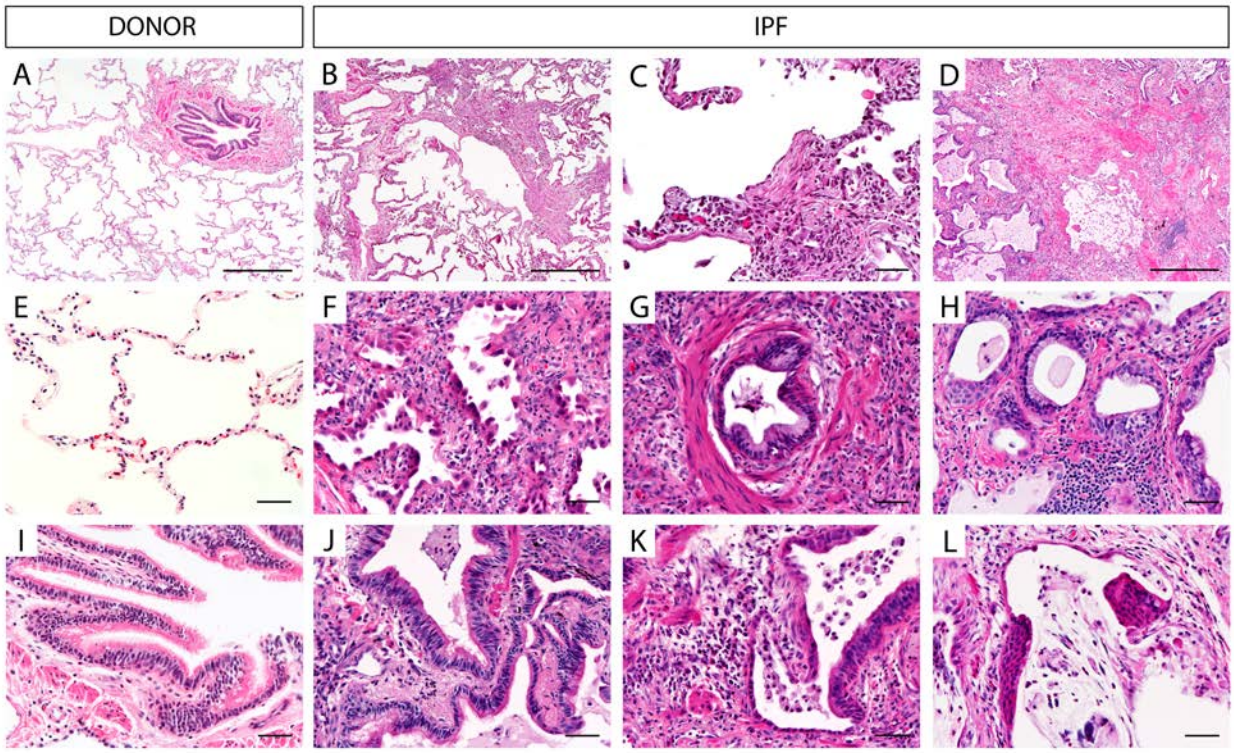

Figure S1

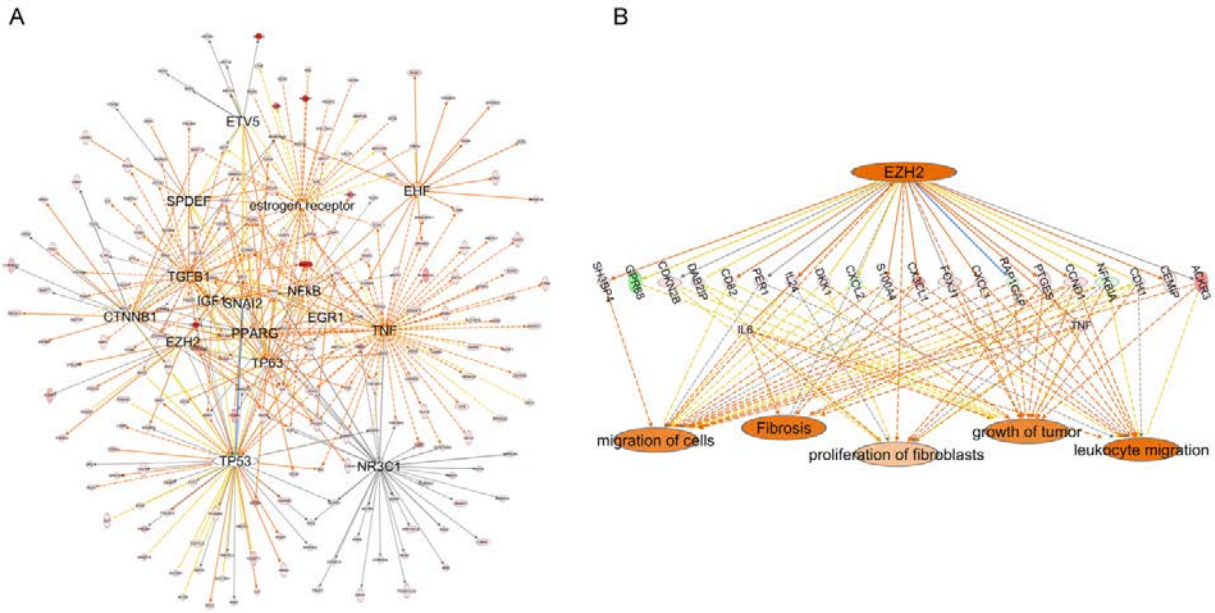
A
ETV5
SPDEF
estrogen receptor
EHF
TGFB1
NFkB
IGF1
SNAI2
CTNNB1
EGR1
TNF
EZH2
PPARG
TP63
TP53
NR3C1
B
EZH2
SH3BP4
GPR68
CDKN2B
DAB2IP
CD82
PER1
IL24
DKK1
CXCL2
S100A4
CXCL11
FOXJ1
CXCL3
RAP1GAP
PTGES
CCND1
NFKBIA
CDK1
CENUP
ADRB3
IL8
TNF
