Article

Differential Lysotracker Uptake Defines Two Populations of Distal Epithelial Cells in Idiopathic Pulmonary Fibrosis

Roxana Wasnick[1*], Irina Shalashova[1], Jochen Wilhelm[1,2,4,5], Ali Khadim[1,5], Nicolai Schmidt[1], Holger Hackstein[3], Andreas Hecker[6], Konrad Hoetzenecker[7], Werner Seeger[1,2,4,5], Saverio Bellusci[1,2,5], Elie El Agha[1,5], Clemens Ruppert[1,2,9], and Andreas Guenther[1, 2, 5, 6,8,9]

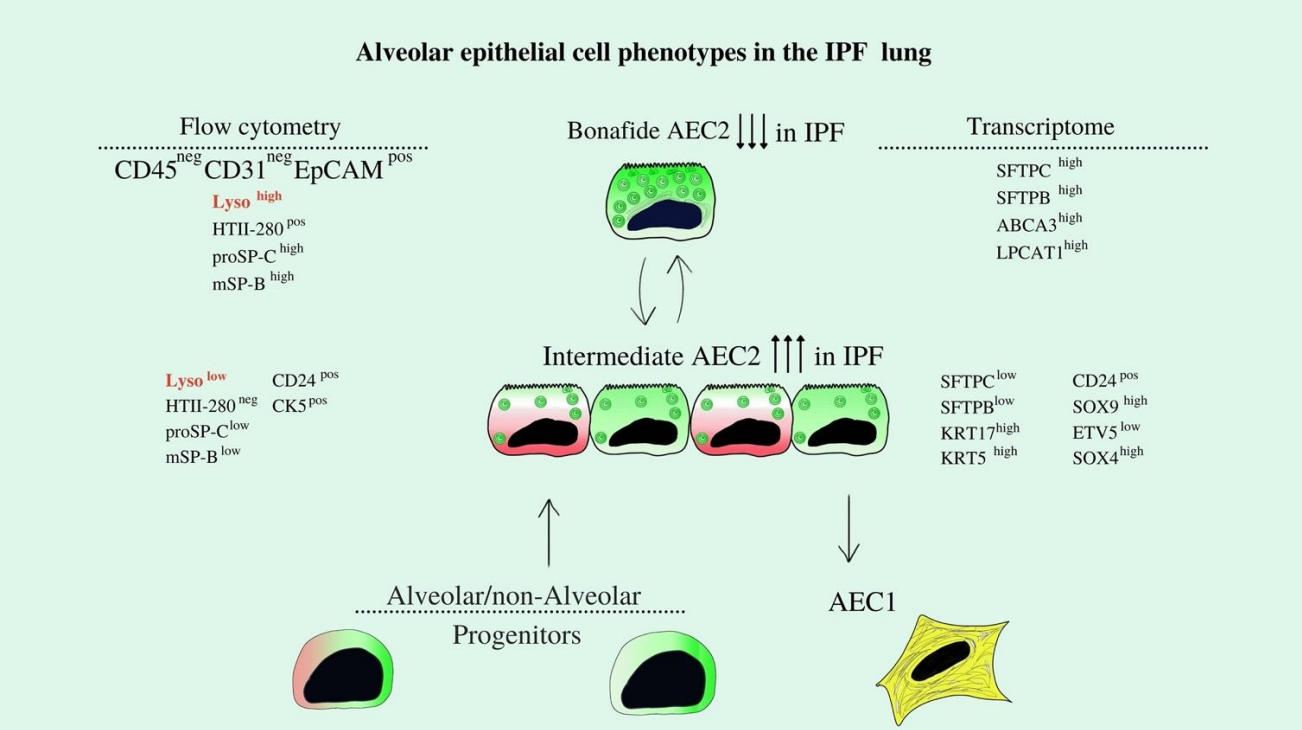
1. Universities of Giessen and Marburg Lung Center (UGMLC); member of the German Center for Lung Research (DZL), 35392 Giessen, Germany
2. Excellence Cluster Cardiopulmonary Institute (CPI), 35392 Giessen, Germany
3. Department of Clinical Immunology and Transfusion Medicine, 35392 Giessen, Germany
4. Max-Planck-Institute for Heart and Lung Research, 61231 Bad Nauheim, Germany
5. Institute for Lung Health (ILH), 35392 Giessen, Germany
6. Department of General and Thoracic Surgery, University Hospital Giessen, Giessen, Germany.
7. Department of Thoracic Surgery, Medical University of Vienna, Vienna, Austria
8. Lung Clinic Waldhof-Elgershausen, 35753 Greifenstein, Germany
9. European IPF Registry / UGLMC Giessen Biobank, 35392 Giessen, Germany

* Correspondence: roxana.wasnick@cellergon.de

Abstract: Idiopathic lung fibrosis (IPF) is a progressive and fatal degenerative lung disease of unknown etiology. Although in its final stages it implicates in a reactive manner all lung cell types, the initial damage involves the alveolar epithelial compartment, in particular the alveolar epithelial type 2 cells (AEC2s). AEC2s serve dual progenitor and surfactant secreting functions, both of which are deeply impacted in IPF. Thus, we hypothesize that the size of the surfactant processing compartment, as measured by Lysotracker incorporation, allows the identification of different epithelial states in the IPF lung. Flow cytometry analysis of epithelial Lysotracker incorporation delineates two populations (Lyso^{high} and Lyso^{low}) of AEC2s which behave in a compensatory manner during bleomycin injury and in the donor/IPF lung. Employing flow cytometry and transcriptomic analysis of cells isolated from donor and IPF lungs, we demonstrate that the Lyso^{high} population expresses all classical AEC2 markers and is drastically diminished in IPF. The Lyso^{low} population, which is increased in proportion in IPF, co-expresses AEC2s and basal cell markers resembling the phenotype of the previously identified intermediate AEC2 population in the IPF lung. In that regard, we provide an in-depth flow-cytometry characterization of Lysotracker uptake, HTII-280, proSP-C, mature SP-B, NGFR, KRT5 and CD24 expression in human lung epithelial cells. Combining functional analysis with extra- and intra- cellular marker expression and transcriptomic analysis, we advance the current understanding of epithelial cell behavior and fate in lung fibrosis.

 doi:10.20944/preprints202112.0033.v1

Keywords: IPF; alveolar epithelial cells; intermediate epithelial cells; transitional states, lysotracker; flow cytometry; lung transcriptomic profile, CK5; NGFR; CD24



1. Introduction

The human lung is a highly complex organ designed specifically for gas exchange. In idiopathic lung fibrosis (IPF), chronic epithelial injury leads to excessive deposition of rigid extra-cellular matrix, progressive decrease in lung compliance and gas-exchange surface, causing inevitable and fatal lung failure within 2-5 years after diagnosis[1–3]. Although new therapies increased significantly the duration and quality of life of IPF patients, a therapeutic regimen that would arrest or, even better, reverse disease progression remains to be discovered[4]. Partly responsible for this situation is our limited understanding of the cellular states and processes that each of the more than 40 cell types in the lung undergo, in an active (causative) or reactive manner in homeostatic and injury context[5,6]. A number of recent studies clearly identified the chronic injury of the alveolar type 2 epithelial cells (AEC2s) as the initial site of injury in the IPF lung[7–9]. AEC2s are facultative progenitors in the distal lung, which in differentiated state serve the vital function of surfactant production and secretion, but can also act as progenitors for other AEC2s and AEC1s in homeostatic and injury-repair situations[10,11].

Pulmonary surfactant is a phospholipoprotein mixture secreted exclusively by AEC2s which reduces the alveolar surface tension necessary for alveoli re-opening during the respiratory cycle. The protein component is represented by the surfactant proteins (SP) A, B, C and D, with two of them - SP-B and C holding tension-active properties. Following the processing from pro- (proSP-B and proSP-C) to mature (mSP-B and mSP-C) forms, they are secreted in their mature forms

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

specifically by AEC2s[12–14]. The processing and assembly of pulmonary surfactant takes place in the lamellar bodies of AEC2s, specialized organelles characterized by very low pH[15].

