Single Cell Reconstruction of Human Basal Cell Diversity in Normal and IPF Lung

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

ONLINE DATA SUPPLEMENT

Supplementary Material and Methods

In vitro cultures

Two thousand FACS enriched cells were resuspended with a 1:1 (v/v) ratio of Matrigel (Corning) and culture medium and 100uL seeded onto the apical surface of a 12 mm, hydrophilic PTFE 0.4 μm pore-size cell culture insert (EMD Millipore) placed inside a 24-well flat-bottom plate. After polymerization of the Matrigel, 500μL of basal medium (PneumaCult™-Ex Medium, STEMCELL Technologies) was added to the well, outside the cell culture insert. Cultures maintained at 37°C in a humidified incubator (5% CO_2) and continued to be maintained in basal medium with media changes every other day. To evaluate differentiation potential and self-renewal capacity of $CD66^+$, $CD66^-$ basal cells, cultures were grown for up to fourteen days, at which time they were either fixed for histological analysis or organoids harvested and dissociated to yield single cell suspensions for serial passage. Colony forming efficiency was determined by counting the number of colonies with a diameter of ≥ 50μm in each culture and shown as a percentage of input epithelial cells from images taken with the EVOS® XL Core transmitted-light inverted imaging system (ThermoFisher Scientific). FIJI was used to quantify colony forming efficiency from three independent experiments. Statistical analysis was performed using GraphPad Prism version 7.0a. For testing significance of the results, one-way ANOVA test was used. For serial passaging experiments, whole cultures dissociated by Liberase (Sigma-Aldrich) at 0.25 Wünsch U/mL in HBSS warmed to 37°C freed organoids from the Matrigel. After 15 minutes incubation at 37°C, cells were placed on a Thermomixer (Eppendorf) and agitated at 1000 rpm at 37°C for 20 minutes. Cells pelleted by centrifugation at 400g for 5 minutes at 4°C, followed by another

dissociation step using 2mL of TrypLE™ Express Enzyme (Gibco) to generate single-cell suspensions. Once single-cell suspensions were attained and visually confirmed under light microscopy, cells were washed in HBSS-2%FBS to stop the reaction. Live cell counts were performed on resuspended cells in 0.5-1mL HBSS-2%FBS using a hemocytometer and Trypan Blue (Sigma-Aldrich). For air liquid interface (ALI) culture experiments, isolated $CD66^+$ and $CD66^-$ basal cells were expanded and seeded onto the apical surface of a 12 mm, hydrophilic PTFE, 0.4 μm cell culture inserts (EMD Millipore) coated with matrigel and placed inside a 24-well flat-bottom plate. Cells were grown with 500μL of basal or ALI medium (STEMCELL Technologies) according to manufacture instructions. Cultures were maintained at 37°C in a humidified incubator (5% CO_2). For studies of NOTCH signaling, the γ-secretase inhibitor (2*S*)-*N*-[(3,5-Difluorophenyl)acetyl]-L-alanyl-2-phenyl]glycine1,1-dimethylethyl ester (DAPT) (Tocris) was added at a concentration of 5 μM. Blocking antibodies for NOTCH1 (NR1), NOTCH2 (NR2), or NOTCH3 (NR3), were added individually at concentrations of 30 μg/ml starting at the beginning of the culture, and media was changed every three days. Notch blocking antibodies were provided by Genentech.

High-throughput FACS analysis

Samples of bronchi, bronchiolar and distal lung tissue were pooled, dissociated into a single cell suspension and used for cell screen analysis. Basal cell subset stained with APC anti-human CD271, AF488 anti-human CD326, Brilliant Violet 421 anti-human CD45 and CD31 combined with a panel of cell-surface proteins targeting 332 PE-conjugated monoclonal antibodies (Biolegend, LEGENDScreen Human PE Kit, 700007) and 10

isotype controls (Biolegend) identified novel candidate markers. Cell preparation for screening analysis were followed per manufacturer's protocol (Biolegend). In brief, cells filtered through a 40µm cell strainer were resuspended at a density of between 1.5 - 2 x 10^6 cells/mL in 30mLs of HBSS-2%FBS. Prior to sample staining, cells pre-incubated with human Fc-receptor blocking solution (BioLegend) to reduce nonspecific staining carried out at a concentration of (0.01%) for 15 minutes on ice. After lyophilized antibodies on plates reconstituted, plates ready for sample staining had 75µL/well of cells aliquoted and incubated in the dark for 30 minutes on ice. Cells were washed twice and fixed for 10 minutes at room temperature prior to flow cytometer analysis. Analyses were performed on labeled cells using an LSRFortessa cell analyzer (Becton Dickinson, BD) and collected data collected were interrogated using FlowJo software.