migration of cells
Fibrosis
proliferation of fibroblasts
growth of tumor
leukocyte migration

Figure S2

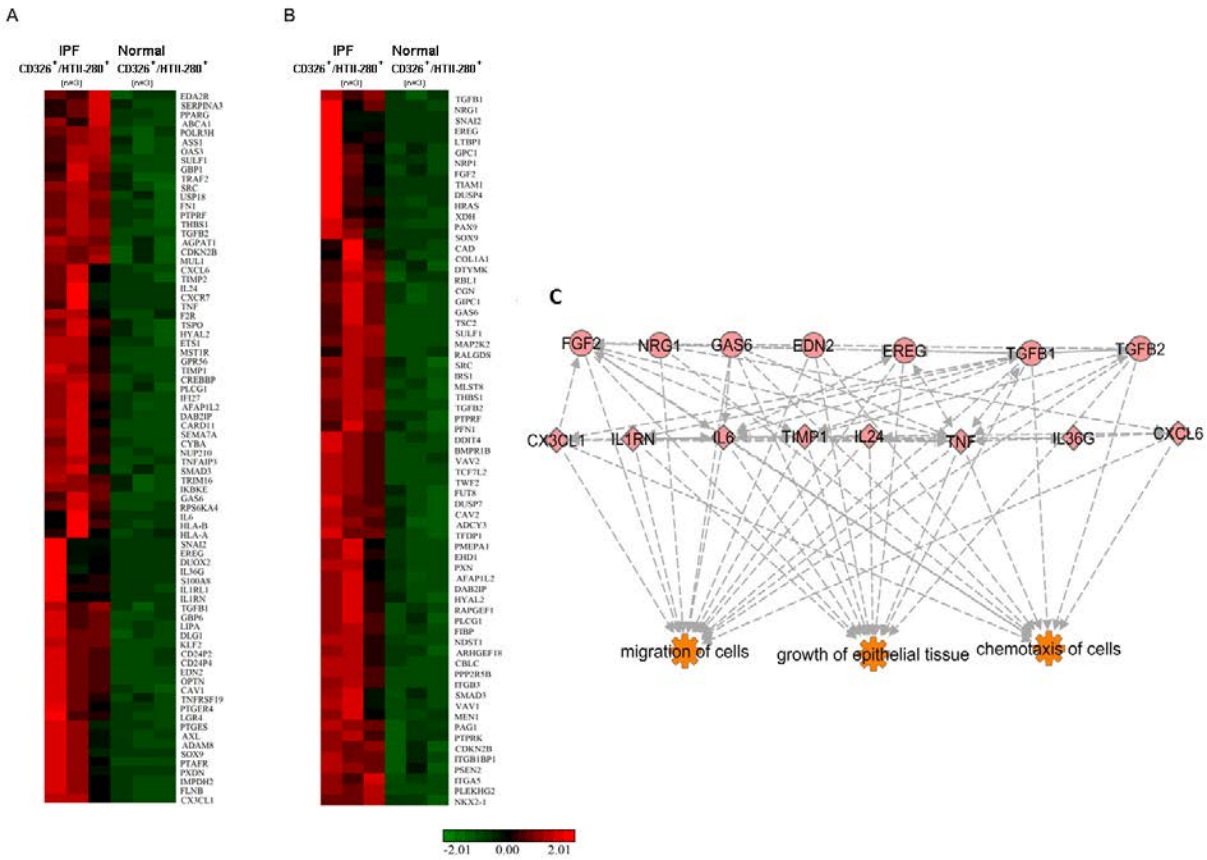
A
IPF
Normal
B
IPF
Normal
C
FGF2
NRG1
GAS6
EDN2
EREG
TGFB1
TGFB2
CX3CL1
IL1RN
IL6
TIMP1
IL24
TNF
IL36G
CXCL6
migration of cells
growth of epithelial tissue
chemotaxis of cells
-2.01
0.00
2.01

