Single cell reconstruction of human basal cell diversity in normal and IPF lung.

Gianni Carraro[1, *], Apoorva Mulay[1, *], Changfu Yao[1, *], Takako Mizuno[1], Bindu Konda[1], Martin Petrov[1], Daniel Lafkas[2], Joe R. Arron[2], Cory M. Hogaboam[1], Peter Chen[1], Dianhua Jiang[1], Paul W. Noble[1], Scott H. Randell[3], Jonathan L. McQualter[1,4], Barry R. Stripp[1, #].

[1]Lung and Regenerative Medicine Institutes, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA; [2]Genentech Inc, San Francisco, CA, USA; [3]Department of Cell Biology and Physiology, University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599; [4]School of Health and Biomedical Sciences, RMIT University, Melbourne, VIC 3001, Australia.

[*]Contributed equally to this manuscript.

[#]Corresponding author: Strippb@cshs.org.

Author contributions

G.C. and B.R.S. designed the project and wrote the manuscript. G.C. performed scRNAseq data analysis. A.M. designed and performed experiments for Notch signaling. T.M. designed and performed organoid experiments and population RNAseq. M.P. and B.K. performed tissue processing and immunostainings. C.Y. performed single-cell RNAseq experiments and popRNAseq data processing, M.P. performed single-cell RNAseq experiments. J.L.M. performed high throuput FACS screening experiments and contributed to the project design. D.L., J.R.A. and S.R. contributed to data interpretation.

C.M.H., P.C., D.J., and P.W.N. contributed with comments on the manuscript. All authors read and reviewed the manuscript.

Abstract

Rationale. Declining lung function in patients with interstitial lung disease is accompanied by epithelial remodeling and progressive scarring of the gas-exchange region. There is a need to better understand the contribution of basal cell hyperplasia and associated mucosecretory dysfunction to the development of idiopathic pulmonary fibrosis (IPF).

Objectives. We sought to decipher the transcriptome of freshly isolated epithelial cells from normal and IPF lung to discern disease-dependent changes within basal stem cells.

Methods. Single cell RNA sequencing was used to map epithelial cell types of the normal and IPF human airway. Organoid and ALI cultures were used to investigate functional properties of basal cell subtypes.

Measurements and Main Results. We found that basal cells included multipotent and secretory primed subsets in control adult lung tissue. Secretory primed basal cells include an overlapping molecular signature with basal cells obtained from distal lung tissue of IPF lungs. We confirmed that NOTCH2 maintains undifferentiated basal cells and restrict basal-to-ciliated differentiation, and present evidence that NOTCH3 functions to restrain secretory differentiation.

Conclusions. Basal cells are dynamically regulated in disease and are specifically biased towards expansion of the secretory primed basal cell subset in idiopathic pulmonary fibrosis. Modulation of basal cell plasticity may represent a relevant target for therapeutic intervention in IPF.

Introduction

Human airways are lined by a mucociliary epithelium that functions in the clearance of inhaled pathogens, particulates and toxic gases. Major portions of this epithelium are pseudostratified, with basal cells located adjacent to the basement membrane and subtending specialized luminal cell types including ciliated cells and various secretory cell types. Both basal and secretory cell types have proliferative capacity, but only basal cells have the capacity for long-term self renewal that is commonly associated with tissue stem cells(1, 2). Pathological remodeling of the airway epithelium is common in chronic lung disease(3, 4). Disease-related changes include hyperplasia of basal cells and altered differentiation of secretory cells leading to overproduction of mucus(5, 6).

Idiopathic pulmonary fibrosis (IPF) is a progressive lung disease with mean survival after initial diagnosis of approximately three years(7). Declining lung function results from scarring and remodeling of the interstitium surrounding the alveoli. Other pathological sequelae of IPF include the appearance of honeycomb cysts at sites of extensive parenchymal remodeling; structures that are composed of basal cells and mucin-producing airway secretory cells (5). These structural defects are considered to be secondary to epithelial progenitor cell dysfunction and defective epithelial-mesenchymal signaling (8, 9); furthermore detection of basal cells within bronchoalveolar lavage fluid of patients with IPF is associated with poor prognosis (10).

Herein, we comprehensively assess the transcriptome of single cells from normal human lung and lung tissue of patients undergoing transplantation for end-stage IPF. We found

71 that basal cells include multiple molecularly distinct states in the normal lung, including
72 multipotent and secretory cell primed, and present evidence suggesting that basal cell
73 subsets form a hierarchy that is regulated by Notch signaling.
74
75

76 **Methods**

77 **Study population**

78 Explant tissue was obtained from patients undergoing transplantation for end-stage IPF
79 at either Duke University or Cedars-Sinai Medical Center, in compliance with consent
80 procedures accepted by the Internal Review Board of Cedars-Sinai Medical Center.
81 Human lung specimens obtained through the International Institute for the Advancement
82 of Medicine (IIAM) were obtained in compliance with consent procedures developed by
83 IIAM and approved by the Cedars-Sinai Medical Center IRB.

84 **Data availability**

85 All transcriptome data were deposited in GEO: GSE143706 and GSE143705. A GitHub
86 repository with a markdown to show our QC and data analysis, is available at
87 https://github.com/gc-github-bio/Normal-and-IPF-lung.

88 **Histology**

89 Lung tissue was formalin-fixed, paraffin-embedded and sectioned at 4 μm. Cultures were
90 fixed with 4% paraformaldehyde and washed with Phosphate-Buffered Saline (PBS) prior
91 to whole-mount staining. Paraffin sections were deparaffinized in xylene and rehydrated

92 through a gradient of ethanol. Prior to antibody staining, antigen retrieval was performed
93 in 10 mM citrate buffer (pH 6.0) using a 2100-Retreiver (Aptum Biologics). Slides were
94 washed with Tris-Buffered Saline (TBS) and blocked for 1 hour at room temperature using
95 blocking buffer containing 3% BSA, 0.1% Triton X-100 and 10% serum (from secondary
96 antibody species) in TBS. Primary antibodies MUC5B (HPA008246), MUC5AC (45M1,
97 Thermofisher), FOXJ1 (14-9965-82, Biolegend), KRT5 (905904, Biolegend), KRT8
98 (TROMA-1, Hybrodoma Bank, Iowa), were diluted in 3% BSA, 0.1% Triton X-100 and
99 1% serum in TBS, applied to tissue sections and incubated overnight at 4ºC. Nuclei were
100 counterstained with 2-(4-Amidinophenyl)-1*H*-indole-6-carboxamidine (DAPI) and
101 appropriate secondary antibodies conjugated to fluorochromes (Thermo Fisher Scientific)
102 were applied for 1 hour at room temperature. Sections were mounted using Fluoromount
103 G mounting medium (Thermo Fisher Scientific). Stained samples were imaged on a Zeiss
104 LSM 780 Confocal Microscope and image processed using Zen Blue software (Zeiss).

105 **Cell isolation**

106 Human IPF explants were processed to obtain epithelial cells from fibrotic regions. Donor
107 explants were processed to obtain airway epithelial cells. The tissue was processed as
108 described previously(1) with the following modifications. Proximal tissue was
109 enzymatically digested with Liberase followed by gentle scraping of epithelial cells off the
110 basement membrane. IPF tissue was finely minced and washed in Ham's F12 (Corning)
111 at 4°C for 5 minutes with rocking, followed by centrifugation for 5 minutes at 600g and
112 4°C. The minced cleaned tissue was then incubated in DMEM/F12 (Thermo Fisher
113 Scientific) containing 1X Liberase (Sigma-Aldrich), incubated at 37°C with rocking for 45
114 minutes. Dissociated single-cell preparations were enriched for epithelial cells and

115 depleted of erythrocytes, leukocytes, and endothelial cells using antibodies against the
116 following molecules: EPCAM (CO17-1A, 369820), CD235a (HI264, 349106), CD45
117 (2D1,368522), and CD31 (WM59,303124) (Biolegend). For the basal cells enrichment
118 experiment and subpopulations isolation, CD271 (ME20.4, 345106) and CD66 (ASL-32,
119 342306) (Biolegend) were also included. Labeled cells were washed in HBSS with 2%
120 FBS, resuspended and placed on ice for fluorescence-activated cell sorting (FACS) using
121 a BD Influx cell sorter (Becton Dickinson). Viability was determined by staining cell
122 preparations with either 7AAD (Biolegend), Propidium Iodide (Biolegend) or DAPI
123 (ThermoFisher Scientific), 15 minutes prior to cell sorting.