In IPF, repeated alveolar injury results in the recruitment of AEC2 progenitors necessary for the repair process[11,16–18]. In this process, the differentiated function of bona-fide AEC2s, defined as AEC2s which synthetize, process and secrete alveolar surfactant, is impaired and leads to increase alveolar surface tension, increased alveolar collapse which propagates the injury even further, thus creating a self-propagating cycle of injury and repair[19–22]. The acute AEC1 / 2 injury creates a micro-environment where other reactive cell types like alveolar macrophages and fibroblasts are quickly activated and recruited to cover the basement membrane and prevent fluid leakage into the air-space[23]. However, as the AEC2 progenitor pool is exhausted by injury or by extensive proliferation, the long-term repair after or during the chronic and repeated injury relies on the recruitment of other local epithelial progenitors, several of which have already been identified in the mouse lung[17,24–26]. In humans, the profound histological changes found in the distal IPF lung are consistent with the expansion of a cytokeratin 5 (CK) progenitor, but its origin and differentiating trajectory remains to be determined[27]. Recent landmark papers described the transcriptomic signatures of disease-free (donor) and IPF epithelial cells at single cell level, leading to the identification of transcriptomic signatures for many known epithelial cell types in the lung, and the identification of novel ones (ionocytes and CK17+/CK5-, aberrant AEC2s)[16,28–32]. It is unclear though how these transcriptomic signatures translate into stable or transitional cellular states and processes responsible for the disease phenotype[33]. The repair process in IPF is ultimately ineffective, underlying the disease progression that leads to organ failure. Thus, the ability to correlate scNGS data with protein expression and functional behavior would greatly increase our understanding of these epithelial fates and states, and turn this into a therapeutically actionable process for the benefit of IPF patients.

In our work, we delineate two functional alveolar epithelial states present in donor and IPF lung, defined by the size of their surfactant processing compartments, as measured by Lysotracker incorporation. Based on known intra-cellular and cell surface proteins together coupled with transcriptomic analysis, we show that the Lyso^{high} population consists of bona-fide AEC2s and is drastically diminished in IPF. A second population of Lyso^{low} cells, which uniformly expresses and processes surfactant proteins, but bears the transcriptional foot-print of a CK5-derived (basal) population is increased in IPF.

2. Materials and Methods

2.1 Animal studies

Animal studies were performed in accordance with the Helsinki convention for the use and care of animals and were approved by the local authorities at Regierungspräsidium Giessen V54-19 c 2015 (1) GI 20/10 Nr. 109/2011 (Bleomycin) or V54-19 c 20 15 h 02 GI 20/10 Nr. A53/2012 (untreated controls).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

2.2 Patient material

The study protocol was approved by the Ethics Committee of the Justus-Liebig-University School of Medicine (No. 31/93, 29/01, and No. 111/08: European IPF Registry), and informed consent was obtained in written form from each subject. Explanted lungs (n=31 for sporadic IPF, IPF_{LTX}; n=6 for COPD) or non-utilized donor lungs or lobes fulfilling transplantation criteria (n=27; human donors) were obtained from the Dept. of Thoracic Surgery in Giessen, Germany and Vienna, Austria and provided by the UGMLC Giessen Biobank, member of the DZL platform Biobanking. All IPF diagnoses were made according to the American Thoracic Society (ATS)/European Respiratory Society (ERS) consensus criteria[34], and a usual interstitial pneumonia (UIP) pattern was proven in all IPF patients.

2.3 Bleomycin Model of Lung Fibrosis

C57BL/6N mice (Charles River Laboratories, Sulzfeld, Germany) mice aged between 10 and 16 weeks old were used. Mice were intubated and bleomycin (Hexal, 2.5U/Kg body weight in 0.9% saline) was aerosolized using a Microsprayer (Penncentury). At each time point saline-treated and/or untreated mice were used as controls. Mice were weighed every day and sacrificed three, seven, 14, 21 and 28 days later for cell dissociation and flow cytometry analysis.

2.4 Lung tissue dissociation

For both mouse and human lung, standard dispase-based dissociation protocols were used as previously described[22,35,36] and detailed in the Supplemental material.

2.5 Flow cytometry analysis and cell sorting

Standard[37], previously published methods[22,35,36] were used for sample preparation, intra- and extracellular staining in preparation for flow cytometry and fluorescence activated cell sorting. Detailed methods and reagents including all antibodies are described in supplemental material. Single color controls were used to compensate for spectral overlap. Fluorescence minus one (FMO) controls were used whenever possible for positive / negative population gating. In the case of indirect intra-cellular staining, no primary control samples, consisting of the FMO control in that particular channel to which only the secondary antibody was added, were used for data interpretation and quantification. Data were acquired on a BD FACSCanto II (BD Biosciences) using BD FACSDiva software (BD Biosciences). Data were further analyzed using FlowJo vX software (FlowJo, LLC).

2.6 Immunofluorescence analysis

The staining procedures were based on standard, previously published techniques and the reagents are listed in Supplemental material and Supplemental Tables 1 and 2. However, given that both the mature SP-B and proSP-B antibodies were raised in the same species (rabbit), the standard protocol was modified as follows. Following standard deparaffinization and blocking (see Supplemental material) , slides were incubated with mature SP-B antibody at a very low concentration (1:2000, 10 times lower than for traditional mature SP-B staining) and the fluorescent signal was amplified using the Alexa Fluor™ 555 Tyramide SuperBoost™ Kit, goat anti-rabbit IgG (Thermo Scientific). This resulted in the covalent attachment of Alexa Fluor 555 Tyramide at the base of the antigen, which allowed consequent stripping of the rabbit anti-mature SP-B antibody using the standard Citrate-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

based antigen retrieval solution as described in Supplemental material. Samples were re-blocked with 5% BSA in PBS solution and incubated with the rabbit anti-proSP-B antibody followed by an Alexa Fluor 488 labeled donkey anti-rabbit secondary. Apropriate controls demostrating the lack of cross-reactivity were used to ascertain the specificity of the two signals (see Supplemental material). Stainings were imaged on a wide-field fluorescence microscope (Axio Observer.Z1 fluorescence microscope, Carl Zeiss MicroImaging, Germany) and a confocal microscope (TCS SP5, Leica Microsystems) and the images were processed and quantified using the Fiji package of ImageJ image analysis software (www.imagej.net).

2.7 Microarray analysis

Purified total RNA was amplified using the Ovation PicoSL WTA System V2 kit (NuGEN Technologies, Bemmel, Netherlands). Per sample, 2µg amplified cDNA was Cy3-labeled using the SureTag DNA labeling kit (Agilent, Waldbronn, Germany). Hybridization to 8x60K 60mer oligonucleotide spotted microarray slides (Human Whole Genome, SurePrint G3 Human GE v3 8x60K Microarray; Agilent Technologies, design ID 072363) and subsequent washing and drying of the slides was performed following the Agilent hybridization protocol in Agilent hybridization chambers, with following modifications: 2 µg of the labeled cDNA were hybridized for 22h at 65°C. The cDNA was not fragmented before hybridization. The dried slides were scanned at 2 µm/pixel resolution using the InnoScan is900 (Innopsys, Carbonne, France). Image analysis was performed with Mapix 8.2.5 software, and calculated values for all spots were saved as GenePix results files. Stored data were evaluated using the R software (www.r-project.org) and the limma package[38] from BioConductor[39]. Log_2 mean spot signals were taken for further analysis. Data was background corrected using the NormExp procedure on the negative control spots and quantile-normalized[38,40] before averaging. Log_2 signals of replicate spots were averaged, and from several different probes addressing the same gene only the probe with the highest average signal was used. Genes were ranked for differential expression using a moderated t-statistic[38]. Pathway analyses were done using gene set tests on the ranks of the t-values[38,41] . Pathways were taken from the KEGG database (http://www.genome.jp/kegg/pathway.html).

Heatmaps are generated from the normalized log_2 spot intensities (I) and show the gene-wise z-values (where $z_j = (I_j - mean(I)) / SD(I)$ for $j = 1 \ldots n$).