Population RNAseq

Basal cell $CD66^+$ and $CD66^-$ populations were processed immediately or after in vitro culturing for ten days. RNA was extracted from the sorted cells using the RNeasy Micro Kit (Quiagen). Total RNA was quantified and sample contamination by proteins or carryover reagents were assessed using both a NanoDrop and a Qubit (Thermo Scientific). Samples were then qualified using the Fragment Analyzer (Advanced Analytical Technologies) to check the integrity of total RNA by measuring the ratio between the 18S and 28S ribosomal peaks. SMART-Seq V4 Ultra Low RNA Input Kit for Sequencing (Takara Bio USA, Inc., Mountain View, CA) was used for reverse transcription and generation of double stranded cDNA for subsequent library preparation using the Nextera XT Library Preparation kit (Illumina, San Diego, CA). Reaction Buffer (1X) was added to lysed cells and the RNA was used directly for oligo(dT)-primed reverse

transcription, followed by cDNA amplification and cleanup. Quantitation of cDNA was performed using Qubit (Thermo Fisher Scientific). cDNA normalized to 80 pg/µl was fragmented and sequencing primers added simultaneously. A limiting-cycle PCR added Index 1 (i7) adapters, Index 2 (i5) adapters, and sequences required for cluster formation on the sequencing flow cell. Indexed libraries were pooled and cleaned up, and the pooled library size was verified on the 2100 Bioanalyzer (Agilent Technologies) and quantified via Qubit. Libraries were sequenced on a NextSeq 500 (Illumina) using with a 1x75 bp read length and coverage of over 25M reads per sample. Population RNAseq reads were processed using packages inside Galaxy (https://usegalaxy.org/). Specifically, alignment to human genome hg38 was performed using STAR 2.6.0b-1 with application of the default setting in Galaxy. Raw counts were measured using FeatureCounts 1.6.0.3 with application of the default setting in Galaxy. The annotation GTF file of USCS hg38 was downloaded from iGenomes (https://support.illumina.com/sequencing/sequencing_software/igenome.html).
Differential gene expression was determined using DESeq2(4) (version 2.11.40.2) using default parameters in Galaxy. The Top DEG for freshly isolated and cultured datasets, along with raw and normalized counts, are shown in Supplementary Table 5.

Heatmaps in Fig. 3C-D were produced selecting the top 20 differentially expressed genes in each group on the base of their log2 fold change among genes with adjusted p-values < 0.05. Heatmaps in Fig. S4A-B were produced as follows: for popRNAseq DEG, only genes with adjusted p-value < 0.05 were used for analysis. A DEG of scRNAseq data of MPB and SPB cells was produced (Supplementary Table 5) and only genes with adjusted p-value < 0.05 were used in the analysis. scRNAseq DEG were used to define

PopRNAseq DEG that were used in the analysis. For freshly isolated cells the top 20 DEG's for CD66+ and CD66- where used for creation of the heatmap. For popRNAseq of cultured basal cells, only 13 genes passed the filtering and were used for creation of the heatmap.

Single cell RNAseq

Single cells were captured using a 10X Chromium device (10X Genomics) and libraries were prepared according to the Single Cell 3' v2 or v3 Reagent Kits User Guide (10X Genomics, Supplementary Table 7 specifies the kit version used for each dataset). Totalseq-A human hashing antibodies (Biolegend) were used to label cells isolated from sample cc05; trachea or extralobar bronchi used TotalSeq-A0253 (394605), proximal intra-lobar bronchi used TotalSeq-A0254 (394607). Hashing was performed for a different purpose and was not considered for data processing in this manuscript. Cellular suspensions were loaded on a Chromium Controller instrument (10X Genomics) to generate single-cell Gel Bead-In-EMulsions (GEMs). Reverse transcription (RT) was performed in a Veriti 96-well thermal cycler (ThermoFisher). After RT, GEMs were harvested, and the cDNA underwent size selection with SPRIselect Reagent Kit (Beckman Coulter). Indexed sequencing libraries were constructed using the Chromium Single-Cell 3' Library Kit (10X Genomics) for enzymatic fragmentation, end-repair, A-tailing, adapter ligation, ligation cleanup, sample index PCR, and PCR cleanup. The barcoded sequencing libraries were quantified by quantitative PCR using the KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA). Sequencing libraries were loaded on a NovaSeq 6000 (Illumina) with a sequencing setting of 26bp and 98bp for v2

and 28bp and 91bp for v3, respectively, to obtain a sequencing depth of at least ~5 x 10^4 reads per cell.

Data analysis

Cell Ranger software (10X Genomics) was used for mapping and barcode filtering. Briefly, the raw reads were aligned to the transcriptome using STAR(5), using a hg38 transcriptome reference from Ensembl 93 annotation. Expression counts for each gene in all samples were collapsed and normalized to unique molecular identifier (UMI) counts. The result is a large digital expression matrix with cell barcodes as rows and gene identities as columns.

Data analysis was mainly performed with Seurat 3.0(6) with some variation that will be described.