124 **In vitro cultures**

125 Two thousand FACS enriched cells were resuspended with a 1:1 (v/v) ratio of Matrigel
126 (Corning) and culture medium and 100uL seeded onto the apical surface of a 12 mm,
127 hydrophilic PTFE 0.4 μm pore-size cell culture insert (EMD Millipore) placed inside a 24-
128 well flat-bottom plate. After polymerization of the Matrigel, 500μL of basal medium
129 (PneumaCult™-Ex Medium, STEMCELL Technologies) was added to the well, outside
130 the cell culture insert. Cultures maintained at 37°C in a humidified incubator (5% CO_2)
131 and continued to be maintained in basal medium with media changes every other day.
132 To evaluate differentiation potential and self-renewal capacity of $CD66^+$, $CD66^-$ basal
133 cells, cultures were grown for up to fourteen days, at which time they were either fixed for
134 histological analysis or organoids harvested and dissociated to yield single cell
135 suspensions for serial passage. Colony forming efficiency was determined by counting
136 the number of colonies with a diameter of ≥ 50μm in each culture and shown as a
137 percentage of input epithelial cells from images taken with the EVOS® XL Core

138 transmitted-light inverted imaging system (ThermoFisher Scientific). FIJI was used to
139 quantify colony forming efficiency from three independent experiments. Statistical
140 analysis was performed using GraphPad Prism version 7.0a. For testing significance of
141 the results, one-way ANOVA test was used. For serial passaging experiments, whole
142 cultures dissociated by Liberase (Sigma-Aldrich) at 0.25 Wünsch U/mL in HBSS warmed
143 to 37°C freed organoids from the Matrigel. After 15 minutes incubation at 37°C, cells were
144 placed on a Thermomixer (Eppendorf) and agitated at 1000 rpm at 37°C for 20 minutes.
145 Cells pelleted by centrifugation at 400g for 5 minutes at 4°C, followed by another
146 dissociation step using 2mL of TrypLE™ Express Enzyme (Gibco) to generate single-cell
147 suspensions. Once single-cell suspensions were attained and visually confirmed under
148 light microscopy, cells were washed in HBSS-2%FBS to stop the reaction. Live cell counts
149 were performed on resuspended cells in 0.5-1mL HBSS-2%FBS using a hemocytometer
150 and Trypan Blue (Sigma-Aldrich). For air liquid interface (ALI) culture experiments,
151 isolated $CD66^+$ and $CD66^-$ basal cells were expanded and seeded onto the apical surface
152 of a 12 mm, hydrophilic PTFE, 0.4 μm cell culture inserts (EMD Millipore) coated with
153 matrigel and placed inside a 24-well flat-bottom plate. Cells were grown with 500μL of
154 basal or ALI medium (STEMCELL Technologies) according to manufacture instructions.
155 Cultures were maintained at 37°C in a humidified incubator (5% CO_2). For studies of
156 NOTCH signaling, the γ-secretase inhibitor (2*S*)-*N*-[(3,5-Difluorophenyl)acetyl]-L-alanyl-
157 2-phenyl]glycine1,1-dimethylethyl ester (DAPT) (Tocris) was added at a concentration of
158 5 μM. Blocking antibodies for NOTCH1 (NR1), NOTCH2 (NR2), or NOTCH3 (NR3), were
159 added individually at concentrations of 30 μg/ml starting at the beginning of the culture,

160 and media was changed every three days. Notch blocking antibodies were provided by
161 Genentech.

162 **High-throughput FACS analysis**

163 Samples of bronchi, bronchiolar and distal lung tissue were pooled, dissociated into a
164 single cell suspension and used for cell screen analysis. Basal cell subset stained with
165 APC anti-human CD271, AF488 anti-human CD326, Brilliant Violet 421 anti-human CD45
166 and CD31 combined with a panel of cell-surface proteins targeting 332 PE-conjugated
167 monoclonal antibodies (Biolegend, LEGENDScreen Human PE Kit, 700007) and 10
168 isotype controls (Biolegend) identified novel candidate markers. Cell preparation for
169 screening analysis were followed per manufacturer's protocol (Biolegend). In brief, cells
170 filtered through a 40µm cell strainer were resuspended at a density of between 1.5 - 2 x
171 10^6 cells/mL in 30mLs of HBSS-2%FBS. Prior to sample staining, cells pre-incubated with
172 human Fc-receptor blocking solution (BioLegend) to reduce nonspecific staining carried
173 out at a concentration of (0.01%) for 15 minutes on ice. After lyophilized antibodies on
174 plates reconstituted, plates ready for sample staining had 75µL/well of cells aliquoted and
175 incubated in the dark for 30 minutes on ice. Cells were washed twice and fixed for 10
176 minutes at room temperature prior to flow cytometer analysis. Analyses were performed
177 on labeled cells using an LSRFortessa cell analyzer (Becton Dickinson, BD) and collected
178 data collected were interrogated using FlowJo software.

179 **Population RNAseq**

180 Basal cell $CD66^+$ and $CD66^-$ populations were processed immediately or after in vitro
181 culturing for ten days. RNA was extracted from the sorted cells using the RNeasy Micro
182 Kit (Quiagen). Total RNA was quantified and sample contamination by proteins or
183 carryover reagents were assessed using both a NanoDrop and a Qubit (Thermo

184 Scientific). Samples were then qualified using the Fragment Analyzer (Advanced
185 Analytical Technologies) to check the integrity of total RNA by measuring the ratio
186 between the 18S and 28S ribosomal peaks. SMART-Seq V4 Ultra Low RNA Input Kit for
187 Sequencing (Takara Bio USA, Inc., Mountain View, CA) was used for reverse
188 transcription and generation of double stranded cDNA for subsequent library preparation
189 using the Nextera XT Library Preparation kit (Illumina, San Diego, CA). Reaction Buffer
190 (1X) was added to lysed cells and the RNA was used directly for oligo(dT)-primed reverse
191 transcription, followed by cDNA amplification and cleanup. Quantitation of cDNA was
192 performed using Qubit (Thermo Fisher Scientific). cDNA normalized to 80 pg/µl was
193 fragmented and sequencing primers added simultaneously. A limiting-cycle PCR added
194 Index 1 (i7) adapters, Index 2 (i5) adapters, and sequences required for cluster formation
195 on the sequencing flow cell. Indexed libraries were pooled and cleaned up, and the
196 pooled library size was verified on the 2100 Bioanalyzer (Agilent Technologies) and
197 quantified via Qubit. Libraries were sequenced on a NextSeq 500 (Illumina) using with a
198 1x75 bp read length and coverage of over 25M reads per sample. Population RNAseq
199 reads were processed using packages inside Galaxy (https://usegalaxy.org/).
200 Specifically, alignment to human genome hg38 was performed using STAR 2.6.0b-1 with
201 application of the default setting in Galaxy. Raw counts were measured using
202 FeatureCounts 1.6.0.3 with application of the default setting in Galaxy. The annotation
203 GTF file of USCS hg38 was downloaded from iGenomes
204 (https://support.illumina.com/sequencing/sequencing_software/igenome.html).
205 Differential gene expression was determined using DESeq2(4) (version 2.11.40.2) using
206 default parameters in Galaxy. The Top DEG for freshly isolated and cultured datasets,
207 along with raw and normalized counts, are shown in Table 6.

208 Heatmaps in Fig. 3C-D were produced selecting the top 20 differentially expressed genes
209 in each group on the base of their log2 fold change among genes with adjusted p-values
210 < 0.05. Heatmaps in Fig. S4A-B were produced as follows: for popRNAseq DEG, only
211 genes with adjusted p-value < 0.05 were used for analysis. A DEG of scRNAseq data of
212 MPB and SPB cells was produced (Table 6) and only genes with adjusted p-value < 0.05
213 were used in the analysis. scRNAseq DEG were used to define PopRNAseq DEG that
214 were used in the analysis. For freshly isolated cells the top 20 DEG's for CD66+ and
215 CD66- where used for creation of the heatmap. For popRNAseq of cultured basal cells,
216 only 13 genes passed the filtering and were used for creation of the heatmap.
217
218 **Single cell RNAseq**
219 Single cells were captured using a 10X Chromium device (10X Genomics) and libraries
220 were prepared according to the Single Cell 3' v2 or v3 Reagent Kits User Guide (10X
221 Genomics, Table 9 specifies the kit version used for each dataset). Totalseq-A human
222 hashing antibodies (Biolegend) were used to label cells isolated from sample cc05;
223 trachea or extralobar bronchi used TotalSeq-A0253 (394605), proximal intra-lobar
224 bronchi used TotalSeq-A0254 (394607). Hashing was performed for a different purpose
225 and was not considered for data processing in this manuscript. Cellular suspensions were
226 loaded on a Chromium Controller instrument (10X Genomics) to generate single-cell Gel
227 Bead-In-EMulsions (GEMs). Reverse transcription (RT) was performed in a Veriti 96-well
228 thermal cycler (ThermoFisher). After RT, GEMs were harvested, and the cDNA
229 underwent size selection with SPRIselect Reagent Kit (Beckman Coulter). Indexed
230 sequencing libraries were constructed using the Chromium Single-Cell 3' Library Kit (10X
231 Genomics) for enzymatic fragmentation, end-repair, A-tailing, adapter ligation, ligation

232 cleanup, sample index PCR, and PCR cleanup. The barcoded sequencing libraries were
233 quantified by quantitative PCR using the KAPA Library Quantification Kit (KAPA
234 Biosystems, Wilmington, MA). Sequencing libraries were loaded on a NovaSeq 6000
235 (Illumina) with a sequencing setting of 26bp and 98bp for v2 and 28bp and 91bp for v3,
236 respectively, to obtain a sequencing depth of at least ~5 x 10^4 reads per cell.
237
238 **Data analysis**
239 Cell Ranger software (10X Genomics) was used for mapping and barcode filtering.
240 Briefly, the raw reads were aligned to the transcriptome using STAR(5), using a hg38
241 transcriptome reference from Ensembl 93 annotation. Expression counts for each gene
242 in all samples were collapsed and normalized to unique molecular identifier (UMI) counts.
243 The result is a large digital expression matrix with cell barcodes as rows and gene
244 identities as columns.
245 Data analysis was mainly performed with Seurat 3.0(6) with some variation that will be
246 described.
247 For all data, quality control and filtering were performed separately for each dataset, to
248 remove cells with low number of expressed genes (threshold n>=200) and elevated
249 expression of apoptotic transcripts (threshold mitochondrial genes < 15%). Only genes
250 detected in at least 3 cells were included. Each dataset was run separately with SoupX
251 analysis package to remove contaminant 'ambient' RNA derived from lysed cells during
252 isolation and capture (Young MD et al., https://doi.org/10.1101/303727). We performed a
253 correction on the basis of genes with a strong bimodal distribution and for which the
254 'ambient' RNA expression was overlapping with a gene signature of a known cell type.
255 The 'adjustCounts' function of SoupX was used to generate corrected count's matrices.