2.8 Data analysis

2.8.1 Flow cytometry data analysis

Frequency of parent and mean fluorescence intensity (MFI) data was exported from FlowJo v.10 and analyzed using MicrosoftExcel, R software (www.r-project.org) or GraphPad Prism (GraphPad Software). For the bleomycin experiments, the percentage of $Lyso^{high}$ and $Lyso^{low}$ cell populations in control vs treated samples were first Log-transformed and the Student's T-test was used to determine the statistical significance of their differences at each time point. To evaluate the dynamic of the $Lyso^{high}$ and $Lyso^{low}$ populations over time (Figure 1C), we related the proportion of each $Lyso^{high}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

or Lyso^{low} population in each group (bleomycin or saline) to the total Lysotracker incorporating population in the control animals at each time-point as follows: each point on the line represents the % change Log10 of the respective Lyso^{high} or Lyso^{low} at DayX calculated as Log10(Lyso$^{high\ or\ low}$) DayX$_{bleo\ or\ control}$/Log10(Lyso^{high}DayX+Lyso^{low}DayX)$_{control}$. To determine the relative difference between the Lyso^{high} and Lyso^{low} populations in each epitheilial group at each time point (Figure 1D), we analyzed the log odds ration of these two populations using the R statistical analysis software. A log odds ratio of 0 means that the two populations are similar and therefore the probability that an epithelial cell is a Lyso^{high} cell is equal to that of it being a Lyso^{low}. Positive log odds ratios indicate that the two populations are different and therefore the probability of a cell to belong to one population is larger than the probability of it belonging to the other one. Flow cytometry data collected from human samples were analyzed in a similar manner. Student's t-test or two- way ANOVA were used as appropriate (stated in the figure legend) to test the null hypotheses that the Log transformed MFI of percentage of parent values were different in each comparison.

2.8.2 Immunoflurescence quantification and analysis

Fluorescence intensity was analyzed using the Fiji/Image J (https://imagej.net) image analysis software and fluorescence intensity data was exported and statistically analyzed and ploted using MicrosoftExcel.

3. Results

3.1. Lysotracker incorporation delineates two populations of epithelial cells in bleomycin induced injury

To determine the dynamic behavior of AEC2s during bleomycin injury, C57BL6 mice were treated with bleomycin (2.5U/kg) and analyzed 3, 7, 14, 21 and 28 days post-administration. Saline or untreated mice (generally termed controls) were used as controls at each time point. Mice were sacrificed and their lungs dissociated into single cell suspension, whose cellular composition was analyzed by flow cytometry at each time-point. To identify AEC2s, dead cells (PI$^+$ or DAPI$^+$) and AEC1 cells (podoplanin - PDPN$^+$), were first excluded, and the epithelial compartment was further identified by EpCAM expression within the CD45$^-$ (non-hematopietic) and CD31$^-$ (non-endothelial) population (full gating path in Supplemental Figure 1A). The proportion of DAPI$^-$ CD45$^-$ CD31$^-$ PDPN$^-$ EpCAM$^+$ population (the epithelial cell compartment from here on) was slightly decreased in bleomycin-treated mice starting day 14 and reached statistical significance at day 28 (Supplemental Figure 1B). To identify AEC2s within the epithelial compartment, we took advantage of their specific ability to uptake Lysotracker dyes[42,43]. The dynamic of Lysotracker uptake was analyzed during the bleomycin recovery time-course, revealing three distinct populations: a Lysotracker^{neg} (Lyso^{neg}), a Lysotracker^{low} (Lyso^{low}) and a Lysotracker^{high} (Lyso^{high}) population (Figure 1A). At all time-points analyzed, the Lyso^{high} population was decreased in number, with the greatest decrease registered at days 7 and 14, when AEC2s injury is maximal[44,45]. This was paralleled by a proportional increase in the Lyso^{low} population, that reached a maximum increase at the same time points (Figure 1B). The time-course analysis of population dynamic, showed that Lyso^{high} and Lyso^{low} populations

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

behaved complementary to each other, with a maximum relative change at day 7 and partial recovery by day 21 and 28. The paired analysis of the log odds ratio of the Lyso^{high} vs Lyso^{low} population (Log Lyso^{high}/Lyso^{low}) further supported our conclusion that the difference between the two populations is maximal in control samples and early time points, but decreases significantly at day 7 and 14 (log value of zero) (Figure 1D). Of note, lysotracker uptake was not completely recovered at day 28, suggesting long-lasting alterations in cellular phenotype (Figure 1B-D).

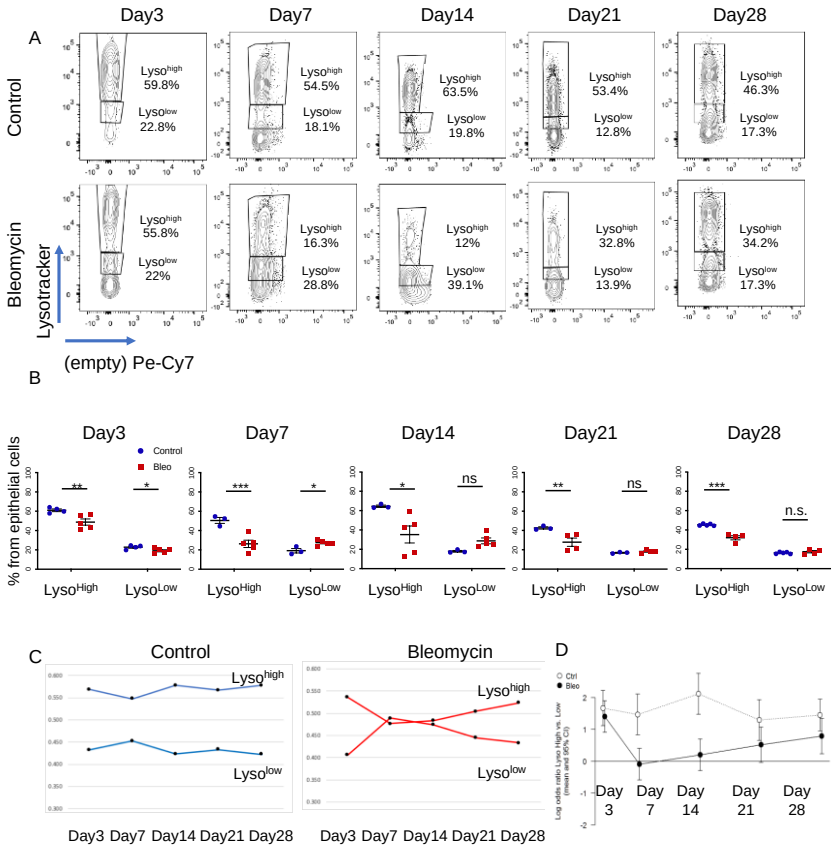


Figure 1. Characterization of AEC2s in bleomycin model of lung fibrosis

(A)-(D) Bleomycin or saline was intratracheally instilled into the lung of C57B6 mice which were analyzed after 3 (control n=4, bleomycin n=5), 7 (control n=3, bleomycin n=5), 14 (control n=3, bleomycin n=5), 21 (control n=3, bleomycin n=4), and 28 (control n=4, bleomycin n=4) days. Shown is the flow cytometry analysis of the AEC2s in the epithelial compartment defined as DAPI$^-$ CD45$^-$ CD31$^-$ PDP$^-$ EpCAM$^+$. (A) Representative panels of Lysotracker uptake (Lyso^{high} and Lyso^{low}) as a percentage of the parent epithelial compartment of bleomycin treated mice.(B) Statistical analysis of the Lyso^{high} and Lyso^{low} populations at each time-point. Data are presented as the mean ± SEM. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, n.s. = not significant by ANOVA. (C) Time-course analysis of the Lyso^{high} and Lyso^{low} populations in control mice (left panel) and bleomycin treated mice (right panel). (D) Analysis of the log odds ratio of the Lyso^{high} vs Lyso^{low} population (Log Lyso^{high}/Lyso^{low}) during the time-course of bleomycin recovery. Data are presented as the mean and 95%confidence interval.

3.2 Lysotracker uptake in human lung epithelium

Next, we asked if the behavior of these populations is similar in the distal human donor and IPF lung. To that end, subpleural tissue from six donor and six end-stage IPF explanted lungs were dissociated into single cell suspension and analyzed by flow cytometry. The gating strategy was

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

similar to that of the mouse lung, where the epithelium was identified as live, $CD45^-$ $CD31^-$ $EpCAM^+$ cells (Supplemental Figure 2A). Fluorescence-minus-one (FMO) samples were used for appropriate gating (Supplemental Figure 2B). There was no statistically different proportion of epithelial cells between the groups (Supplemental Figure 2D). Similar to the mouse data, the proportion of IPF $Lyso^{high}$ cells was dramatically decreased compared to donors, from an average of 50.3% in donors to 10.1% in IPF. The $Lyso^{low}$ population behaved in a complementary fashion, increasing from an average of 15.2% in donors to 37.1% in IPF patients (Figure 2A, B). Individual panels from each patient are shown in Supplemental Figure 2C.