For all data, quality control and filtering were performed separately for each dataset, to remove cells with low number of expressed genes (threshold n>=200) and elevated expression of apoptotic transcripts (threshold mitochondrial genes < 15%). Only genes detected in at least 3 cells were included. Each dataset was run separately with SoupX analysis package to remove contaminant 'ambient' RNA derived from lysed cells during isolation and capture (Young MD et al., https://doi.org/10.1101/303727). We performed a correction on the basis of genes with a strong bimodal distribution and for which the 'ambient' RNA expression was overlapping with a gene signature of a known cell type. The 'adjustCounts' function of SoupX was used to generate corrected count's matrices. Supplementary Table7 show the percentage of ambient RNA identified in each dataset. To minimize doublet contamination for each dataset we performed a quantile thresholding

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

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

Top differentially expressed genes were fed to STRING (https://string-db.org/) to visualize their protein association network and extract their functional enrichments. Gene list of specific terms were used to create a score, as described above, and visualized as UMAP.

IPF datasets were first processed as described for normal datasets, from QC analysis to integration. For comparative analysis of control and IPF, a common integration of all datasets was performed following the same integration strategy described for control datasets. The integrated Seurat object containing control and IPF data was used to produce differential gene expression analysis using MAST (Supplementary Table 6). This analysis was used to generate visualizations in Fig.4 B and G.

To estimate the presence of cellular transitions between basal and secretory clusters, we used diffusion map with 'Destiny'(12), followed by trajectory analysis with 'Slingshot'(13), following their Bioconductor vignettes. The normalized counts from Seurat were imported into an ExpressionSet for Destiny and into a SingleCellExperiment for Slingshot. The 3D visualization was obtained using the 'rgl' library in R. Notch signaling ligand-receptor interactions were analyzed using iTALK (Wang Y et al., https://doi.org/10.1101/507871). Supplementary Table 8 show the Notch signalling database. All analyses were performed using R 3.6.

Statistical analysis

FIJI was used to quantify colony forming efficiency from three biological samples in at least eight technical replicates. Statistical analysis was performed using GraphPad Prism version 7.0a. Data presented are shown as mean ratio ± SEM. For testing significance of the results, one-way ANOVA test was used. For quantification of immunofluorescence

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.

Legends

Supplementary Tables

Supplementary Table 1. Differential gene expression of major cell types. pValue, average fold change, pct, and adjusted pValue are shown.

Supplementary Table 2, A.B.Jaffe gene list of major human airway cell types.

Supplementary Table 3. Differential gene expression of subsets of basal, secretory, and ciliated cells. pValue, average fold change, pct, and adjusted pValue are shown.

Supplementary Table 4. List of CD markers screened against scRNAseq data.

Supplementary Table 5. Raw counts, normalized counts, and differential gene expression of population RNAseq from Galaxy. It contains also DEG from comparison with MPB and SPB cells.

Supplementary Table 6. Differential gene expression of Control vs IPF clusters. pValue, average fold change, pct, and adjusted pValue are shown.

Supplementary Table7. QC information for control and IPF datasets. Included are details of 10X kit version per dataset, % of ambient RNA correction for each sample, and other metrics.

Supplementary Table 8. Database of Notch signalling, ligand-receptor pathway.

Fig S1. Supplementary to figure 1. (A) Characteristics of donor and patients for six normal and seven IPF samples used for scRNAseq. (B-E) Normal and IPF sample

contribution to major cell populations and basal cell subclusters, visualized by a stacked column chart. (F, G) Visualization of the distribution of the six normal and seven IPF samples of origin within clustering groups visualized by UMAP. (H) *PCNA* expression in IPF datasets visualized by UMAP. (I) Scoring derived from Serpin family gene signature visualized by UMAP. (J) Scoring of early ciliogenesis markers expressed at the time of FOXJ1 initial expression. visualized by UMAP. (K) Scoring of p38-SAPK signaling, visualized by UMAP.

Fig S2. Supplementary to figure 1. (A) Large clustering community of normal lung datasets obtained with Leiden algorithm (resolution set to 3) visualized by UMAP. (B) Normal lung datasets segregation in major cell type at the end of the refinement score analysis, visualized by UMAP. (C-L) Progressive assignment of major cell types to the clustering assigned in A. Initial assignment rely on previously published gene sets. Differential gene expression within assigned clusters is used to refine the assignment score. The procedure was reiterated when necessary.

Fig. S3. Supplementary to figure 2. In the representative FACS data in Fig 2D we sequentially gated for viable cells (FSC and propidium iodide exclusion; 54.6% of all events), side scatter singlets (61.8%), forward scatter singlets (96.1%), lineage negative (CD45- and CD31-; 22.5%), epithelial cells (Epcam+; 24.6%) and basal cells (NGFR+; 42.1%).