256 Table 8 show the percentage of ambient RNA identified in each dataset. To minimize
257 doublet contamination for each dataset we performed a quantile thresholding of high UMI
258 using a fit model generated using the multiplet's rate to recovered cells proportion, as
259 indicated by 10X Genomics (https://kb.10xgenomics.com/hc/en-
260 us/articles/360001378811-What-is-the-maximum-number-of-cells-that-can-be-profiled-).
261 The raw expression matrix was processed with SCTransform wrapper in Seurat. Each
262 dataset was first processed separately with Principal Component Analysis (PCA),
263 followed by clustering using the first 30 independent components and a resolution of 0.5.
264 Two-dimensional visualization was obtained with Uniform Manifold Approximation and
265 Projection (UMAP)(8). Identified AT2 (SFTPC+), immune (PTPRC+), and endothelial
266 (PECAM1+) contaminating clusters were removed by subsetting the Seurat object, using
267 the 'subset' function, before proceeding to merging of the data and integration. In the
268 merged raw expression matrix, each sample was processed with SCTransform.
269 Mitochondrial and ribosomal mapping percentages were regressed to remove them as
270 source of variation. We obtained log1p logarithmically transformed data for each dataset
271 and scaled data as Pearson residuals. Pearson residual was then used to integrate
272 datasets following Seurat workflow, using the PrepSCTIntegration function. Integrated
273 datasets were used for downstream analysis. The integrated datasets didn't show batch
274 effects from sample of origin allowing datasets clustering on the base of biological
275 significance (Fig. S1F-G). Datasets were processed with PCA using the 5000 most
276 variable genes as input, followed by clustering with Leiden algorithm using the first 30
277 independent components and a resolution of 3 for fine clustering. Two-dimensional
278 visualization was obtained with UMAP. To identify differentially expressed genes between
279 clusters we used Model-based Analysis of Single-cell Transcriptomics (MAST)(9) within

280 Seurat's FindMarkers and FindAllMarkers functions. For this analysis the p-value
281 adjustment is performed using Bonferroni correction based on the total number of genes.
282 Results for major clusters and subclusters are reported as supplementary tables
283 (Supplementary Table 2, 3). To identify major cell types in our normal integrated datasets,
284 we used previously published lung epithelial cell type specific gene lists(10)
285 (Supplementary Table 1) and created a scoring, using a strategy previously
286 described(11). This method calculates the expression of each gene signature list as
287 average, and this is subtracted by the aggregated expression of control feature sets. All
288 analyzed features are binned based on averaged expression, and the control features
289 are randomly selected from each bin. Thirty-seven clusters identified with the Leiden
290 algorithm were assigned to major cell types on the base of rounds of scoring and
291 refinement (FigS2). Each refinement was produced using transcripts differentially
292 expressed within the best identified clusters from the previous scoring. Within each major
293 cell type, Leiden clustering and differential gene expression were used to infer
294 subclustering. Subclustering segregation was visualized by heatmap showing the top
295 differentially expressed genes, and by violin plot, showing a signature score obtained as
296 described above (Fig 1). The gene list used as signature are shown in supplementary
297 tables (Supplementary Table 2, 3). Violin plots show expression distribution and contain
298 a boxplot showing median, interquartile range, and lower and upper adjacent values. To
299 identify CD markers that can discriminate basal cell subtypes, we checked all CD markers
300 (from HGNC, https://www.genenames.org) (Supplementary Table 4) with differential
301 expression in our datasets, and selected markers that were confirmed by FACS
302 screening.

303 Top differentially expressed genes were fed to STRING (https://string-db.org/) to visualize
304 their protein association network and extract their functional enrichments. Gene list of
305 specific terms were used to create a score, as described above, and visualized as UMAP.
306 IPF datasets were first processed as described for normal datasets, from QC analysis to
307 integration. For comparative analysis of control and IPF, a common integration of all
308 datasets was performed following the same integration strategy described for control
309 datasets. The integrated Seurat object containing control and IPF data was used to
310 produce differential gene expression analysis using MAST (Supplementary Table 7). This
311 analysis was used to generate visualizations in Fig.4 B and G.
312 To estimate the presence of cellular transitions between basal and secretory clusters, we
313 used diffusion map with 'Destiny'(12), followed by trajectory analysis with 'Slingshot'(13),
314 following their Bioconductor vignettes. The normalized counts from Seurat were imported
315 into an ExpressionSet for Destiny and into a SingleCellExperiment for Slingshot. The 3D
316 visualization was obtained using the 'rgl' library in R. Notch signaling ligand-receptor
317 interactions were analyzed using iTALK (Wang Y et al., https://doi.org/10.1101/507871).
318 All analyses were performed using R 3.6.

---

319 **Statistical analysis**

320 FIJI was used to quantify colony forming efficiency from three biological samples in at
321 least eight technical replicates. Statistical analysis was performed using GraphPad Prism
322 version 7.0a. Data presented are shown as mean ratio ± SEM. For testing significance of
323 the results, one-way ANOVA test was used. For quantification of immunofluorescence
324 experiments with at least three biological replicates Mann-Whitney test ($P<0.1$) was used.

For analysis of single-cell transcriptomics the p-value adjustment was performed using Bonferroni correction based on the total number of genes. For population RNAseq all samples were collected and stored to allow processing at the same time. For single-cell RNAseq, samples were processed based on sample availability, so balanced batch preparation could not be performed.

References for Methods

1. Xu Y, Mizuno T, Sridharan A, Du Y, Guo M, Tang J, Wikenheiser-Brokamp KA, Perl AT, Funari VA, Gokey JJ, Stripp BR, Whitsett JA. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016; 1: e90558.
2. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; 25: 1105-1111.
3. Langmead B, Wilks C, Antonescu V, Charles R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 2019; 35: 421-432.
4. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014; 15: 550.
5. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; 29: 15-21.
6. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018; 36: 411-420.
7. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 2019; 9: 5233.
8. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol* 2018.
9. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, Linsley PS, Gottardo R. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 2015; 16: 278.
10. Plasschaert LW, Zilionis R, Choo-Wing R, Savova V, Knehr J, Roma G, Klein AM, Jaffe AB. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 2018; 560: 377-381.
11. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, Fisher JM, Rodman C, Mount C, Filbin MG, Neftel C, Desai N, Nyman J, Izar B, Luo CC, Francis JM, Patel AA, Onozato ML, Riggi N, Livak KJ, Gennert D, Satija R, Nahed BV, Curry WT, Martuza RL, Mylvaganam R, Iafrate AJ, Frosch MP, Golub TR, Rivera MN, Getz G, Rozenblatt-Rosen O, Cahill DP, Monje M, Bernstein BE, Louis DN, Regev A, Suva ML. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* 2016; 539: 309-313.

12. Angerer P, Haghverdi L, Buttner M, Theis FJ, Marr C, Buettner F. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* 2016; 32: 1241-1243.
13. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 2018; 19: 477.

Results

Mapping the molecular phenotype of normal human lung airway epithelial cells.

Single-cell RNA-sequencing (scRNA-seq) was performed on fractionated epithelial cells from six donors with no preexisting chronic lung disease and no or mild smoking history (Fig S1A). A single-cell transcriptome of the aggregated datasets containing 9731 cells was generated after initial filtering. We defined 37 clusters (Fig S2A) encompassing 12 cellular phenotypes, including 4 basal, 5 secretory, 2 cilated, and ionocyte cell types, that were visualized by Uniform Manifold Approximation and Projection (UMAP) and by their top differentially expressed genes (Fig1A, B, Table 2). Clusters were first assigned to major cell populations using previously published gene lists(11) (Table 1) to create a signature score(12). Clustering and subclustering were than refined by multiple iterations of scoring, based on differential gene expression (Fig S2C-L). Clusters 18, 28, and 31, initially uncharacterized, were recognized as secretory cell types after the first round of score refinement (Fig S2D), and were than manually annotated as mucous and serous cells on the base of published markers for submucosal gland(13) (Fig 1A, J, K). Using the same approach two separate clusters of ciliated cells were identified. A relatively immature (ciliated2) cluster that contains genes of early ciliogenesysis such as *LRRC6*,

388 *DNAAF1*, and *DNAAF5* (Fig 1A, L, M, FigS1J) and a mature ciliated1 cluster (Fig 1A, L,
389 M). Each cell type classification was represented by single cells from each donor (Fig
390 S1B).
391
392 **Normal human airway basal cells comprise specialized subpopulations.**
393 Basal cells segregated into 4 clusters on the basis of their gene signatures (Fig 1A, B, H,
394 I; Table 3). MPB cluster represents a 'classical' Multipotent Basal (MPB) cell that was
395 defined by a strong WNT signaling signature (Fig 1A-C). PB cluster represents
396 proliferating basal (PB) cells (Fig 1A, B, E). Cells within the SPB cluster express various
397 members of the serpin family (Fig 1H, S1I) that regulate protein folding in the early
398 secretory pathway(14). SPB cells also show enrichment of transcripts involved in O-
399 glycan biosynthesis (Fig 2A), a pathway involved in the initial steps of mucin production
400 by secretory cells (15). The Activated Basal (AB) cluster is characterized by a stress-
401 response signature (Fig1A, B, D) including p38-SAPK MAP kinase cellular stress (Fig
402 S1K). SAPK has been involved in centriolar satellite remodeling and ciliogenesis (16).
403 Furthermore AB cluster is also characterized by activation of the TP53 pathway (Fig1G)
404 and contain a cell-cycle arrest signature (Fig 1F).
405 In addition to scRNAseq we used a high content flow cytometry screening to detect basal
406 cell heterogeneity based on surface marker expression. This discovery screen was
407 performed on pooled samples of epithelial cells isolated from bronchi, bronchiolar and
408 distal lung regions from a single IPF patient. The requirement for very high cell numbers
409 for this type of high throughput screen restricted our capacity to perform this screen on
410 separate spatially localised samples. A total of 332 individual surface markers (plus
411 isotype controls) were screened in combination with a viability die and antibodies against