Figure S3

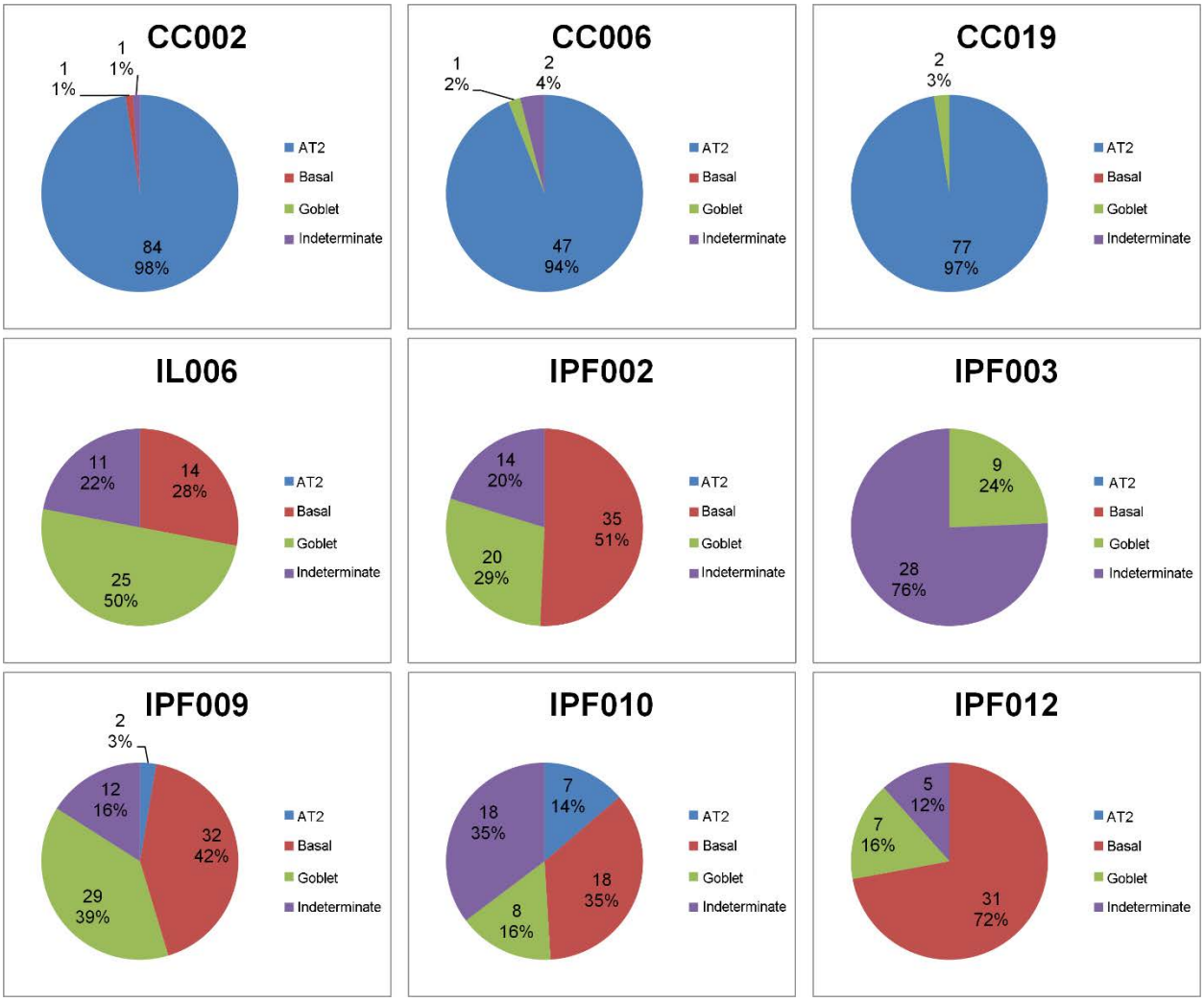
CC002
1
1%
1
1%
84
98%
AT2
Basal
Goblet
Indeterminate
CC006
1
2%
2
4%
47
94%
AT2
Basal
Goblet
Indeterminate
CC019
2
3%
77
97%
AT2
Basal
Goblet
Indeterminate
IL006
11
22%
14
28%
25
50%
AT2
Basal
Goblet
Indeterminate
IPF002
14
20%
35
51%
20
29%
AT2
Basal
Goblet
Indeterminate
IPF003
9
24%
28
76%
AT2
Basal
Goblet
Indeterminate
IPF009
2
3%
12
16%
32
42%
29
39%
AT2
Basal
Goblet
Indeterminate
IPF010
7
14%
18
35%
18
35%
8
16%
AT2
Basal
Goblet
Indeterminate
IPF012
5
12%
7
16%
31
72%
AT2
Basal
Goblet
Indeterminate