Fig S4. Supplementary to figure 2. (A, B) Transcriptomic DEG overlap between popRNAseq and scRNAseq is shown. DEG of $CD66^{+}$ versus $CD66^{-}$ basal cells from popRNAseq, were compared to DEG of SPB versus MPB cells from scRNAseq. For

sorted cells, the top 20 shared genes are shown. For cultured cells only 13 genes were shared and are shown. An adjusted p-value < 0.05 was adopted for genes cut-off. Top DEG genes were selected from their log2 fold change values. (C) Normal lung cross-section shows CD66 (green), KRT5 (red), and KRT8 (purple). Immunoreactivity for CD66 is localized in a basal cell subpopulation (arrowheads), transitional KRT8+ cells (arrows), and luminal secretory cells (green arrowheads). Ciliated cells, recognizable in phase contrast, do not show immunoreactivity for CD66. Scale bar: 20 mm.

Fig S5. Supplementary to figure 3. (A, B) Bright field representative image of entire well for $CD66^{+}$ and $CD66^{-}$ basal cells grown in mesenchyme-free three-dimensional cultures. (C, D) Immunofluorescence representative tiling image for detection of mucins MUC5B and MUC5AC in organoid cultures of $CD66^{+}$ and $CD66^{-}$ basal cells.

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.

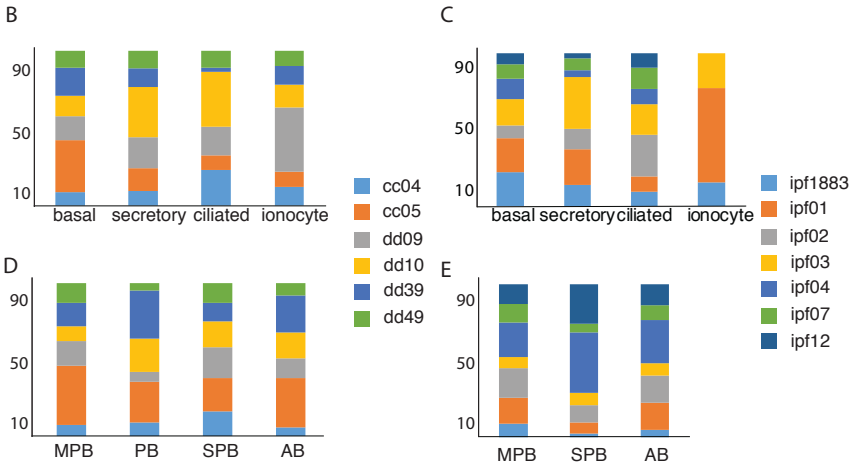
12. Angerer P, Haghverdi L, Buttner M, Theis FJ, Marr C, Buettner F. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* 2016; 32: 1241-1243.

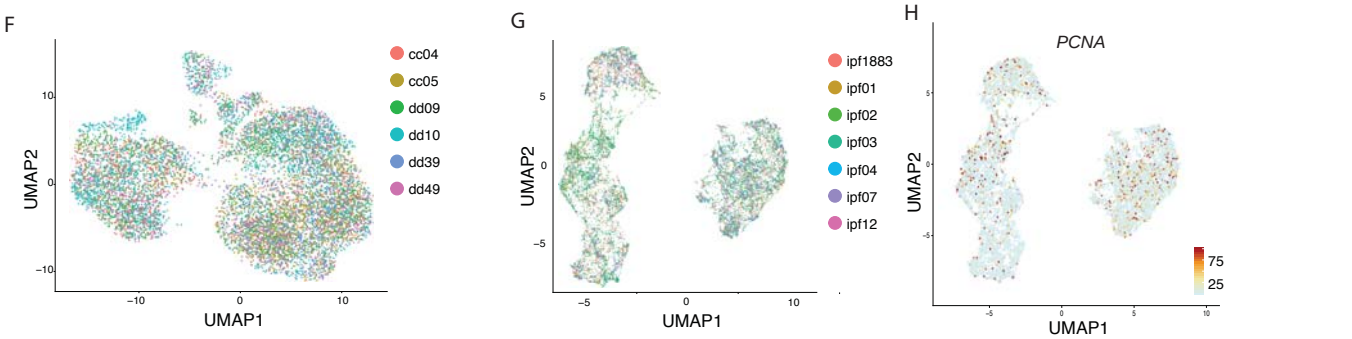
13. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, Dudoit S. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 2018; 19: 477.