412 CD31 (endothelial cells), CD45 (hematopoietic cells), EPCAM (epithelial cells) and NGFR
413 (basal cells). Analysis for distribution of cells in the CD31- CD45- EpCAM+ NGFR+ basal
414 cell fraction revealed 30 surface markers that could be used for basal cell subclustering
415 (Fig 2C, S3). Of these, CD66 recognition of CEACAM1,5,6 isoforms was one of the
416 markers that was identified to correlate with selective expression of *CEACAM 5* and *6* by
417 the SPB cluster in the scRNAseq data (Fig 2F) and was thus chosen for further validation
418 in normal (n=3) and IPF (n=3) lungs (Fig 2D, E). This confirmed differential surface
419 reactivity of CD66 among basal cells in the normal proximal lung, and showed a trend for
420 an elevated proportion of CD66+ basal cells in IPF. Immunofluorescent staining of normal
421 lung airways showed that CD66 immunoreactivity was observed at the plasma membrane
422 in a rare population of basal cells in addition to previously described subpopulations of
423 luminal transitional and secretory cells (Fig 2K, L, S4C) (17). Expression of *CEACAM 5*
424 and *6* isoforms that contribute to CD66 immunoreactivity of SPB and secretory cells is
425 also observable at the transcriptional level by UMAP (Fig 2G-J). Interestingly, dimensional
426 reduction by diffusion map with 'Destiny' and lineage visualization with 'Slingshot' show
427 that SPB cells have a molecular phenotype that is transitional between basal and
428 secretory cells (Fig 2B). However, our analysis does not allow determination of whether
429 this population represents a transitory or phenotypically stable subset of basal cells.
430
431 **SPB cells have limited capacity for self-renewal and are primed to assume**
432 **secretory cell fates.**
433
434 We next sought to test the functional potential of SPB cells and validate computational
435 predictions from scRNAseq data. Total basal cells were sorted for either EPCAM^{+}, NGFR

436 $^{+}$, CD66^{+} (CD66^{+} cells) or EPCAM^{+}, NGFR^{+}, CD66^{-} (referred to as CD66^{-} cells).
437 Mesenchyme-free three-dimensional in vitro cultures were used to assess the capacity of
438 basal cell subsets to form clonal organoids and evaluated differentiation potential. Both
439 CD66^{+} and CD66^{-} populations generated clonal organoids in the absence of
440 mesenchymal support. However, CD66^{+} organoids displayed reduced capacity for self-
441 renewal starting at passage 2, compared to their CD66^{-} counterparts (Fig. 3A). We
442 performed bulk RNA-seq on either freshly isolated or cultured basal cell fractions. Top
443 differentially expressed genes (log2 fold change, adjusted p-value < 0.05) for CD66^{+} and
444 CD66^{-} are shown (Fig. 3C, D, Table 6). Top differentially expressed genes (log2 fold
445 change, adjusted p-value < 0.05) that were also differentially expressed in SPB and MPB
446 cells (scRNAseq data), are shown (Fig. S4A,B, Table 6). PCA analysis showed that
447 CD66^{+} and CD66^{-} cells diverged significantly seven days after organoid culture reflecting
448 their disparate fates (Fig 3B). Immunofluorescence showed marked up-regulation of
449 secretory cell markers MUC5AC and MUC5B (Fig.3E, G, S5). When seeded at high
450 density on Transwell® cultures and grown at air-liquid interface CD66^{+} basal cells also
451 demonstrated a more pronounced propensity for secretory cell differentiation compared
452 to their CD66^{-} counterparts (Fig 3F, H). These data are consistent with the notion that
453 SPB CD66^{+} basal cells are primed to assume secretory cell fates.
454
455
456
457 **IPF distal lung contains an expanded pool of SPB cells.**

458 We previously showed that decreases in the abundance of alveolar type 2 cells (AEC2)
459 within explant tissue of patients undergoing transplant for end-stage IPF was
460 accompanied by a corresponding increase in the abundance of NGFR$^+$ basal cells(18).
461 These findings were of particularl interest in light of GWAS data linking MUC5B promoter
462 polymorphism and susceptibility to IPF, which suggest potential roles for altered airway
463 secretory cells in disease progression(19). Immunofluorescent staining to localize CD66$^+$
464 KRT5$^+$ basal cells in end-stage IPF explant tissue revealed an increase in SPB cells (Fig
465 3I-K). To better define changes to basal cells in the IPF lung we evaluated scRNAseq
466 data from IPF distal lung epithelium. Single cell transcriptomes of 7 IPF samples (Fig S1A,
467 C, G) were generated and the initial quality control and filtering were performed as for
468 control subjects. IPF datasets were classified by the gene signatures obtained for each
469 of the major epithelial cell types defined in the normal lung (Fig 4A). Shared transcripts
470 between IPF and control datasets were evaluated as integrated data, showing similarity
471 among major cell types (Fig 4B). IPF datasets contained representative cells for AB, MPB,
472 and SPB cells identified in the normal lung. Control and IPF cells showed an incomplete
473 signature overlap. A specific proliferative cluster of IPF basal cells was not identified.
474 Instead, proliferating cells were present in all IPF datasets (Fig S1H). Furthermore, two
475 secretory cell clusters specifically enriched in IPF were named on the basis of their
476 specificity for selected secretoglobins: SCGB3A2+ and SCGB3A2+/SCGB3A1+ (Fig.4A).
477 IPF datasets were enriched in CD66 expression, in SPB and in all the secretory clusters
478 (Fig.4 C-F, Table 7). IPF SPB cells were specifically enriched in *CEACAM6* expression
479 (Fig 4E), and could be distinguisged from normal SPB cell by their gene signature (Fig
480 4G, Table 7). Dimensional reduction by diffusion map with 'Destiny' and lineage

481 visualization with 'Slingshot' show that also in IPF datasets basal and secretory cells form
482 a continuum, with SPB cells being the connection to secretory cells (Fig 4H).
483
484 **Notch signaling regulates the maintenance of SPB cells.**
485 Given the importance of Notch signaling in regulating lung airway maintenance and
486 differentiation, we interrogated Notch pathway gene expression between basal cell
487 subsets of control lung. Analysis of Notch ligand-receptor interactions suggest that both
488 signaling and receiving cells were present within the total basal cell population (Fig 5).
489 Interestingly the SPB cluster posses the largest amount of both autocrine and juxtacrine
490 Notch signaling (Fig 5A). *NOTCH1* was the most broadly involved receptor and was
491 predicted to play a significant role in the regulation of all basal cell subsets with the
492 exception of MPB cells (Fig 5B). In contrast, signaling by *NOTCH3* was the most
493 restricted, showing significant regulation of only the SPB subset (Fig 5B, R-T). The PB
494 cells (proliferating basal cells) showed the least signaling activity and were defined
495 predominantly as Notch receving cells.
496 To test the importance of Notch signaling in SPB cells maintenance we established an
497 ALI culture with a 1:1 proportion of $CD66^+$ (SPB) and $CD66^-$ cells. Treatment with DAPT,
498 a γ-secretase inhibitor that blocks release of the NOTCH intracellular domain, dramatically
499 reduced the abundance of $CD66^+$ SPB cells 48h after treatment, suggesting a functional
500 role for Notch signaling in the maintenance of SPB cells (Fig 5C-E). To specifically
501 interrogate the significance of the three Notch signaling receptors for maintenance and
502 differentiation of basal cells, the fate of basal cells in primary cultures of human tracheal
503 epithelium was evaluated following treatment with specific blocking antibodies for

504 NOTCH1 (NR1), NOTCH2 (NR2), and NOTCH3 (NR3). Cells grown in ALI culture were
505 harvested after 24h, 8 days, and 14 days of treatment. From day 8 to day 14, NR2 treated
506 ALI culture showed a dramatic reduction in KRT5$^+$ cells compared to control (Fig 5F-I).
507 Interestingly, neither NR1 nor NR3 treatment impacted the abundance of KRT5$^+$ cells,
508 suggesting that NOTCH2 is the main ligand involved in basal cell maintenance.
509 Furthermore, NOTCH2 inhibition also promoted expansion of ciliated cell differentiation
510 as shown by an increase in FOXJ1-immunoreactive cells (Fig 5J-M). These data are
511 consistent with the previously reported effects γ-secretase inhibitors, which promote
512 commitment of secretory progenitors to multiciliated cell fates(20) and suggest that
513 NOTCH2 is the relevant receptor that mediates these effects. Even though inhibition of
514 NOTCH3 had no impact on the abundance of KRT5$^+$ basal or FOXJ1$^+$ ciliated cells in
515 culture, NR3 treatment led to an increase in the proportion of mucin secreting cells at 8
516 days (Fig 5N-Q) and 14 days (Fig 5U-W) in culture. These data suggest that NOTCH3
517 may specifically restrains the differentiation of SPB cells, while NOTCH2 plays a more
518 global role in regulating basal cell maintenance and restrains ciliated cell differentiation.
519
520
521
522 **Discussion**
523
524 Airway basal cells function as stem cells that maintain the pseudostratified epithelium of
525 the mouse and presumably also the human lung. Even though basal cell heterogeneity
526 has been implied from studies of the pseudostratified epithelium lining the mouse trachea,

527 little is known of basal cell heterogeneity in the normal human lung and how this changes
528 in the setting of pathological tissue remodeling that accompanies disease. Employing
529 single cell transcriptomics of freshly isolated human airway epithelial cells we were able
530 to discern four broad classes of epithelial cells that included basal, secretory, ciliated, and
531 ionocyte. Even though heterogeneity was observed within each of the abundant epithelial
532 cell types, significant heterogeneity was observed among basal cells. Basal cell subsets
533 included a Wnt signaling multipotent cluster, proliferating basal cells, a cluster that
534 showed evidence of secretory priming, and a cluster with activation of stress response
535 genes. Secretory Primed Basal (SPB) cells may either represent a transitory state of
536 basal cells that are in the process of secretory cell differentiation or a phenotypically stable
537 subset. Diffusion mapping pseudotime analysis of differentiation trajectory show a
538 continuum of differentiation, suggesting a transitional state. However,
539 immunofluorescence analysis reveals rare basal cells that are positive for the SPB marker
540 CD66 in control adult lung tissue and we also found that these cells can be clonally
541 expanded in vitro for multiple passages in the absence of mesenchymal support. Further
542 work is therefore necessary to exaustively determine the features of these cells and their
543 contribution to homeostasis and disease. A recent study performing single-cell
544 transcriptomic analysis of epithelial cells isolated from air-liquid interface cultures of nasal
545 epithelium(21) described a KRT5+ cell type, termed 'suprabasal', that may represent a
546 precursor of secretory cells. Suprabasal cells were characterized by expression of
547 *KRT13*, a transcript observed in SPB cells in our study. Furthermore, suprabasal cells
548 share additional similarity with SPB cells in their expression of *NOTCH3*. However,
549 differences between transcriptomes of suprabasal and SPB cells were observed, such as
550 expression of *KRT4* in suprabasal cells but absence of this cytokeratin in SPB cells.