To further understand the identity of the $Lyso^{high}$ and $Lyso^{low}$ incorporating cells, we analyzed the expression of HTII-280, a well-known AEC2 marker, as a function of Lysotracker incorporation [46]. In addition to donor (n=6) and IPF (n=7) samples, COPD samples (n=7) were added as non-IPF related controls. Quadrant gating of Lysotracker versus HTII-280 expression in the epithelial cell population gated as in Figure 2A led to the identification of four populations: Q1 ($Lyso^{pos}$/HTII-280^{neg}), Q2 ($Lyso^{pos}$/HTII-280^{pos}), Q3 ($Lyso^{neg}$/HTII-280^{pos}) and Q4 ($Lyso^{neg}$/HTII-280^{neg}) (Figure 2C). In donor and COPD samples, the largest proportion of epithelial cells comprised of bona-fide AEC2s (Q2: 79.64% Donor and 67.39% COPD), which were $Lyso^{pos}$/HTII-280^{pos}. In contrast, in the IPF samples, the proportion of Q2 cells was markedly reduced to an average of 14.33%, consistent with the well establish chronic injury of AEC2s characteristic of IPF. This decrease in Q2 was paralleled by a marked increase in Q1, which represent the $Lyso^{pos}$/HTII-280^{neg}, from 5.12% in donor samples and 9.43% in COPD samples, to 47.73% in IPF (Figure 2D and E). Populations Q3 and Q4 were not significantly altered in all comparisons, with the exception of a slight but statistically increase in Q4 in the comparison COPD/IFP (Supplemental Figure 2E). The analysis was very consistent from patient to patient, with some variability noted in the Q4 ($Lyso^{neg}$/HTII-280^{pos}) population as shown in Supplemental Figure 2F.

At a first glance, the gating strategy suggests that the Q2 population consists mostly of $Lyso^{high}$ cells, while the $Lyso^{low}$ cells belong to Q1. Thus, we compared the mean fluorescence intensity (MFI) of the Lysotracker incorporating populations Q1 and Q2, which showed a constant and statistically significant increase in Lysotracker incorporation in Q2 compared to Q1, demonstrating that Q2 comprises of mostly $Lyso^{high}$ cells and Q1 of mostly $Lyso^{low}$. This difference was maintained in all three groups regardless of their disease status (Figure 2F), suggesting that these two parameters define two distinct cellular states and, in this regard, functionally homogenous populations (Figure 2G and H). Taken together our data suggests the existence of two distinct epithelial populations with distinct levels of Lysotracker uptake characteristics, which vary in inversely correlated manner, suggestive of compensatory behavior in IPF patients compared to donors. Moreover, the $Lyso^{high}$ population is marked by the well-established HTII-280 antibody, confirming its bona-fide AEC2 identity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

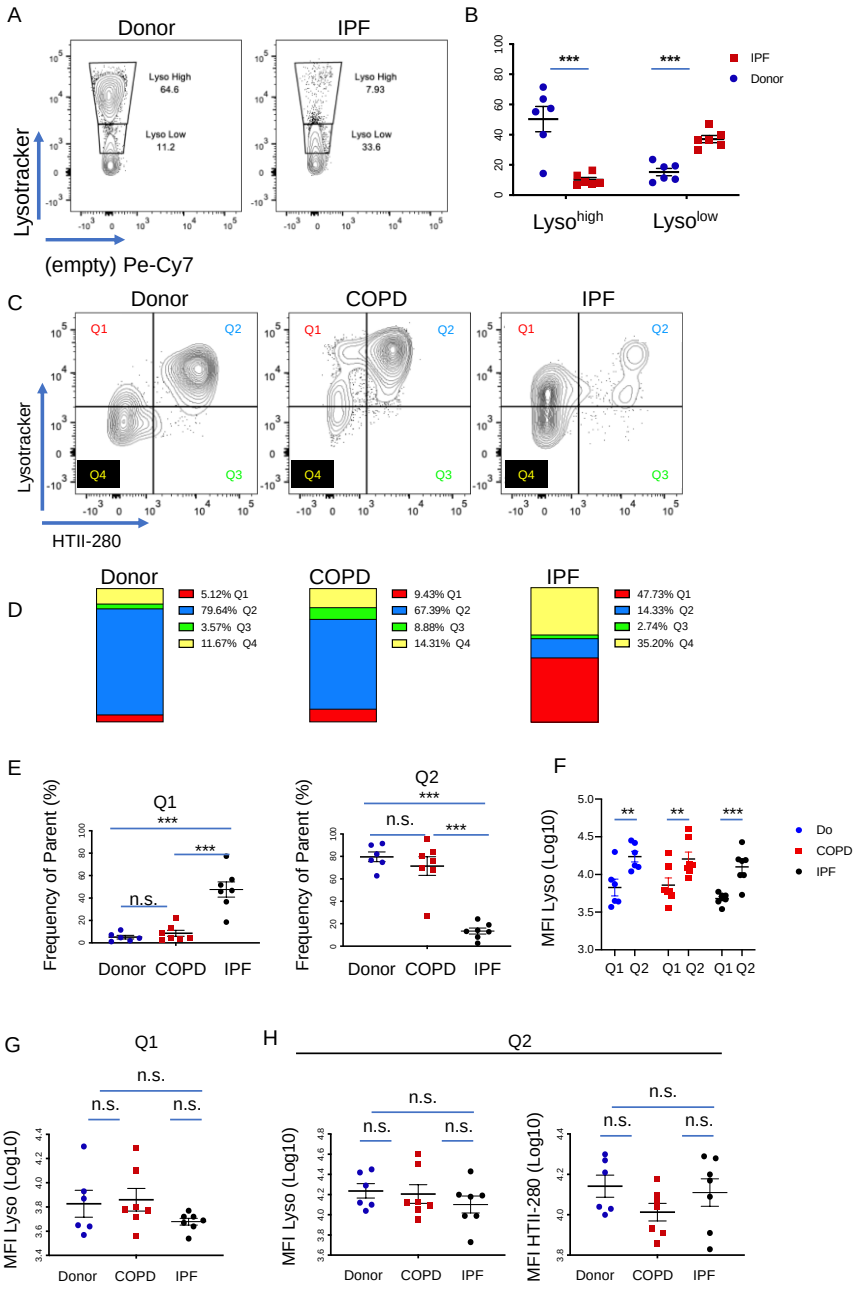


Figure 2. Lysotracker uptake in the epithelial compartment of the human lung. (A) Representative panels of flow cytometry analysis of the Lysotracker uptake ($Lyso^{high}$ and $Lyso^{low}$) in the epithelial compartment ($DAPI^{-}$ $CD45^{-}$ $CD31^{-}$ $EpCAM^{+}$) of human donor (n=6) and IPF (n=6) lungs. (B) Quantification of the $Lyso^{high}$ and $Lyso^{low}$ populations in donor and IPF samples in (A). (C) Representative panels of Lysotracker uptake (y-axis) as a function of HTII-280 reactivity (x-axis) in the epithelial compartment of donor (n=6), COPD (n=7) and IPF (n=7) lungs. Quadrant gating identifies four different populations as follows: Q1 ($Lyso^{pos}$/HTII-280^{neg}), Q2 ($Lyso^{pos}$/HTII-280^{pos}), Q3 ($Lyso^{neg}$/HTII-280^{pos}) and Q4 ($Lyso^{neg}$/HTII-280^{neg}). (D) Quantification of the data shown in (C) showing the relative contribution of the Q1 to Q4 populations to the epithelial compartment of donor, COPD und IPF lungs. (E) Quantification of the Q1 (left diagram) and Q2 (right diagram) as the frequency of parent ($DAPI^{-}$ $CD45^{-}$ $CD31^{-}$ $EpCAM^{+}$) population. (F) Comparison of the Lysotracker uptake in the Lysotracker positive populations Q1 and Q2 in donor, COPD and IPF patients measured by the MFI of the respective populations. (G) Comparison of the Lysotracker uptake in the Q1 population in donor, COPD and IPF patients measured by the MFI of the respective populations. (H) Comparison of the Lysotracker uptake (left panel) and HTII280 reactivity (right panel) in the Q2 population in donor, COPD and IPF patients measured by the MFI of the respective populations. Data are presented as the mean ± SEM of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

percentage cells from the parent population. Statistical analysis was performed on Log(10) values. *p < 0.05, **p < 0.01, ***p < 0.001, n.s. = not significant by ANOVA