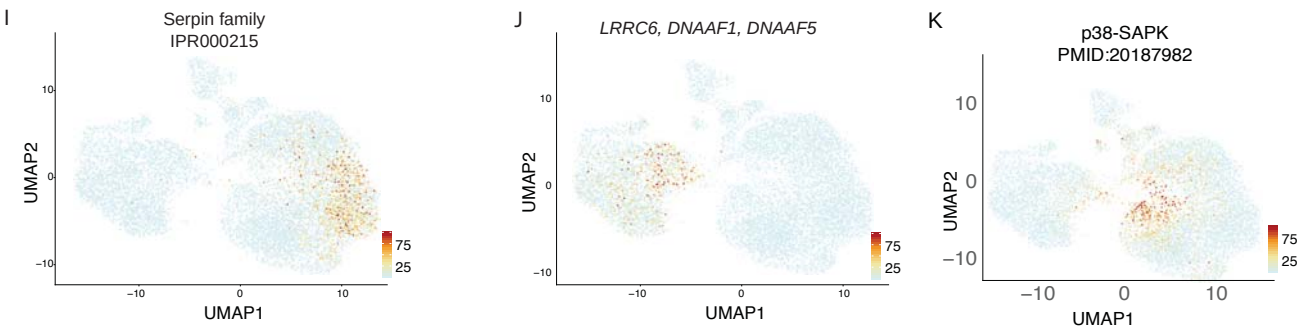
Figure S1

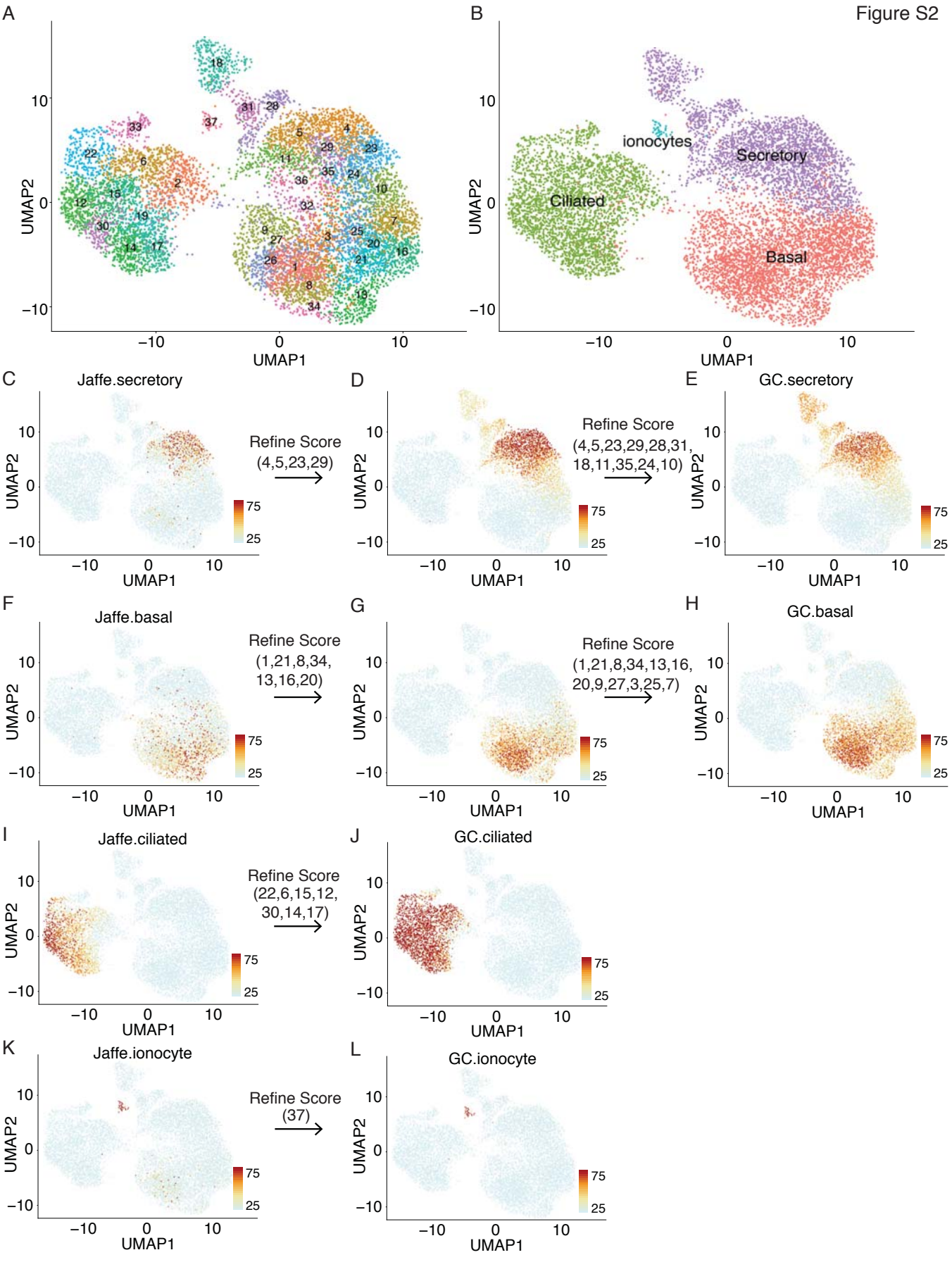
A

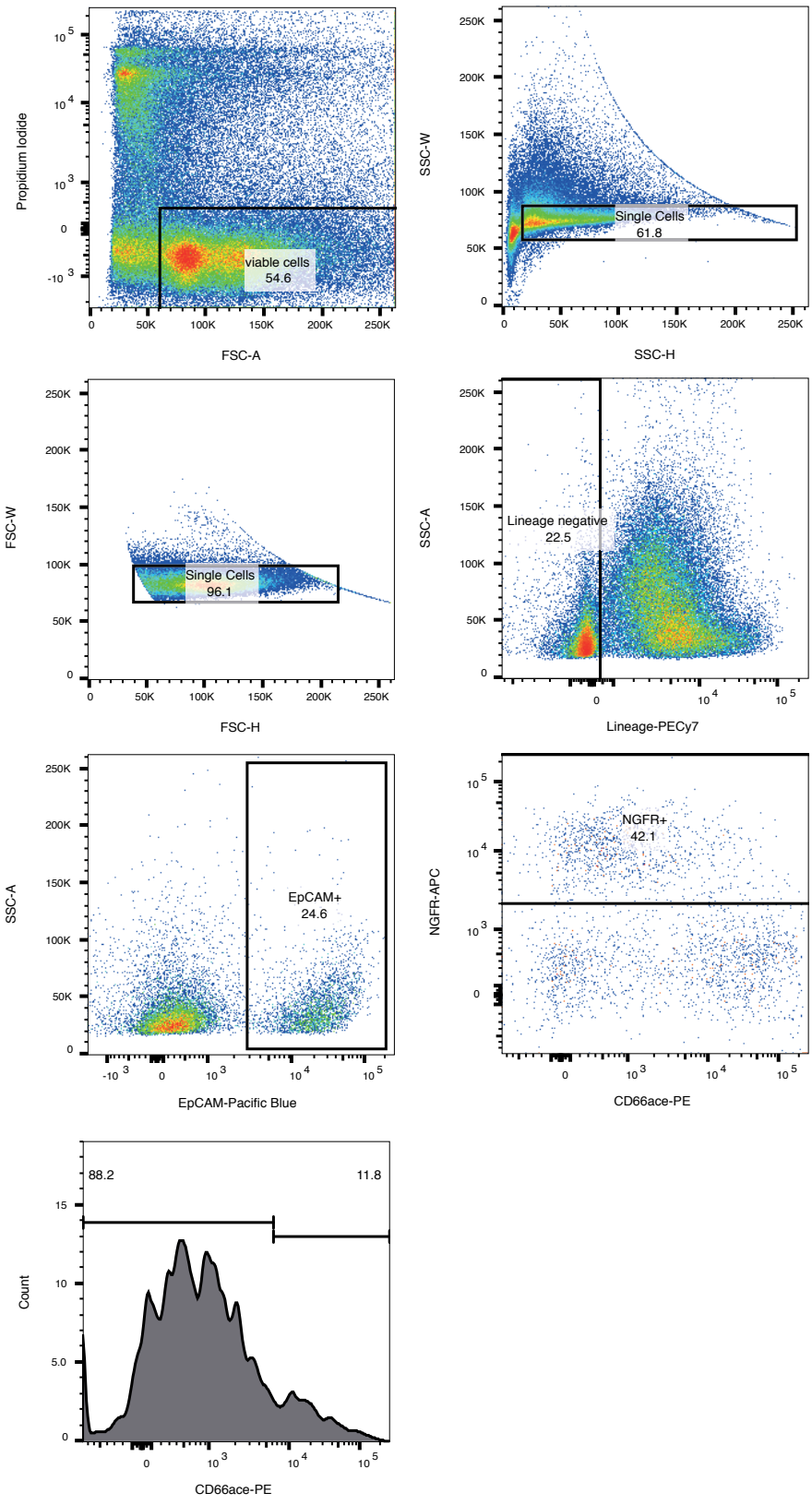
Platform	10X Kit	Type	ID	Age	Gender	Smoking History
10X	v3	Normal	cc04	47	M	smoker (½ py)
10X	v3	Normal	cc05	63	F	smoker (½ py)
10X	v2	Normal	dd09	53	F	smoker (4.5 py)
10X	v3	Normal	dd10	24	M	non-smoker
10X	v3	Normal	dd39	52	M	non-smoker
10X	v3	Normal	dd49	48	F	smoker (13 py)
10X	v2	IPF	ipf1883	64	F	non-smoker
10X	v2	IPF	ipf01	64	F	non-smoker
10X	v2	IPF	ipf02	62	F	smoker (9 py)
10X	v3	IPF	ipf03	72	M	smoker (quit 1/1/15)
10X	v2	IPF	ipf04	65	F	non-smoker
10X	v2	IPF	ipf07	68	F	non-smoker
10X	v3	IPF	ipf12	64	M	non-smoker