551 These may represent cell type differences between epithelia of the upper and lower
552 respiratory tracts or represent differences in cell states associated with isolation from
553 fresh tissue versus cultures.
554
555 Studies in mice have established the existence of subsets of quiescent basal cells in the
556 pseudostratified epithelium that lines the mouse trachea(22). Notch2-ICD expression was
557 shown to define a subpopulation of basal cells that preferentially generate luminal
558 secretory cells following injury. In contrast, c-Myb expression defined basal cells that were
559 primed to undergo multiciliogenesis after injury(23). Our analysis of basal cells in the
560 human lung suggests conservation of Notch signaling as a regulator of basal cell fate. In
561 cultures of freshly isolated human basal cells we found that either γ-secretase inhibitors
562 or isoform-specific Notch blocking antibodies biased basal cell differentiation towards
563 secretory or ciliated phenotypes. We specifically found that NOTCH3 restrains secretory
564 primed basal cells from assuming mature secretory cell fates and that is a global regulator
565 of basal cell maintenance. These findings are consistent with studies in mice showing
566 that NOTCH3 signaling regulates the pool of epithelial progenitors that are competent to
567 respond to NOTCH receptors 1 and 2 (24). In basal cells of the human airway NOTCH2
568 is required for their differentiation (25). This is consistent with our observation that
569 blocking NOTCH2 with anti-NOTCH2 antibody blocks secretory and enhances ciliated
570 cell differentiation. Our results with anti-NOTCH3 differ from those by Danahay et al(23)
571 in that we see an increase in secretory cell differentiation, an effect that was not observed
572 in the Danahay study. One explanation for this discrepancy could be that we co-culture
573 CD66+ and CD66- basal cells, hence increasing the proportion of CD66+ basal cells in

574 culture and allowing more precise evaluation of this basal cell subset. JAG blocking
575 antibodies were reported to produce loss of club cells and a gain in ciliated cells(26). This
576 result is consistent with our observed effect of anti-NOTCH2 on ciliated cell expansion.
577 We provide evidence that basal cells in fibrotic explant lung tissue obtained from patients
578 undergoing transplant for end-stage IPF include secretory primed basal (SPB) cells that
579 are found in distal airways and honeycomb regions of IPF lungs that are the site of
580 excessive MUC5B expression(5). Even though the pathophysiological significance of
581 honeycombing and excessive mucus secretion is not fully understood, Seibold et al
582 showed that a polymorphism of the MUC5B gene is associated with increased
583 susceptibility to IPF but surprisingly IPF patients that carry this polymorphism show better
584 outcomes compared to those with the wildtype MUC5B allele(19). We found that IPF SPB
585 cells express genes previously identified as serum biomarkers for higher risk of
586 mortality(27) or proposed to be involved in disease development(28), such as *MUC1*,
587 *MMP7*, and *ICAM1*.
588 This work defines a basal cell hierarchy that is dynamically regulated between health and
589 disease, presenting new therapeutic targets for modulation of normal tissue maintenance
590 and remodeling.
591
592
593 **Acknowledgments**
594 We would like to thank Adrianne Kurkciyan and Stephen Beil for technical assistance, Dr.
595 Edo Israely for assistance with FACS and Cedars-Sinai Medical Center Flow Cytometry,
596 Genomics cores, and Biobank. Dr. Allon M. Klein and Caleb Weinreb for helpful
597 discussions.This work was supported by grants from the California Institute for

598 Regenerative Medicine (LA1-06915), the Cystic Fibrosis Foundation (CARRAR19G0,
599 STRIPP15XX0), Celgene, the NHLBI (R01 HL135163, HL138540, P01 HL108793, T32
600 HL134637 and NIDDK (P30 DK065988).

601

602 **References**

603 1. Wells JM, Watt FM. Diverse mechanisms for endogenous regeneration and repair in
604 mammalian organs. *Nature* 2018; 557: 322-328.

605 2. Teixeira VH, Nadarajan P, Graham TA, Pipinikas CP, Brown JM, Falzon M, Nye E, Poulsom R,
606 Lawrence D, Wright NA, McDonald S, Giangreco A, Simons BD, Janes SM. Stochastic homeostasis
607 in human airway epithelium is achieved by neutral competition of basal cell progenitors. *Elife*
608 2013; 2: e00966.

609 3. Crystal RG, Randell SH, Engelhardt JF, Voynow J, Sunday ME. Airway epithelial cells: current
610 concepts and challenges. *Proc Am Thorac Soc* 2008; 5: 772-777.

611 4. Noble PW, Barkauskas CE, Jiang D. Pulmonary fibrosis: patterns and perpetrators. *J Clin Invest*
612 2012; 122: 2756-2762.

613 5. Seibold MA, Smith RW, Urbanek C, Groshong SD, Cosgrove GP, Brown KK, Schwarz MI,
614 Schwartz DA, Reynolds SD. The idiopathic pulmonary fibrosis honeycomb cyst contains a
615 mucocilary pseudostratified epithelium. *PLoS One* 2013; 8: e58658.

616 6. Rock JR, Randell SH, Hogan BL. Airway basal stem cells: a perspective on their roles in epithelial
617 homeostasis and remodeling. *Dis Model Mech* 2010; 3: 545-556.

618 7. George PM, Patterson CM, Reed AK, Thillai M. Lung transplantation for idiopathic pulmonary
619 fibrosis. *Lancet Respir Med* 2019; 7: 271-282.

620 8. Bagnato G, Harari S. Cellular interactions in the pathogenesis of interstitial lung diseases. *Eur*
621 *Respir Rev* 2015; 24: 102-114.
622 9. Knudsen L, Ruppert C, Ochs M. Tissue remodelling in pulmonary fibrosis. *Cell Tissue Res* 2017;
623 367: 607-626.
624 10. Prasse A, Binder H, Schupp JC, Kayser G, Bargagli E, Jaeger B, Hess M, Rittinghausen S, Vuga
625 L, Lynn H, Violette S, Jung B, Quast K, Vanaudenaerde B, Xu Y, Hohlfeld JM, Krug N, Herazo-Maya
626 JD, Rottoli P, Wuyts WA, Kaminski N. BAL Cell Gene Expression Is Indicative of Outcome and
627 Airway Basal Cell Involvement in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 2019;
628 199: 622-630.
629 11. Plasschaert LW, Zilionis R, Choo-Wing R, Savova V, Knehr J, Roma G, Klein AM, Jaffe AB. A
630 single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* 2018;
631 560: 377-381.
632 12. Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, Fisher JM, Rodman C,
633 Mount C, Filbin MG, Neftel C, Desai N, Nyman J, Izar B, Luo CC, Francis JM, Patel AA, Onozato ML,
634 Riggi N, Livak KJ, Gennert D, Satija R, Nahed BV, Curry WT, Martuza RL, Mylvaganam R, Iafrate AJ,
635 Frosch MP, Golub TR, Rivera MN, Getz G, Rozenblatt-Rosen O, Cahill DP, Monje M, Bernstein BE,
636 Louis DN, Regev A, Suva ML. Single-cell RNA-seq supports a developmental hierarchy in human
637 oligodendroglioma. *Nature* 2016; 539: 309-313.
638 13. Widdicombe JH, Wine JJ. Airway Gland Structure and Function. *Physiol Rev* 2015; 95: 1241-
639 1319.
640 14. Pan S, Iannotti MJ, Sifers RN. Analysis of serpin secretion, misfolding, and surveillance in the
641 endoplasmic reticulum. *Methods Enzymol* 2011; 499: 1-16.

642 15. Strous GJ, Dekker J. Mucin-type glycoproteins. *Crit Rev Biochem Mol Biol* 1992; 27: 57-92.

643 16. Plotnikova OV, Golemis EA, Pugacheva EN. Cell cycle-dependent ciliogenesis and cancer.
644 *Cancer Res* 2008; 68: 2058-2061.

645 17. Rock JR, Gao X, Xue Y, Randell SH, Kong YY, Hogan BL. Notch-dependent differentiation of
646 adult airway basal stem cells. *Cell Stem Cell* 2011; 8: 639-648.

647 18. Xu Y, Mizuno T, Sridharan A, Du Y, Guo M, Tang J, Wikenheiser-Brokamp KA, Perl AT, Funari
648 VA, Gokey JJ, Stripp BR, Whitsett JA. Single-cell RNA sequencing identifies diverse roles of
649 epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016; 1: e90558.