Figure S4

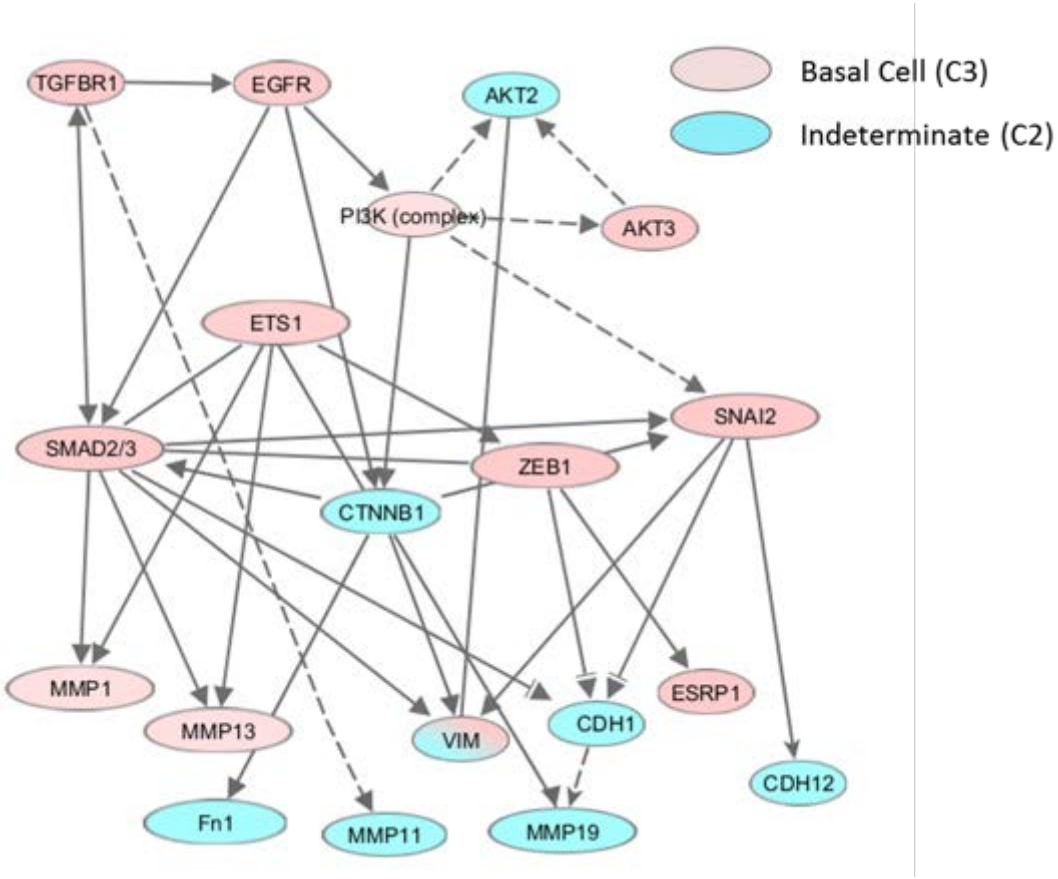
Basal Cell (C3)
Indeterminate (C2)
TGFBR1
EGFR
AKT2
PI3K (complex)
AKT3
ETS1
SNAI2
SMAD2/3
ZEB1
CTNNB1
MMP1
MMP13
VIM
CDH1
ESRP1
CDH12
Fn1
MMP11
MMP19