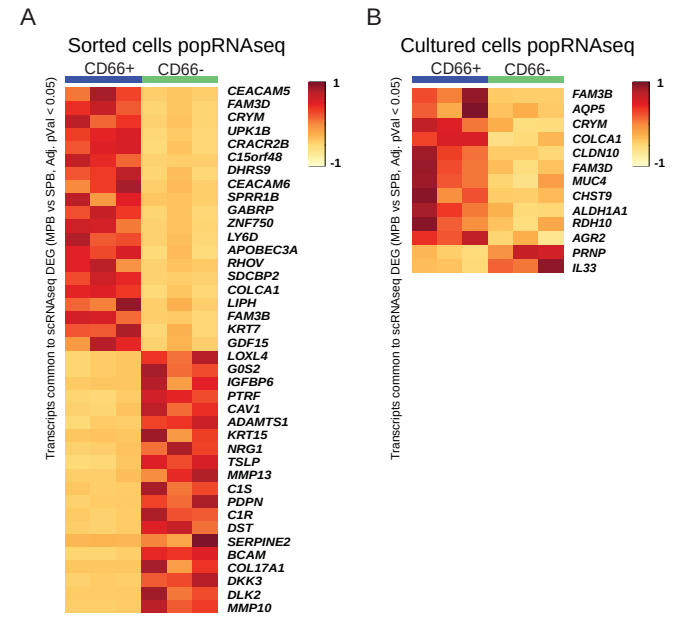




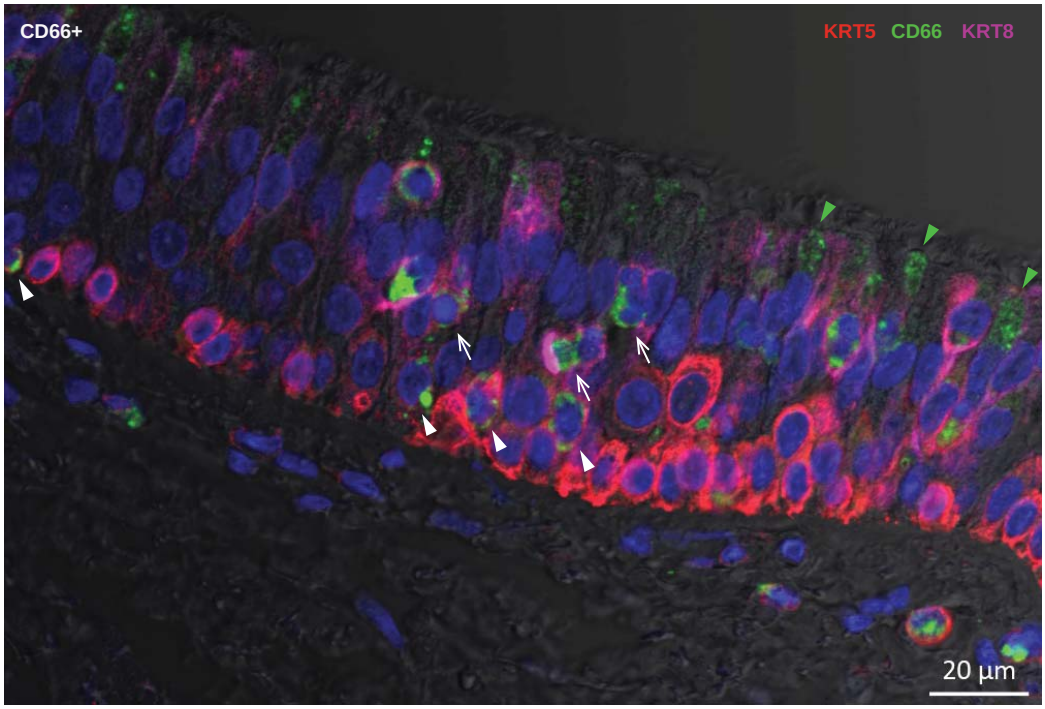


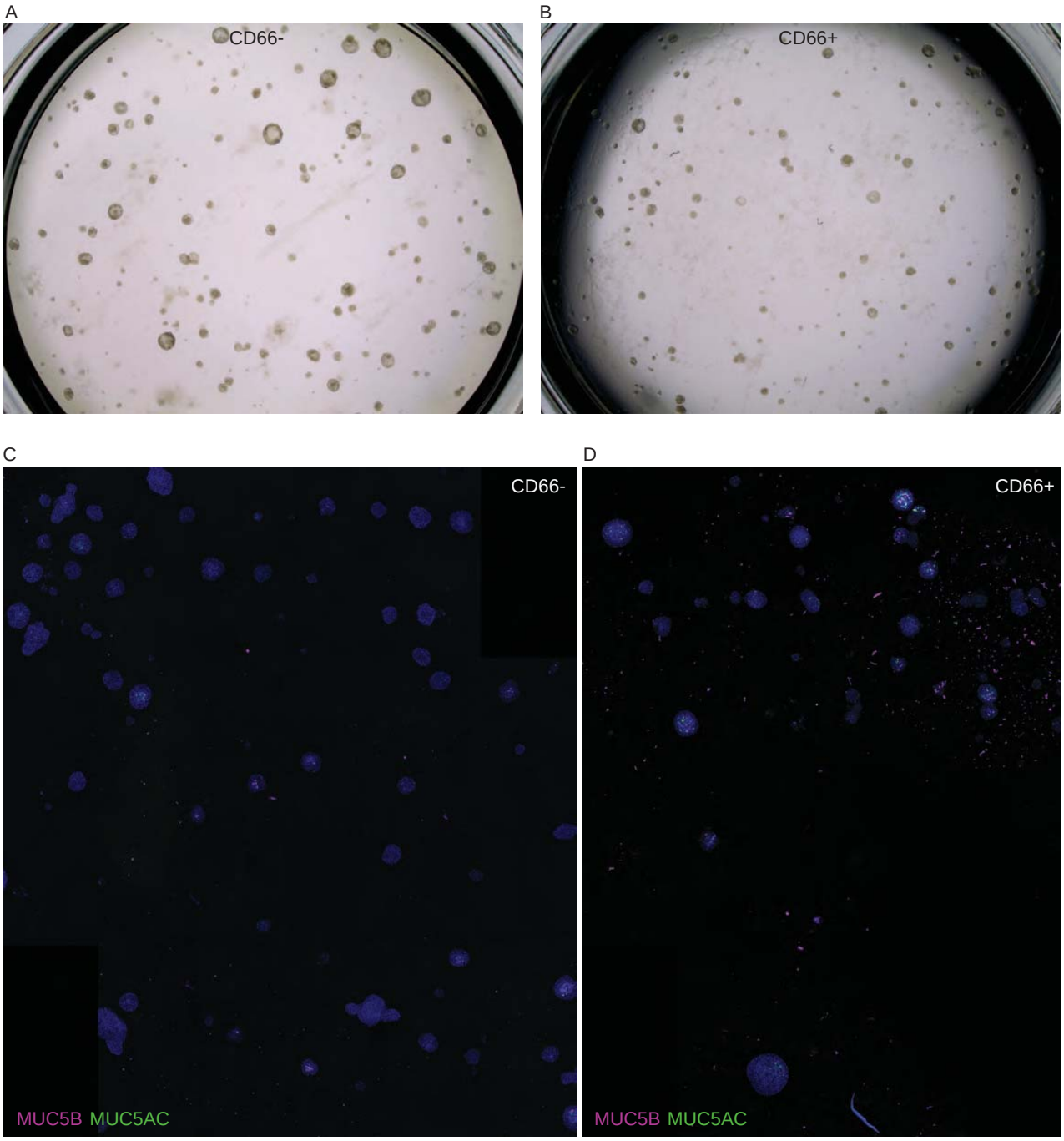

A
B
ionocytes
Secretory
Ciliated
Basal
C
Jaffe.secretory
Refine Score
(4,5,23,29)
D
Refine Score
(4,5,23,29,28,31,
18,11,35,24,10)
E
GC.secretory
F
Jaffe.basal
Refine Score
(1,21,8,34,
13,16,20)
G
Refine Score
(1,21,8,34,13,16,
20,9,27,3,25,7)
H
GC.basal
I
Jaffe.ciliated
Refine Score
(22,6,15,12,
30,14,17)
J
GC.ciliated
K
Jaffe.ionocyte
Refine Score
(37)
L
GC.ionocyte
UMAP1
UMAP2
75
25


Propidium Iodide
FSC-A
viable cells
54.6
SSC-W
SSC-H
Single Cells
61.8
FSC-W
FSC-H
Single Cells
96.1
SSC-A
Lineage-PECy7
Lineage negative
22.5
SSC-A
EpCAM-Pacific Blue
EpCAM+