650 19. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, Fingerlin TE, Zhang W,
651 Gudmundsson G, Groshong SD, Evans CM, Garantziotis S, Adler KB, Dickey BF, du Bois RM, Yang
652 IV, Herron A, Kervitsky D, Talbert JL, Markin C, Park J, Crews AL, Slifer SH, Auerbach S, Roy MG,
653 Lin J, Hennessy CE, Schwarz MI, Schwartz DA. A common MUC5B promoter polymorphism and
654 pulmonary fibrosis. *N Engl J Med* 2011; 364: 1503-1512.

655 20. Tsao PN, Vasconcelos M, Izvolsky KI, Qian J, Lu J, Cardoso WV. Notch signaling controls the
656 balance of ciliated and secretory cell fates in developing airways. *Development* 2009; 136: 2297-
657 2307.

658 21. Ruiz Garcia S, Deprez M, Lebrigand K, Cavard A, Paquet A, Arguel MJ, Magnone V, Truchi M,
659 Caballero I, Leroy S, Marquette CH, Marcet B, Barbry P, Zaragosi LE. Novel dynamics of human
660 mucociliary differentiation revealed by single-cell RNA sequencing of nasal epithelial cultures.
661 *Development* 2019; 146.

662 22. Carraro G, Stripp BR. A new notch for lung stem cells. *Cell Stem Cell* 2015; 16: 107-109.

663 23. Pardo-Saganta A, Law BM, Tata PR, Villoria J, Saez B, Mou H, Zhao R, Rajagopal J. Injury
664 induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell
665 subpopulations. *Cell Stem Cell* 2015; 16: 184-197.
666 24. Mori M, Mahoney JE, Stupnikov MR, Paez-Cortez JR, Szymaniak AD, Varelas X, Herrick DB,
667 Schwob J, Zhang H, Cardoso WV. Notch3-Jagged signaling controls the pool of undifferentiated
668 airway progenitors. *Development* 2015; 142: 258-267.
669 25. Danahay H, Pessotti AD, Coote J, Montgomery BE, Xia D, Wilson A, Yang H, Wang Z, Bevan L,
670 Thomas C, Petit S, London A, LeMotte P, Doelemeyer A, Velez-Reyes GL, Bernasconi P, Fryer CJ,
671 Edwards M, Capodieci P, Chen A, Hild M, Jaffe AB. Notch2 is required for inflammatory cytokine-
672 driven goblet cell metaplasia in the lung. *Cell Rep* 2015; 10: 239-252.
673 26. Lafkas D, Shelton A, Chiu C, de Leon Boenig G, Chen Y, Stawicki SS, Siltanen C, Reichelt M,
674 Zhou M, Wu X, Eastham-Anderson J, Moore H, Roose-Girma M, Chinn Y, Hang JQ, Warming S,
675 Egen J, Lee WP, Austin C, Wu Y, Payandeh J, Lowe JB, Siebel CW. Therapeutic antibodies reveal
676 Notch control of transdifferentiation in the adult lung. *Nature* 2015; 528: 127-131.
677 27. Zhang Y, Kaminski N. Biomarkers in idiopathic pulmonary fibrosis. *Curr Opin Pulm Med* 2012;
678 18: 441-446.
679 28. Ballester B, Milara J, Cortijo J. Idiopathic Pulmonary Fibrosis and Lung Cancer: Mechanisms
680 and Molecular Targets. *Int J Mol Sci* 2019; 20.
681

Figure 1

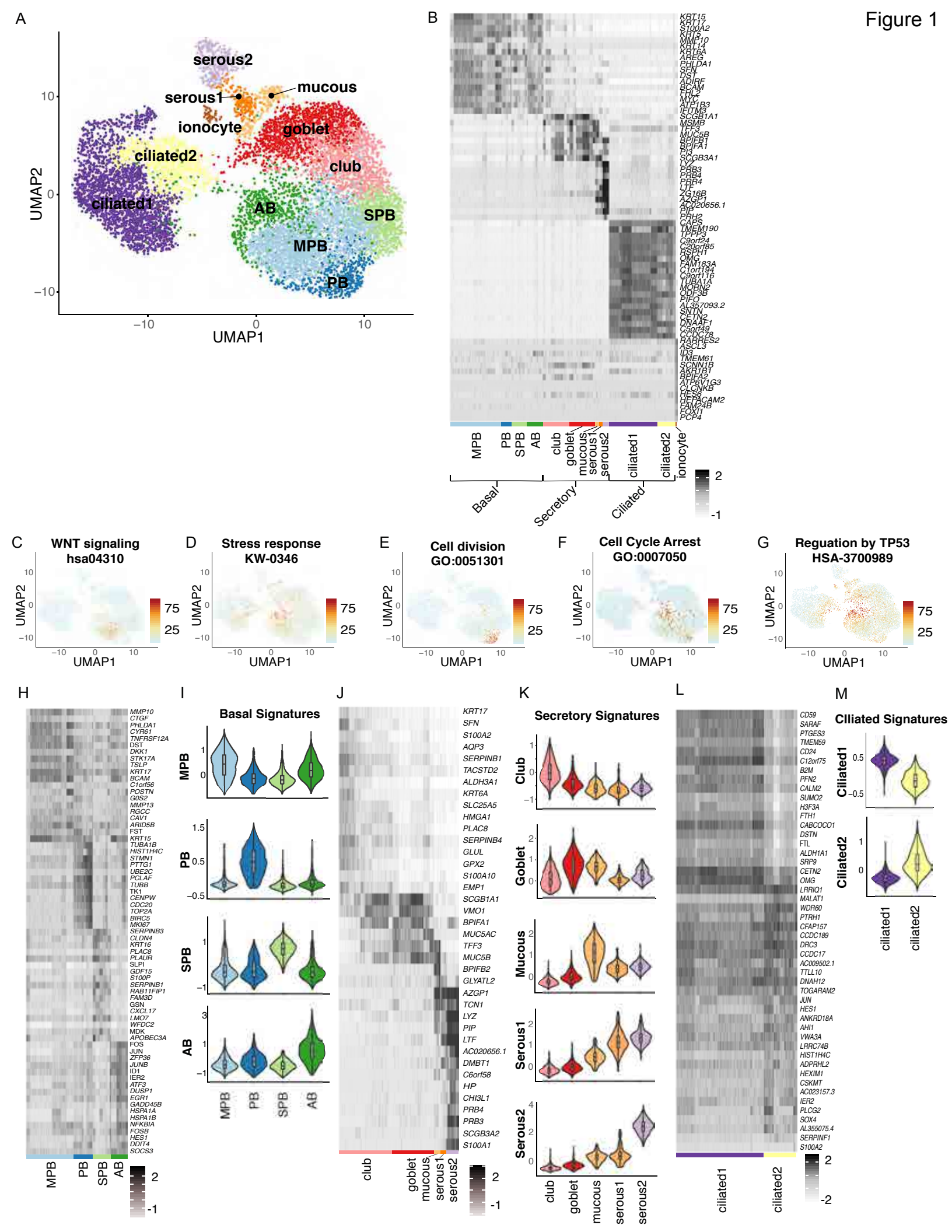

A
serous2
serous1
mucous
ionocyte
goblet
ciliated2
club
ciliated1
AB
SPB
MPB
PB
UMAP2
UMAP1
B
MPB
PB
SPB
AB
club
goblet
mucous
serous1
serous2
ciliated1
ciliated2
ionocyte
Basal
Secretory
Ciliated
C
WNT signaling
hsa04310
D
Stress response
KW-0346
E
Cell division
GO:0051301
F
Cell Cycle Arrest
GO:0007050
G
Reguation by TP53
HSA-3700989
H
I
Basal Signatures
J
K
Secretory Signatures
L
M
CIliated Signatures