3.3 Surfactant protein expression defines two populations of AEC2s in donor and IPF Lung

Surfactant protein production, processing and secretion is the most defining characteristic of AEC2s. Thus, we asked what is the pattern of proSPC expression in relation to HTII-280. To that end, following the usual cell surface staining (CD45, CD31, EpCAM, HTII-280), the same six donor and six IPF single cell preparations used in the previous analysis were fixed, permeabilized and stained intracellularly with a proSP-C specific antibody. Because the lysotracker signal is lost during the fixation process, we relied on HTII-280 reactivity for the identification of the bona-fide AEC2s (DAPI^{neg} CD45^{neg} CD31$_{neg}$ EpCAM^{pos} HTII-280^{pos}). Analysis of HTII-280 vs proSPC expression in these samples resulted in four populations: Q1 (proSP-C^{pos}/HTII-280^{neg}), Q2 (proSP-C^{pos}/HTII-280^{pos}), Q3 (proSP-C^{neg}/HTII-280^{pos}) and Q4 (proSP-C^{neg}/HTII-280^{neg}) (Figure 3A and Supplemental Figure 3A and B). Similar to the Lyso/HTII-280 analysis (Figure 2), donor bona-fide AEC2s (Q2, proSP-C^{pos}/HTII-280^{pos}) were the highest represented population (Q2=66.7%) and their proportion was markedly decreased in IPF samples to 6.75% (Figure 3B). Also, a population that was HTII-280^{neg}, but expressed lower levels of proSP-C than Q2 was present in both patient groups, and it was markedly increased in IPF (donor Q1=17.31% vs IPF Q1=33.84% - Figure 3B, C). Confirming previously known data, analysis of the amount of proSP-C expressed, as measured by the proSP-C MFI of each population showed that in IPF the bona-fide AEC2s (Q2) expressed significantly less proSP-C compared to donors. However, the Q1 (proSP-C^{pos}/HTII-280^{neg}) population which was increased proportionally in IPF patients, did not differ in the amount of proSP-C expressed (Figure 3B-D). This suggests that while in IPF the number and SP producing function of AEC2s is decreased, the potentially compensatory proSP-C^{low} HTII-280^{neg} (Q1) population, expresses lower levels of proSP-C. Additionally, there was a statistically significant increase in the proSP-C^{neg} HTII-280^{neg} (Q4) population which suggested the increased presence of non-AEC2 cells in the distal IPF lung (Figure 3D). The Q3 population, representing proSP-C^{neg} HTII-280^{pos} was negligible and did not vary significantly with the disease state (Supplemental Figure 3C).

Although characteristic for the alveolar epithelium, expression of proSP-C and proSP-B has been previously noted in the non-alveolar compartment of the human lung. However, only AEC2s have the unique ability to process and secrete the mature forms (mSP-C and mSP-B). Thus, we analyzed the expression of mature SP-B (mSP-B) by intracellular staining of the same donor (n=6) and IPF (n=6) samples as in the previous analyses, in conjunction with the usual cell surface markers. Similar to the proSP-C data, mSP-B was expressed in the majority of the bona-fide AEC2s (Q2: HTII-280^{pos} mSP-B^{pos}) in both donor and IPF, and their proportion was drastically reduced in IPF (Figure 3E and F) . However, the IPF Q1 (HTII-280^{neg} mSP-B^{pos}) population expressed higher levels of mSP-B than donor Q1, suggesting an up-regulation of the surfactant processing ability in this population in disease conditions. Of note, the expression level of mSP-B in Q1 of IPF remained below that of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

Q2 (HTII-280^{pos} mSP-B^{pos}), suggesting a distinct functional state of this population (Figure 3E-H). Similar to the proSP-C data, there was no significant change in the proportion of Q3 (HTII-280^{pos} mSP-B^{neg}) population but a significant increase in the Q4 (HTII-280^{neg} mSP-B^{neg}) cells.

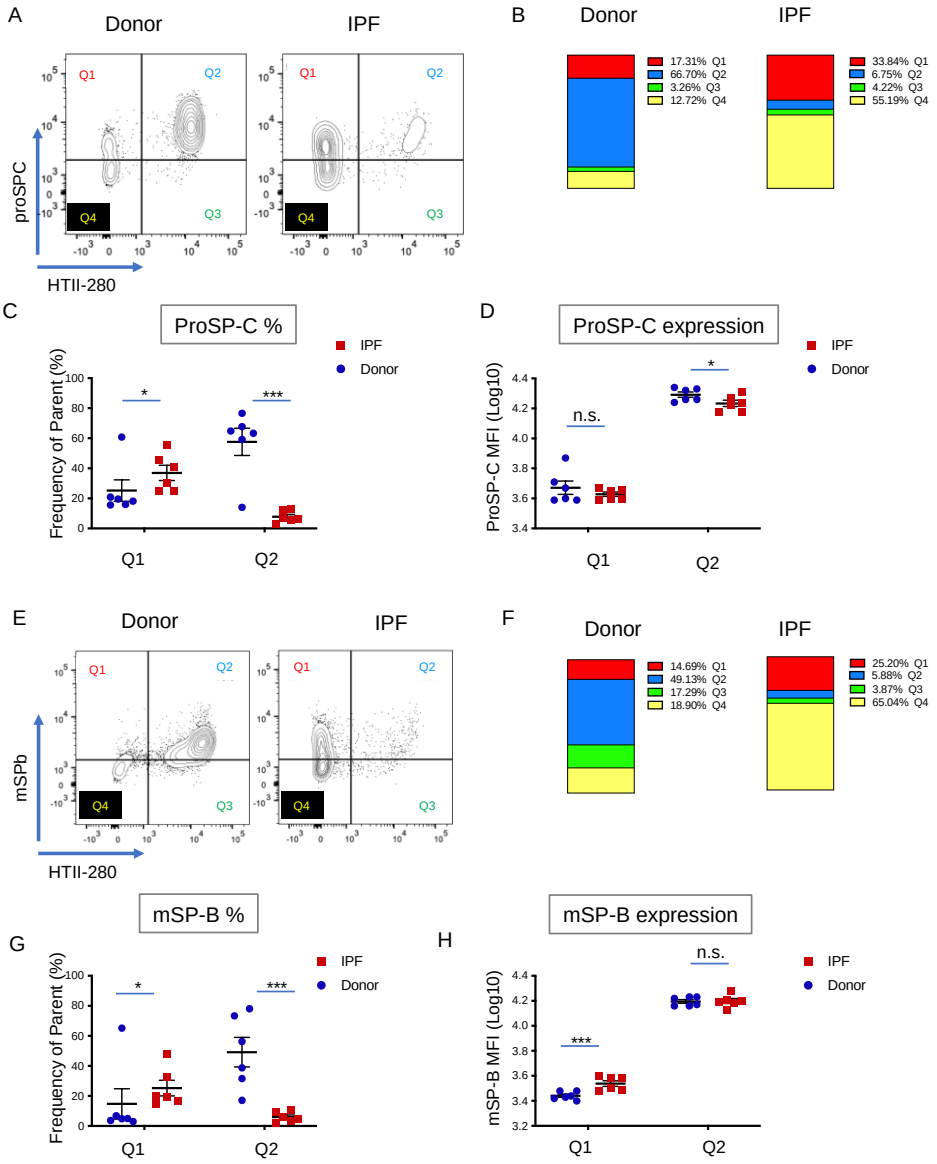


Figure 3. Surfactant protein expression in the epithelial compartment of donor and IPF lung. (A) Representative flow cytometry panels of proSP-C and HTII-280 expression in the epithelial compartment (DAPI$^-$ CD45$^-$ CD31$^-$ EpCAM$^+$) of donor (n=6) and IPF (n=6) lung preparations. (B) Average contribution of the Q1-Q4 populations to the epithelial compartment of the samples shown in (A), showing the change in epithelial composition in IPF lung compared to donors. Left column: donors; right column: IPF. (C) Quantification of the population frequency of Q1 and Q2 in donor (blue dot) and IPF (red square) lung samples shown in (A). (D) Quantification of the MFI as a measure of proSP-C expression level (Log10 MFI) in the Q1 and Q2 populations of the samples shown in (A). (E) Representative flow cytometry panels of mSP-B and HTII-280 expression in the epithelial compartment (DAPI$^-$ CD45$^-$ CD31$^-$ EpCAM$^+$) of donor (n=6) and IPF (n=6) lung preparations. (F) Average contribution of the Q1-Q4 populations to the epithelial compartment of the samples shown in (E). Left column: donors; right column: IPF. (G) Quantification of the population frequency of Q1 and Q2 in donor and IPF lung samples in (E). (H) Quantification of the MFI as a measure of mSP-B expression level in the Q2 population of the samples in (E) Data are presented as the mean ± SEM of the Log10 (MFI) values. *$p < 0.05$, ***$p < 0.001$, ns = not significant by Student t-test.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