24.6
NGFR-APC
CD66ace-PE
NGFR+
42.1
Count
CD66ace-PE
88.2
11.8


A
Sorted cells popRNAseq
CD66+
CD66-
Transcripts common to scRNAseq DEG (MPB vs SPB, Adj. pVal < 0.05)
CEACAM5
FAM3D
CRYM
UPK1B
CRACR2B
C15orf48
DHRS9
CEACAM6
SPRR1B
GABRP
ZNF750
LY6D
APOBEC3A
RHOV
SDCBP2
COLCA1
LIPH
FAM3B
KRT7
GDF15
LOXL4
G0S2
IGFBP6
PTRF
CAV1
ADAMTS1
KRT15
NRG1
TSLP
MMP13
C1S
PDPN
C1R
DST
SERPINE2
BCAM
COL17A1
DKK3
DLK2
MMP10
1
-1
B
Cultured cells popRNAseq
CD66+
CD66-
Transcripts common to scRNAseq DEG (MPB vs SPB, Adj. pVal < 0.05)
FAM3B
AQP5
CRYM
COLCA1
CLDN10
FAM3D
MUC4
CHST9
ALDH1A1
RDH10
AGR2
PRNP
IL33
1
-1

C


CD66+
KRT5 CD66 KRT8
20 µm


A
CD66-
B
CD66+
C
CD66-
MUC5B MUC5AC
D
CD66+
MUC5B MUC5AC