Figure 2

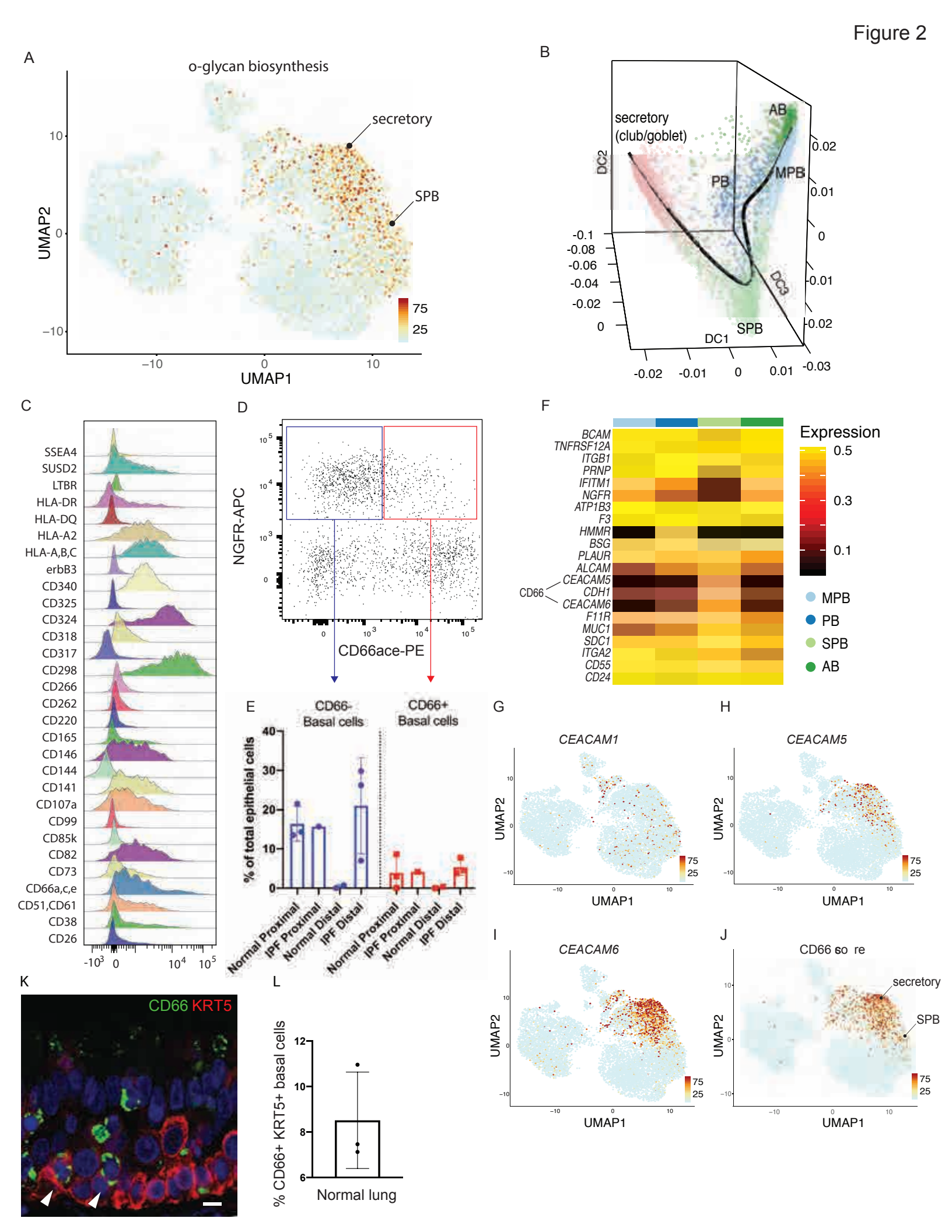
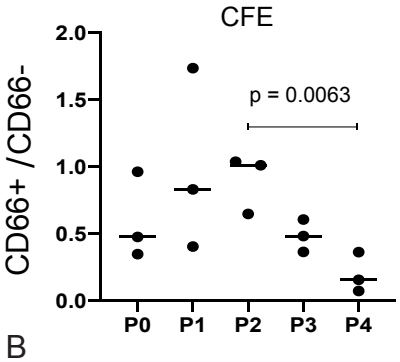
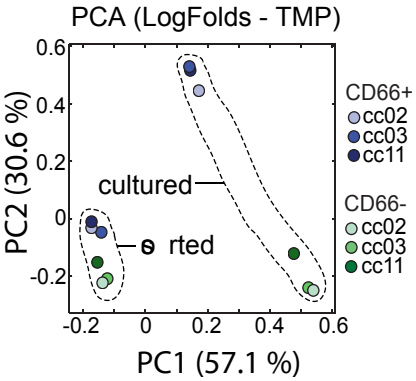

A
o-glycan biosynthesis
secretory
SPB
UMAP2
UMAP1
75
25
B
secretory
(club/goblet)
AB
MPB
PB
SPB
DC1
DC2
DC3
C
SSEA4
SUSD2
LTBR
HLA-DR
HLA-DQ
HLA-A2
HLA-A,B,C
erbB3
CD340
CD325
CD324
CD318
CD317
CD298
CD266
CD262
CD220
CD165
CD146
CD144
CD141
CD107a
CD99
CD85k
CD82
CD73
CD66a,c,e
CD51,CD61
CD38
CD26
D
NGFR-APC
CD66ace-PE
E
CD66-
Basal cells
CD66+
Basal cells
% of total epithelial cells
Normal Proximal
IPF Proximal
Normal Distal
IPF Distal
F
Expression
BCAM
TNFRSF12A
ITGB1
PRNP
IFITM1
NGFR
ATP1B3
F3
HMMR
BSG
PLAUR
ALCAM
CEACAM5
CDH1
CEACAM6
F11R
MUC1
SDC1
ITGA2
CD55
CD24
CD66
MPB
PB
SPB
AB
G
CEACAM1
H
CEACAM5
I
CEACAM6
J
CD66 so re
secretory
SPB
K
CD66 KRT5
L
% CD66+ KRT5+ basal cells
Normal lung