Figure S5

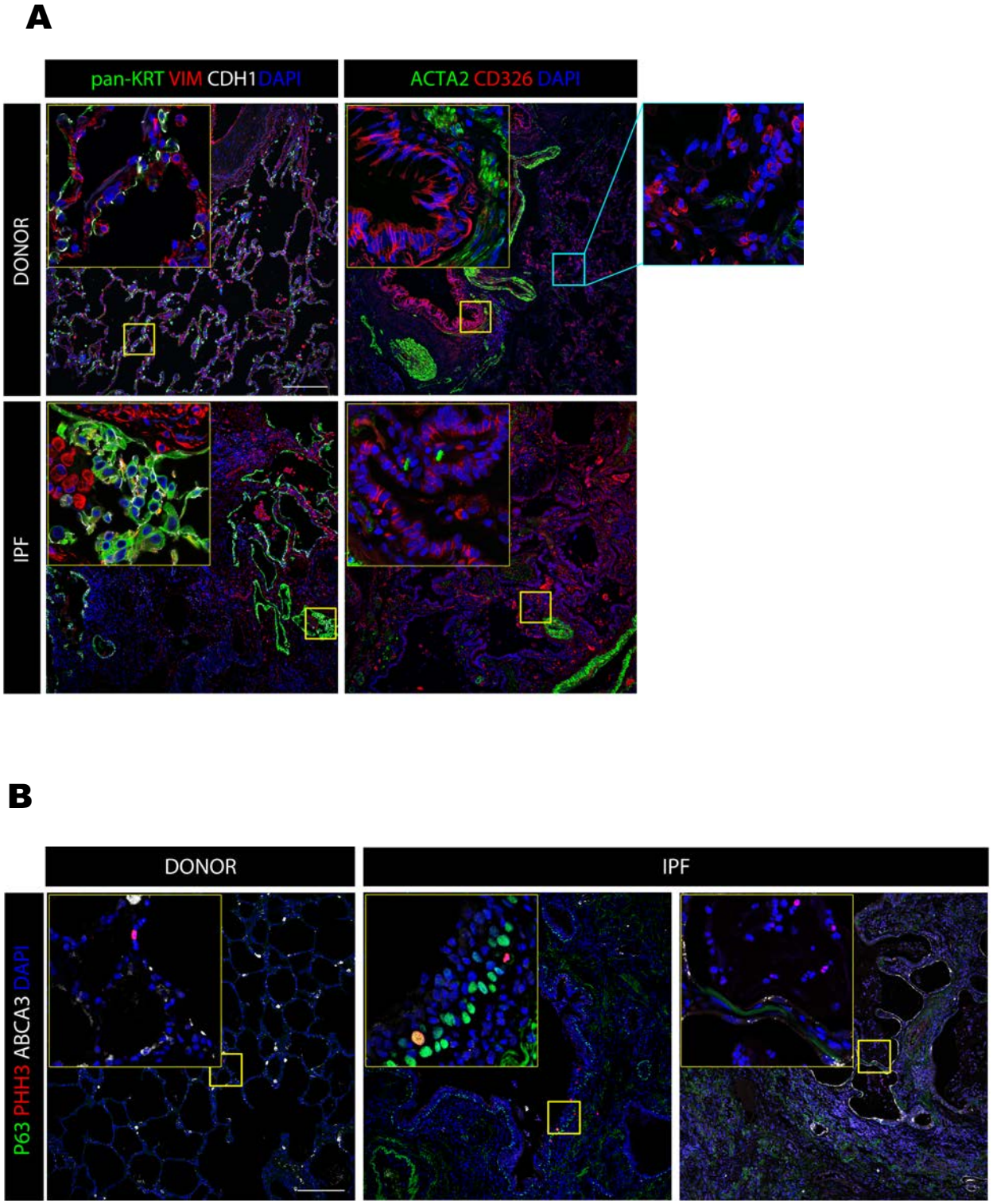
A
pan-KRT VIM CDH1 DAPI
ACTA2 CD326 DAPI
DONOR
IPF
B
DONOR
IPF
P63 PHH3 ABCA3 DAPI