Throughout our analysis we noticed very consistent similarities among the Lysotracker, proSP-C and mSP-B expression pattern in relation to HTIII-280: the Q1 and Q2 populations behaved similarly in all samples in each analysis. The co-staining of Lysotracker with intra-cellular markers is technically not feasible because of the loss of Lysotracker fluorescence during the fixation/permeabilization process necessary for intra-cellular staining. However, comparative and concomitant analysis of the Q1-Q4 profile with the three markers in the same donor (n=6) and IPF (n=6) patient samples showed that the proportion of cells belonging to Q1-4 in each population was very similar in the three parallel analyses (Figure 4A). This suggested, in a correlative manner that the Q1 population represents a $Lyso^{low}$, $proSP\text{-}C^{low}$, $mSP\text{-}B^{low}$ population of AEC2-like cells, while the Q2 population represents the $Lyso^{high}$, $proSP\text{-}C^{high}$, $mSP\text{-}C^{high}$ population of bona-fide AEC2s. To confirm the existence of mSP-B expressing cells outside of the lysotracker incorporating compartment, donor and IPF peripheral lung tissue sections were co-stained for mSP-B and ABCA3, a protein specifically expressed in the lamellar bodies of mature AEC2s. Indeed, in the donor lung mSP-B was present in most all ABCA3 expressing AEC2s, while in IPF extensive epithelial areas (identified morphologically) were characterized by mSP-B expression in the absence of ABCA3 (Figure 4B). The expression of Lysotracker and proSP-C was very consistent within each patient group (Do vs IPF and Supplemental Figure 4). However, the intensity of the mSP-B staining was highly variable within the same patient group, suggesting that the existence of variable mSP-B processing capacity (Figure 4A and Supplemental Figure 4A). To confirm that the variability processing ability of AEC2 cells was not an artefact of the cell isolation procedure, donor (n=4) and IPF (n=4) paraffin embedded tissue sections were co-stained for pro- and mSP-B, and the fluorescence intensity of each was quantified. Linear regression analysis showed that the processing ability of each sample, represented by the regression's slope was variable within each group, but an overall flattening of the slope was noted between IPF and donor samples. Moreover, the two values yielded were positively correlated in most donor samples (positive R^2 values), but this correlation was lost in three out of the four IPF samples (R^2=0, Fig 4C and D). Analzying the spatial distribution of the two signals, we noticed that in donors they were tightly co-expressed, but in IPF there was a heterogenous distribution of areas of co-localization (particularly in non-affected areas) and areas where the mSP-B was low or absent in cells that expressed the pro- forms (Figure 4D). Together, these data show that albeit variable in donor samples, mSP-B processing ability is decreased in IPF.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

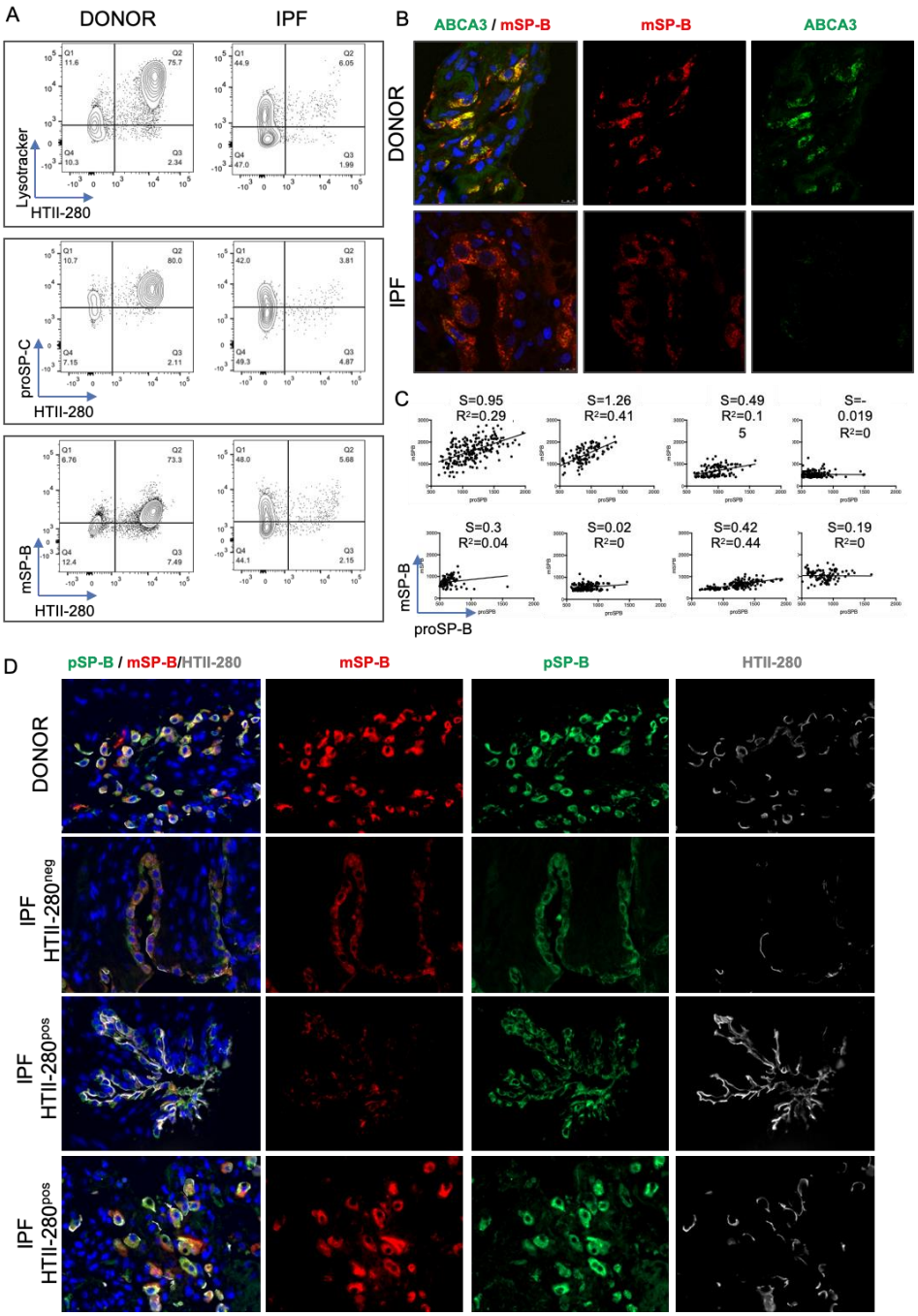


Figure 4. Comparative expression of Lysotracker, proSP-C and mSP-B expression in donor and IPF lung. (A) Six donor and six IPF lung preparation were co-stained in parallel with HTII-280, Lysotracker, proSP-C and mSP-B. Representative flow cytometry panels of Lysotracker (upper), proSP-C (middle) and mSP-B (lower) vs HTII-280 expression in the epithelial compartment of one donor (left column) and one IPF (right column) lung. (B) Representative immunofluorescence images of mature (red) and ABCA3 (green) in donor and IPF paraffin embedded lung tissues. (C) Quantification of the fluorescence intensity of mature SP-B and proSP-B immunofluorescence signal in 4 donor and 4 IPF patients showing the slope (s) of the linear regression and the correlation index (R^2) for each patient. (B) Representative immunofluorescence images of mature (red) and proSP-B (green) in donor and IPF tissue shown in (C).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