Figure 3

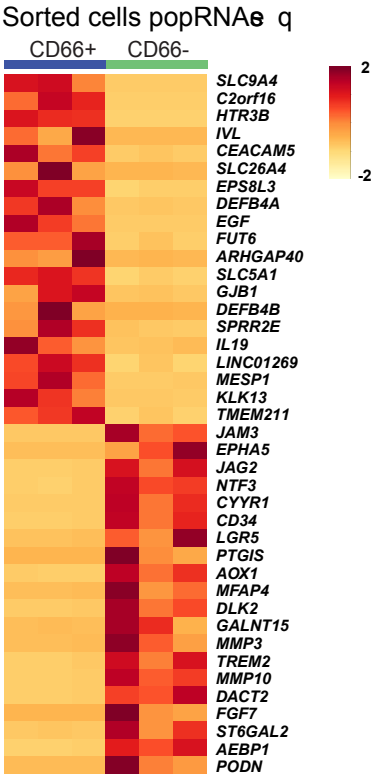
A



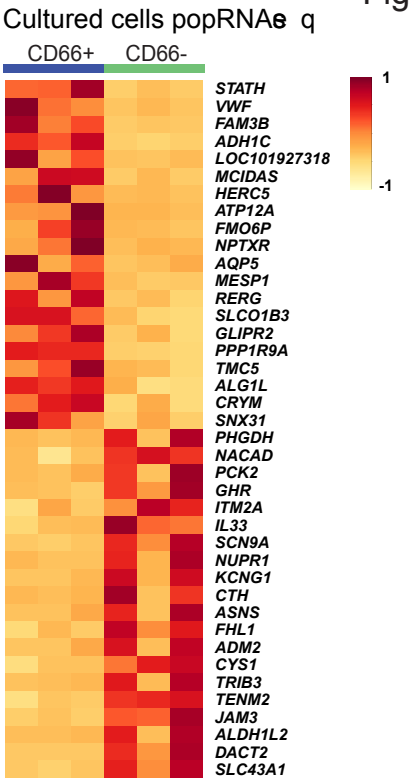
B



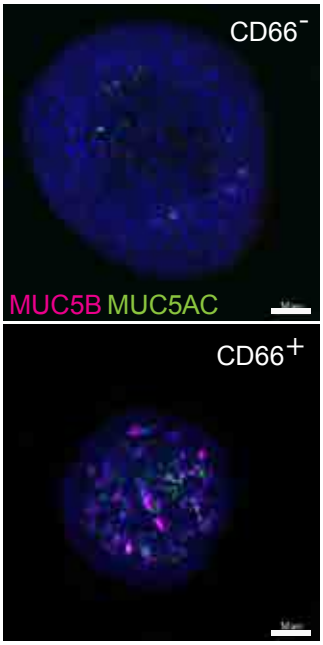
C



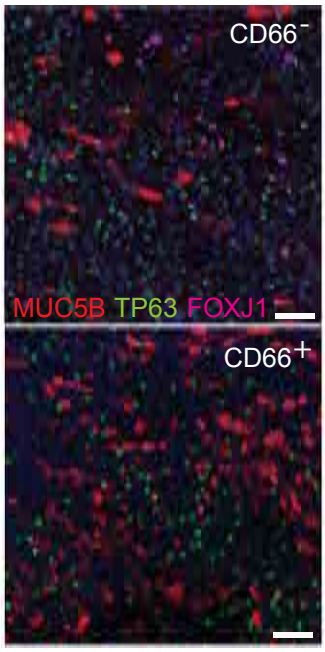
D



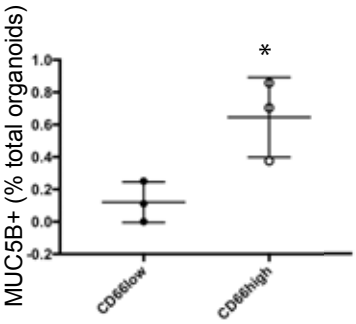
E



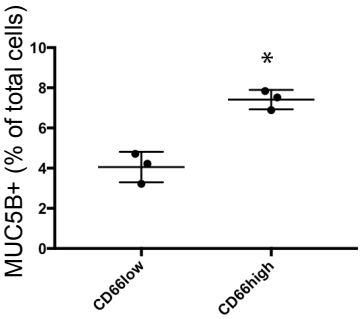
F



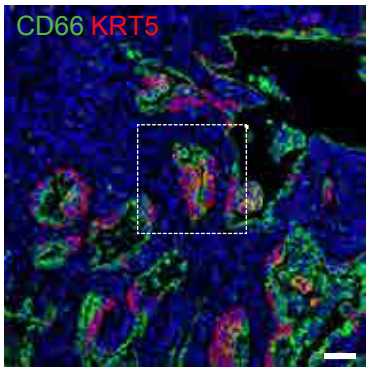
G



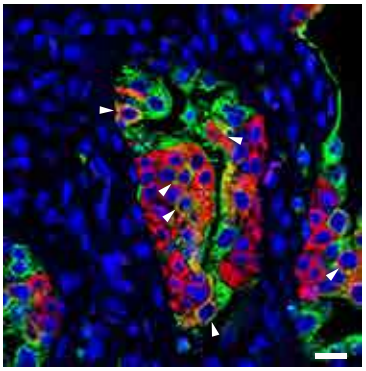
H



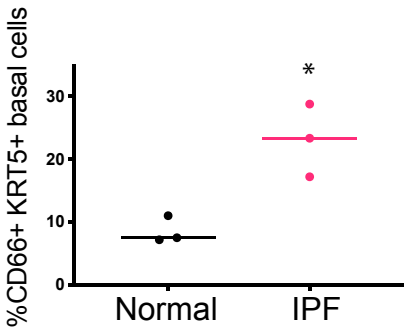
I



J



K



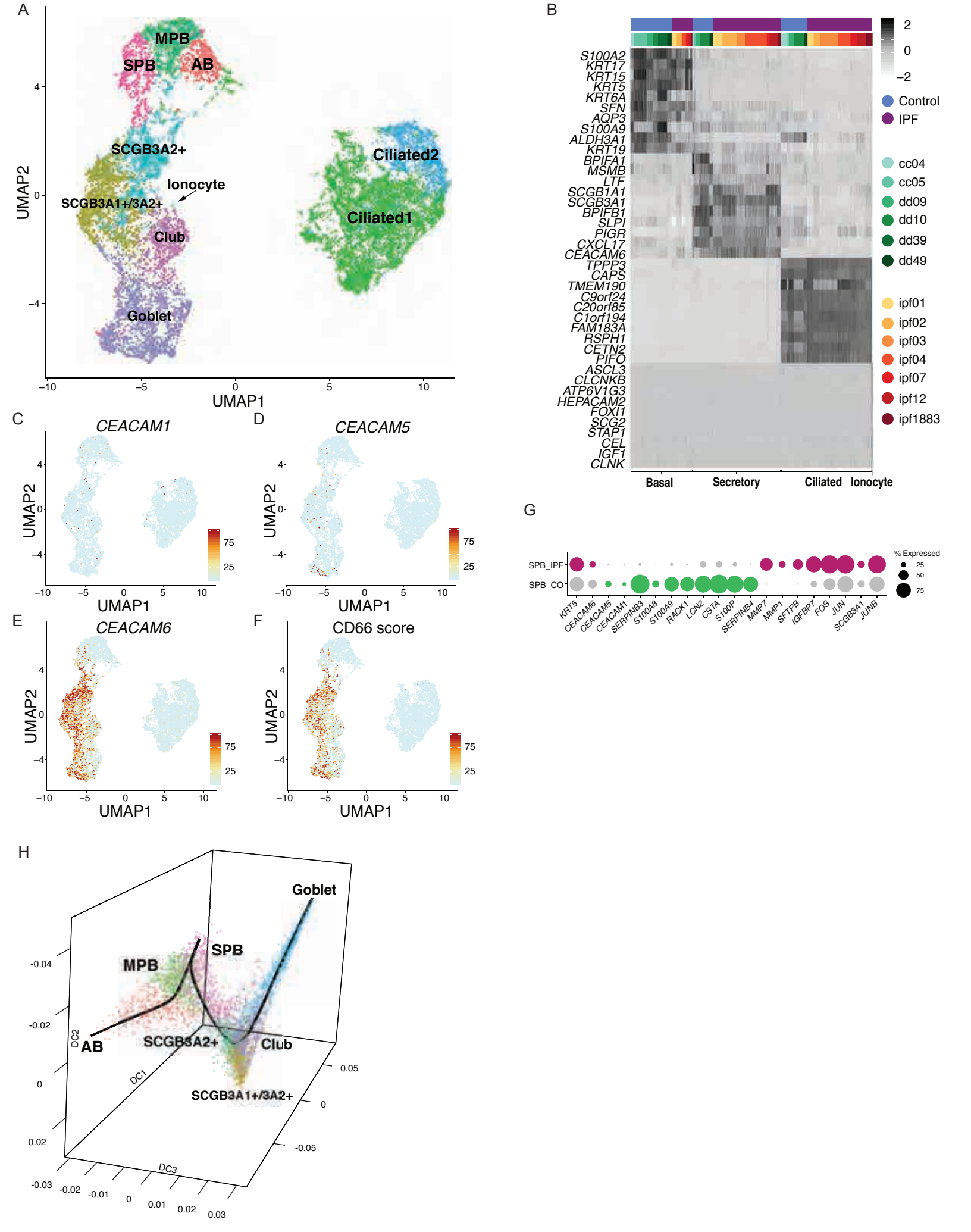
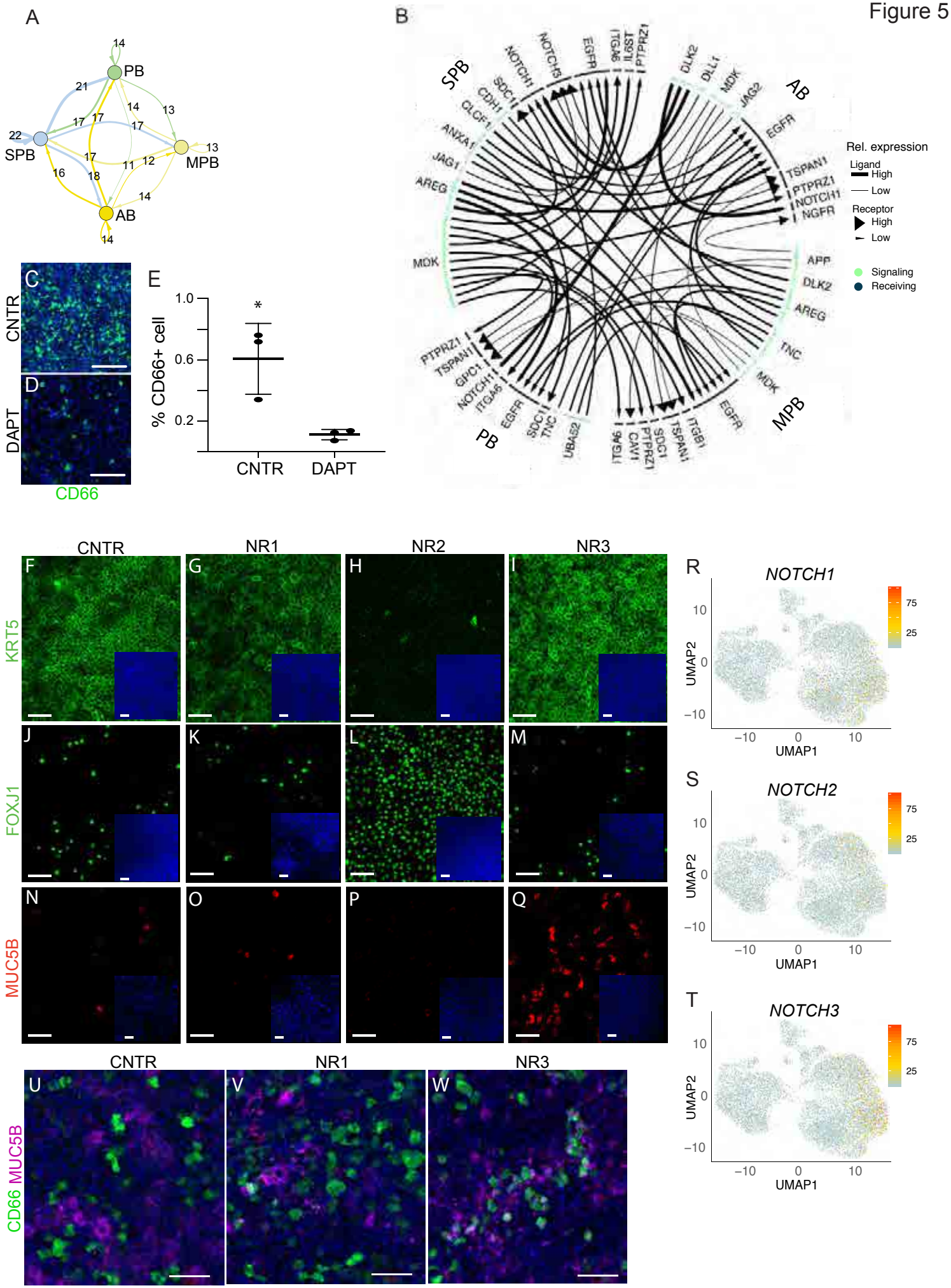

A
MPB
SPB
AB
SCGB3A2+
Ionocyte
SCGB3A1+/3A2+
Club
Goblet
Ciliated2
Ciliated1
UMAP2
UMAP1
B
S100A2
KRT17
KRT15
KRT5
KRT6A
SFN
AQP3
S100A9
ALDH3A1
KRT19
BPIFA1
MSMB
LTF
SCGB1A1
SCGB3A1
BPIFB1
SLPI
PIGR
CXCL17
CEACAM6
TPPP3
CAPS
TMEM190
C9orf24
C20orf85
C1orf194
FAM183A
RSPH1
CETN2
PIFO
ASCL3
CLCNKB
ATP6V1G3
HEPACAM2
FOXI1
SCG2
STAP1
CEL
IGF1
CLNK
Basal
Secretory
Ciliated
Ionocyte
Control
IPF
cc04
cc05
dd09
dd10
dd39
dd49
ipf01
ipf02
ipf03
ipf04
ipf07
ipf12
ipf1883
C
CEACAM1
D
CEACAM5
E
CEACAM6
F
CD66 score
G
SPB_IPF
SPB_CO
% Expressed
KRT5
CEACAM6
CEACAM5
CEACAM1
SERPINB3
S100A8
S100A9
RACK1
LCN2
CSTA
S100P
SERPINB4
MMP7
MMP1
SFTPB
IGFBP7
FOS
JUN
SCGB3A1
JUNB
H
Goblet
SPB
MPB
AB
SCGB3A2+
Club
SCGB3A1+/3A2+
DC1
DC2
DC3

Figure 5


A
PB
SPB
MPB
AB
B
SPB
AB
MPB
PB
Rel. expression
Ligand
High
Low
Receptor
High
Low
Signaling
Receiving
C
D
CNTR
DAPT
CD66
E
% CD66+ cell
1.0
0.6
0.2
CNTR
DAPT
CNTR
NR1
NR2
NR3
KRT5
FOXJ1
MUC5B
CD66 MUC5B
R
NOTCH1
S
NOTCH2
T
NOTCH3
UMAP1
UMAP2