Supplementary Methods

Human diseased and control explanted lungs were procured from donors with end-stage lung disease undergoing transplant or control lungs rejected for transplant, per protocols approved by the Partners IRB with informed consent.

Tissue digestion and biobanking

Lungs were sliced and washed with cold sterile PBS several times. For each explant, approximately 100-150 g of parenchymal lung tissue was obtained from the peripheral third of left and/or right lungs (including all lobes). Approximately 80% of this tissue was digested to generate the cell suspensions described below, while the remaining 20% was snap frozen as small pieces of whole lung tissue.

To generate cell suspensions, visible airway structures, vessels, blood clots and mucin were removed from whole lung tissue. Remaining tissue was transferred to a sterile 15 cm dish and ~12.5 ml of digestion medium, consisting of 30 U/ml Elastase (Elastin Products Company, Owensville, MO), 0.2 mg/ml DNAse I (Sigma, St. Louis, MO), 0.3 mg/ml Liberase (Roche, Basel, Switzerland) and 1% Penicillin/Streptomycin diluted in DMEM/F12 medium (Lonza, Basel, Switzerland) were added. Tissue was minced mechanically into small pieces (<5 mm). The mixture was transferred to a 50 ml tube and then incubated on a rocker for 45 minutes in 37°C. in Digested tissue was filtered using a metal strainer (mesh size: 40-100 μm) (Sigma). The filter solution was kept on ice. Unfiltered tissue was incubated a second time in digestion medium for 30 minutes or less, followed by repeat filtration and addition of 10% FBS to stop the enzymatic reaction. Flow-through from both filtrations was combined and centrifuged at 600G, 4°C for 10 minutes. The pellet was resuspended in up to 20 ml of red cell lysis buffer (VWR, Radnor, PA) for <5 minutes in 37°C. Up to 20 ml PBS with 10% FBS were added to stop the reaction. The solution was then centrifuged again at 600G, 4°C for 10 minutes. The pellet was resuspended in DMEM/F12 medium and filtered using a 100 μm strainer (Fisher Scientific, Waltham, MA). Freezing medium (10% FBS and 10% DMSO in DMEM/F12) was added to the filtrate. Cell suspensions were then aliquoted into cryovials (typically 81 aliquots with 10^7 – 10^8 cells/ml), frozen overnight in freezing containers at -80 °C, and then stored in liquid nitrogen until further use.

Flow cytometry and cell sorting

Biobanked cell suspensions were thawed rapidly in a 37°C water bath, centrifuged at 600G and washed in cold PBS, then filtered through a 100 μm strainer to remove cell clumps. Filtrates were centrifuged at 600g, 4°C for 10 minutes. Samples, isotype and single stain controls were prepared in buffer containing using magnesium chloride 5 mM, DNase I 50 ug/ml, and 2% FBS. Cells were stained using the following antibodies and molecules: PE anti-human CD326 (EpCAM) (eBioscience, San Diego, CA), FITC anti-human CD45 (BD bioscience 340664), Alexa Fluor® 647 anti-human podoplanin (PDPN) (BioLegend, San Diego, CA), and DAPI (BioLegend) or Propidium Iodide. Isotype-matched antibodies were used as negative controls. Stained cells were examined by BD FACSCanto II (BD Biosciences, Franklin Lakes, NJ) or sorted by BD FACSAria II (BD Biosciences). Gates were defined as: $EpCAM^{+}/PDPN^{low}$ (P1 gate) and $EpCAM^{high}/PDPN^{-}$ (P2 gate). Flow cytometric data were analyzed by FlowJo software version 10.1.

RNA extraction and purification
Digested unsorted cell suspensions and sorted cells were lysed using QIAzol (Qiagen) per the manufacturer's instructions. The aqueous RNA phase was collected after treatment with chloroform and RNase free DNase I (Qiagen) to deplete genomic DNA contamination for 10 min at room temperature. RNA was purified using RNeasy MinElute Cleanup (Qiagen). Eluted RNA was stored in -80°C. Frozen whole lung tissue was ultrasonically homogenized in QIAzol and RNA was extracted in same manner described above.

Library construction for RNA sequencing
The Illumina TruSeq RNA Access Library Prep kit (San Diego, CA) was used for library preparation. Library construction was performed as follows: a minimum of 20 ng of total RNA was fragmented and primed for cDNA synthesis. RNA was converted to double stranded cDNA using reverse transcriptase with random priming. cDNA was purified and ligated to adapters. Successfully ligated products were amplified and enriched using PCR. Following purification by AmPure beads (Agilent Genomics, Santa Clara, CA), libraries were validated on TapeStation (Agilent Genomics) and Qubit (Thermo Fisher Scientific) before being pooled. The pools underwent two sets of hybridizations, capture washes, and elutions to capture and enrich the libraries and eliminate nonspecific binding. The enriched libraries were then amplified with a final PCR and cleaned up again using AmPure purification. Libraries were quantitated using Qubit and KAPA (Wilmington, MA). KAPA results were used to determine the flowcell loading concentration. Libraries were loaded onto the flowcell at 18 pM. Libraries were sequenced on a 75 bp paired end flowcell on the HiSeq 2500. Each lane was spiked with 5% PhiX control libraries, which served as an internal control.

Processing and analysis of RNA sequencing data
Demultiplexed sample FastQ files were downloaded onto a Grid engine (Oracle, Burlington, MA)-based high performance computing cluster sharing an NFS (Sun Microsystems, Oracle, Burlington MA) storage system with a Eucalyptus (Hewlett Packard, Palo Alto, CA)-based private cloud. Starcluster (MIT, Cambridge MA) and Chef (Seattle, WA) were used to create a compute environment where samples could be aligned in parallel. FastQC was used to visualize aggregate Phred scores. Adapters were trimmed from reads using Skewer.[1] Trimmed reads were aligned to the GRCh38 primary assembly using STAR.[2] The assembly was concatenated with the ERCC plasmid spike in genome (Thermo Fisher Scientific, Cambridge, MA) to facilitate spike-in alignment. For QC, SAMtools pileup[3] was used to provide basic variant calls for input into KING[4] to verify sample identity relationships between replicates, with the covariance matrix produced by RNASeQC.[5] Sequence quality was assessed by means of the Phred scores, mapping rate, and composition of the RNA (primarily mRNA). The R package Rsubreads[6] and DESeq2[7] were used to quantify and normalize gene expression counts. Only genes with coverage greater than 10x at a position 300 bases from the 3′ end were considered. We modeled the effects of the decay rate and RIN values by linear regression. PCA plots were generated and inspected for outliers, which were removed in downstream analysis. Statistical inference testing was performed using linear regression models in DESeq2, controlling for age, gender, and cell type percentage (based on the percentage of each cell population as measured during the flow cytometry run).

1. Jiang H, Lei R, Ding SW, et al. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC bioinformatics 2014;**15**:182.
2. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;**29**(1):15-21.
3. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;**25**(16):2078-9.
4. Manichaikul A, Mychaleckyj JC, Rich SS, et al. Robust relationship inference in genome-wide association studies. Bioinformatics 2010;**26**(22):2867-73.
5. DeLuca DS, Levin JZ, Sivachenko A, et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 2012;**28**(11):1530-2.
6. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic acids research 2013;**41**(10):e108.
7. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 2014;**15**(12):550.