3.5 Transcriptional characterization of the IPF Lyso^{low} population

Given the recent single cell NGS data that identified the existence of transitional AEC2 states with distinct transcriptomic signatures in normal and IPFlungs[28–31,47], we asked if the Lyso^{low} population in IPF, which has an intermediate expression profile in terms of Lysotracker and surfactant protein expression, resembled any of the previously mentioned intermediate populations. Thus, we used microarray analysis to compare the transcriptomic profile of eigth FACS sorted donor Lyso^{pos} AEC2s, composed in majority of Lyso^{high} cells (see Figure 2A) , and Lyso^{pos} cells from six IPF lungs, consisting in majority of Lyso^{low} cells (Figure 2A and 5A). Principal component analysis of the data showed the lack of variance between the two groups (Supplemental Figure 5A) suggesting great similarities between the two populations. However, in this analysis 612 genes were up-regulated (LFC > 2), and 1382 genes were down-regulated (LFC < -2) in IPF Lyso^{pos} compared to donor Lyso^{pos} AEC2s. Interestingly, the first 50 upregulated genes in the Lyso^{pos} population of IPF patients included several genes known to be upregulated in IPF, while several surfactant related genes were noted in the 50 most downregulated genes (Figure 5B and Supplemental Figure 5B). Validating our data, KEGG analysis identified metabolic pathways and pathways related to protein synthesis/processing and oxidative phosphorylation as being the most significantly down-regulated pathways in IPF (Figure 5C). In order to determine the phenotype of the Lyso^{pos} IPF population, we superimposed the transcriptomic signatures of several relevant cell types from two recent publications onto our differentially expressed gene expression data[29,31]. First, we defined the signatures of all relevant cell types in each data set using the first 30 most differentially expressed genes for each cell-type: AEC2, signaling AEC2, basal, differentiating and proliferating basal, AEC1, ciliated and club cells (Travaglini et al), and AEC2, AEC1, basal, aberrant basal, ciliated and club (Adams et al). These signatures were then superimposed onto our donor Donor / IPF Lysotracker comparison, showing an overall down-regulation of the AEC2 signature in the IPF Lyso^{pos} population. However, a closer look at the surfactant compartment genes revealed the significant down-regulation of several surfactant synthesis and processing genes (*NAPSA, ABCA3, LAMP3, LPCAT1*), while the surfactant proteins genes *SFTPB* and *SFTPC* were not significantly regulated (*SFTPC* LFC=-0,25, LOG(p) = 0.24 and *SFTPB* LFC=-0.30, LOG(p)=0.50). This suggested that the two populations which homogenously express proSP-B and proSP-C mRNA (Figure 5D) and protein (Figure 3), differ in the expression of the processing machinery that would normally commit them to a bona-fide AEC2 fate. In addition, two fundamental regulators of AEC2 fate had opposite patterns of expression: IPF Lyso^{pos} cells expressed markedly decreased levels of *ETV5* (LFC =-3.42, LOG(p)=2.32), but *SOX9* (LFC=3.25, LOG(p)=8.93) was one of the top overexpressed genes in our data set (Figure 5D and Supplemental Figure 5 C,D). Further analysis showed the upregulation of basal, differentiating basal and aberrant basal transcriptomic signatures in IPF, suggesting the presence of cells belonging to the basal cell lineage. Ciliated, club and AEC1 signatures did not unequivocally superimposed with any of the up or down-regulated transcriptomic profiles (Figure 5E and Supplemental Figure 5C,D). Taken together, this data demonstrates that the Lyso^{pos} population in IPF represents most likely a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

heterogenous population of basal-derived cells with the common property of surfactant proteins B and C expression, but lacking a mature surfactant processing compartment necessary to compensate the surfactant defects known to occur in IPF.

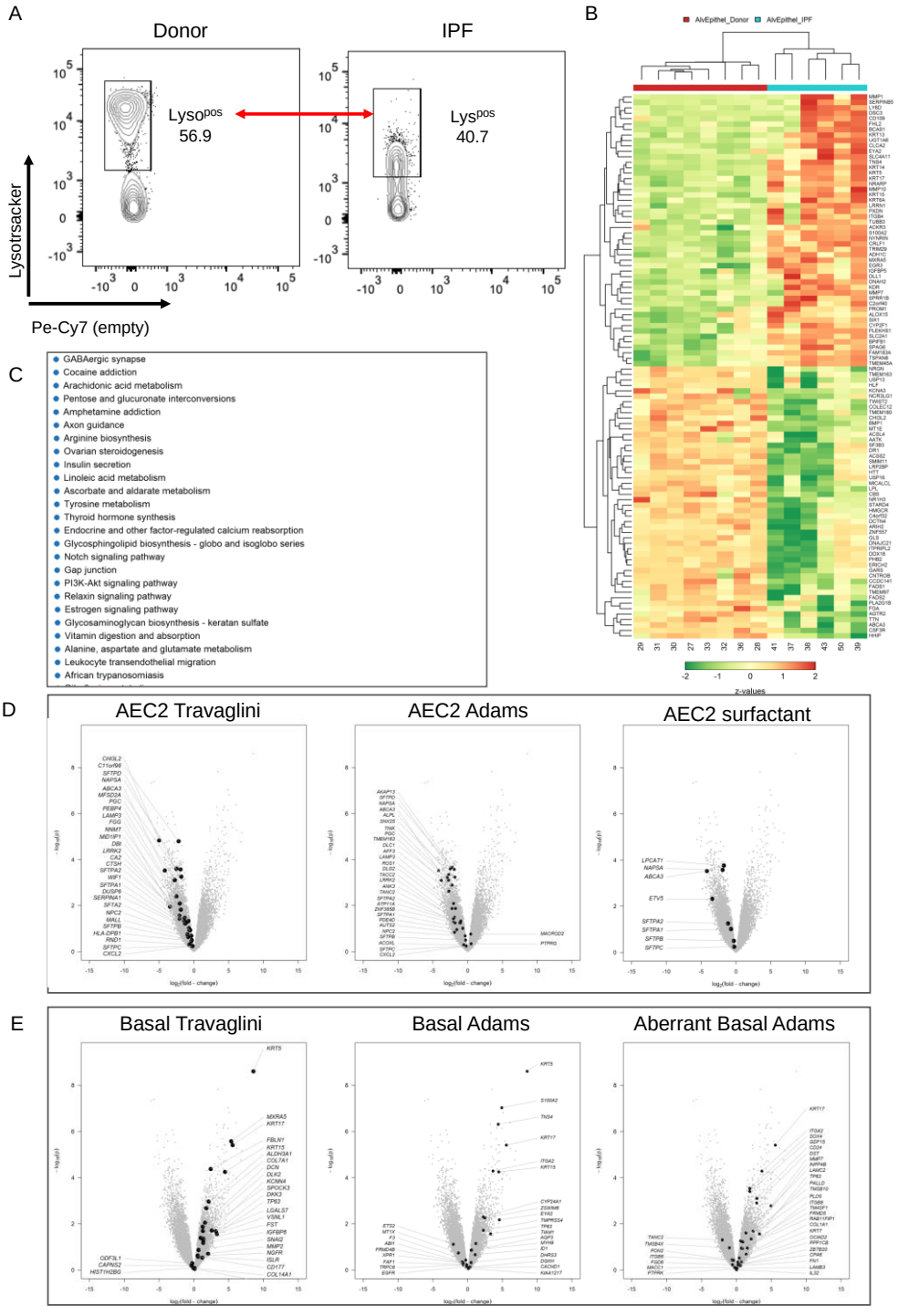


Figure 5. Transcriptomic profiling of the Lysopos population in IPF. Donor Lysopos (n=8) and IPF Lysopos (n=6) epithelial cells were isolated by flow cytometry and their transcriptomic profiles were determined by microarray analysis. (A) Flow cytometry panels showing the sorting strategy for the two populations. (B) Heat map of the first up and downregulated genes in IPF Lysopos vs donor Lysopos. (C) KEGG pathway analysis showing the first 20 most differential regulated pathways. (D) The transcriptomic signatures of AEC2 identified by Travaglini et al. and Adams et al. were superimposed on the vulcano plots depiction of the up and downregulated genes in IPF Lysopos compared to donor Lysopos (left and middle plots). Right plot shows the distribution of surfactant production and processing genes. (E) The transcriptomic signatures of various populations of donro and IPF basal cells identified by Travaglini et al. and Adams et al. were superimposed on the vulcano plots depiction of the up and downregulated genes in IPF Lysopos compared to donor Lysopos.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