Figure S6

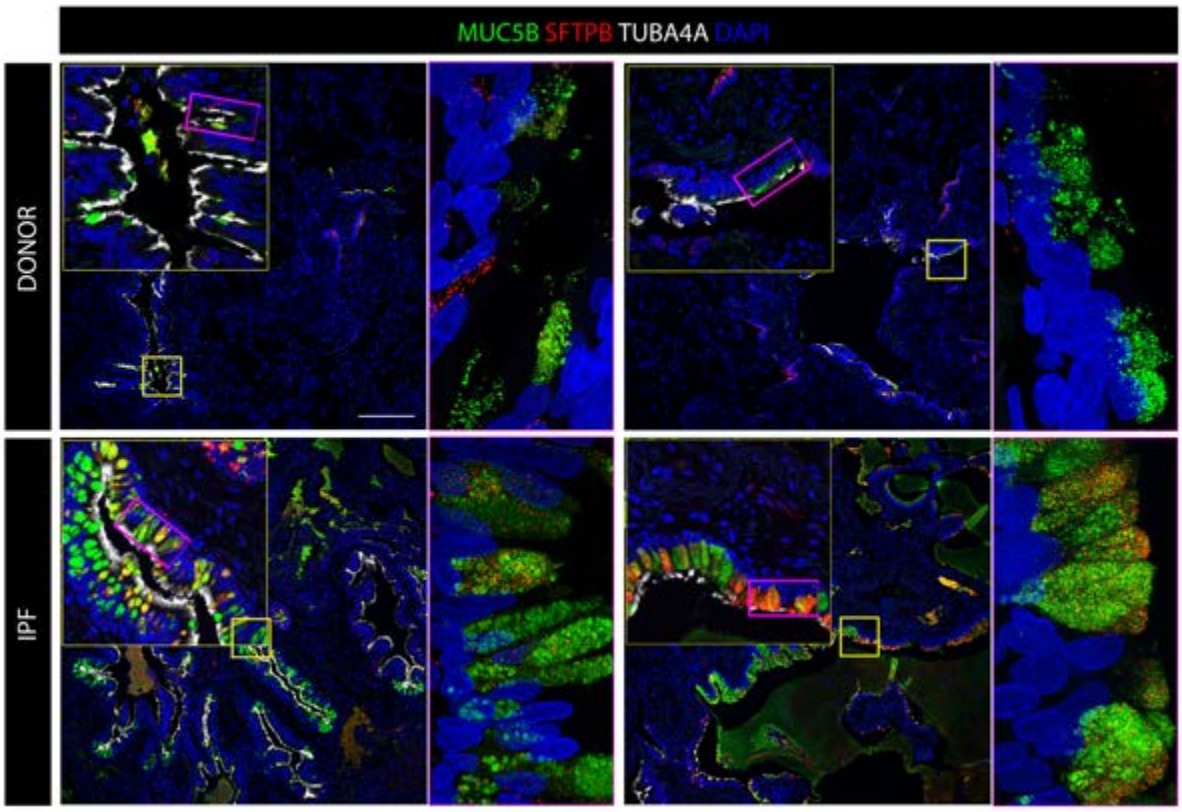

Figure S7

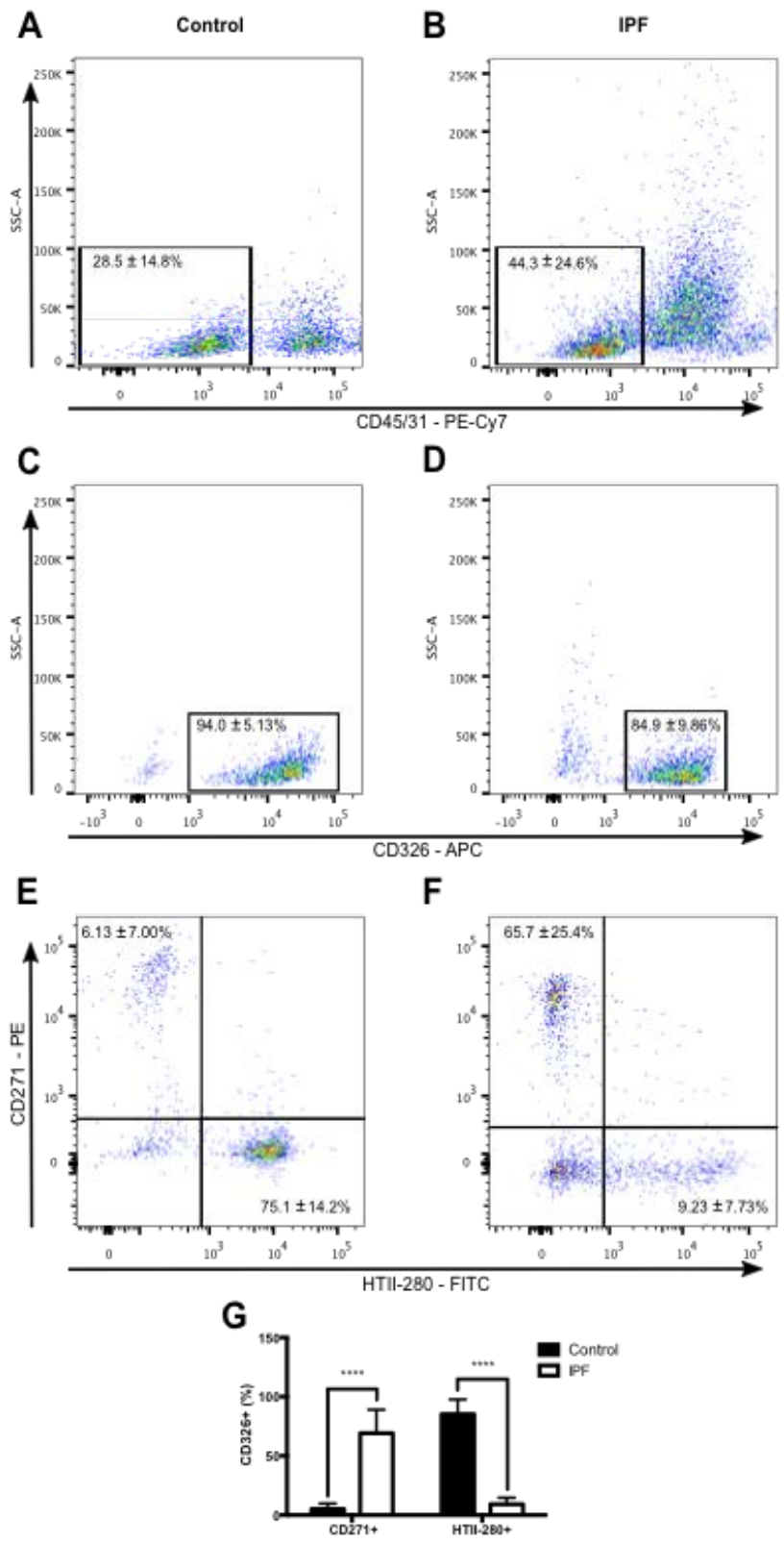
A
Control
B
IPF
SSC-A
28.5 ± 14.8%
44.3 ± 24.6%
CD45/31 - PE-Cy7
C
D
94.0 ± 5.13%
84.9 ± 9.86%
CD326 - APC
E
F
6.13 ± 7.00%
75.1 ± 14.2%
65.7 ± 25.4%
9.23 ± 7.73%
CD271 - PE
HTII-280 - FITC
G
CD326+ (%)
Control
IPF
****
CD271+
HTII-280+

Figure S8