3.6 Basal cell markers expression in donor and IPF lung

Our data, in consensus with the existing literature, suggested an increase in basal and aberrant basaloid cells in the distal lung epithelium of IPF patients, which have been in the past identified by their cytokeratin 5 (CK5, the protein product of *KRT5* mRNA) intracellular or NGFR cell surface expression[50]. First, we determined by flow cytometry the expression of the CK5 protein in the epithelial compartment of six donor and six IPF lungs. Of note, CK5 expression as a function of HTII-280 demonstrated that HTII-280^{pos} cells did not express CK5 in neither donor nor IPF lung, and the CK5 up-regulation was strictly limited to the HTII-280^{neg} compartment (Figure 6A and Supplemental Figure 6A). Indeed, the number of CK5^{pos} cells was greatly increased in the epithelial compartment of IPF lung (average Q3 = 37.86%) compared to donor (average Q3 = 12.15% Figure 6A, B and Supplemental Figure 6B). We also determined in a similar manner the cell surface expression of NGFR in six donor and six IPF lung cell preparations (Figure 6C, D and Supplemental Figure 6 C, D). Albeit the number of NGFR^{pos} cells also increased significantly in proportion in IPF compared to donor epithelial cells (average 20.1% IPF vs 3% donor), their proportion was much lower than that of CK5^{pos} cells in both groups (donor and IPF), suggesting the existence of a population of CK5^{pos} cells that do not express NGFR. Of note, in our transcriptomic analysis, NGFR showed levels of expression in both AEC2 populations below the threshold above which a gene was considered to be expressed (Figure 6B and Supplemental Figure 6B). We next asked if the expression of the two markers defines distinctly localized populations of epithelial cells in either donor and/or IPF lung. Thus we analyzed by immunofluorescence staining the pattern of expression of CK5 and NGFR together with the AEC2 marker ABCA3 in six donor and six IPF lung samples. Representative confocal images (Figure 6 E) show that in both donor and IPF, ABCA3 expressing AEC2s expressed neither CK5 nor NGFR, confirming the flow cytometry data in Figure 6A and C. Additionally, in donor lung, extensive areas of basal cells were labeled with either CK5 alone or co-expressed with NGFR (CK5^{pos} NGFR^{pos} cells) in a clonal fashion. In the IPF lung, CK5^{pos} NGFR^{pos} cells were found in either normal appearing basal cells in the conducting airways or in simple or pseudo-stratified epithelium lining epithelial cysts in distal fibrotic areas. CK5^{pos} NGFR^{neg} cells were present predominatly as highly metaplastic areas in the fibrotic distal lung (Figure 6D). Together these data demonstrate that NGFR expression differentiate two populations of CK5^{pos} basal cells with different behavior in IPF.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

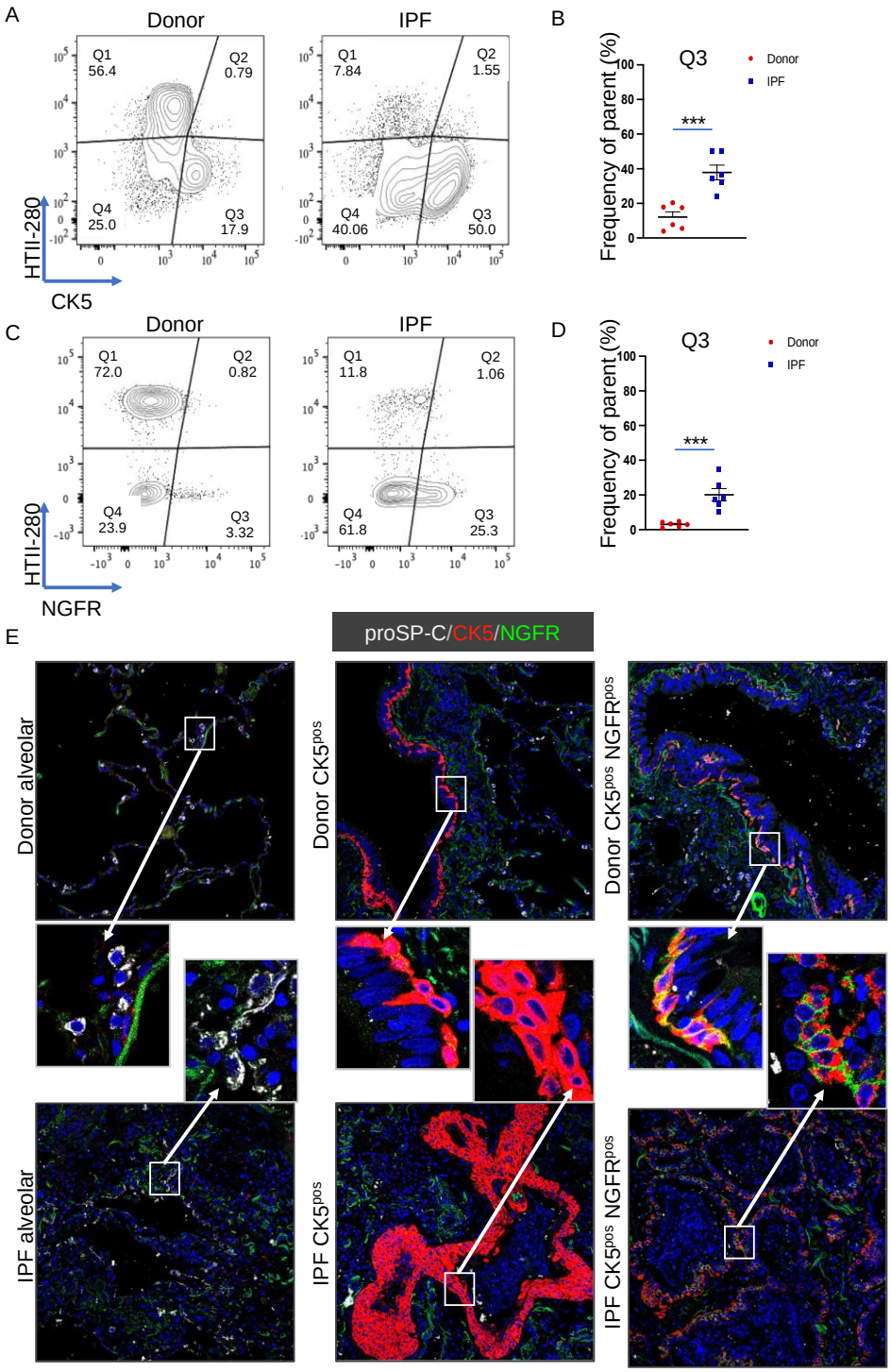


Figure 6. CK5 and NGFR expression in donor and IPF epithelial cells. (A) Representative flow cytometry panels of CK5 vs HTII-280 expression in the epithelial cell compartment of donor (n=6, left) and IPF (n=6, right) lungs. (B) Quantification of the $CK5^{pos}$ $HTII\text{-}280^{neg}$ (Q3) population shown in (A). (C) Representative flow cytometry panels of NGFR vs HTII-280 expression in the epithelial cell compartment of donor (n=6, left) and IPF (n=6, right) lungs. (D) Quantification of the $NGFR^{pos}$ $HTII\text{-}280^{neg}$ (Q3) population shown in (C). Data are presented as the mean ± SEM of percentage cells from the parent population. Statistical analysis was performed on Log(10) values. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, n.s. = not significant by ANOVA (E) Representative confocal images of proSP-C (white signal), CK5 (red signal) and NGFR (green signal) in different locations of donor (upper images) and IPF (lower images) lung.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 2 December 2021 doi:10.20944/preprints202112.0033.v1

3.7. CD24 upregulation in IPF

CD24 was identified as a cell surface marker highly expressed by aberrant basaloid cells[31] and in the KRT5-/KRT17+ intermediate cells[32] (Figure 7A and Supplemental Figure 7A). In our transcriptomic analysis, CD24 was also markedly increased in IPF Lyso^{low} cells (LFC =2.92, LOG(p)=3.1). To determine which epithelial cells express CD24, we stained four donor and three IPF lung cell preparations with the usual cell surface markers in combination with a CD24 antibody and its expression was analyzed in the epithelial compartment as a function of lysotracker uptake. In donor epithelial cells, the proportion of CD24^{pos} cells was very small (average 1.4% in Lyso^{high} and 3.41% in Lyso^{low} population), but its expression was markedly increased in both IPF Lyso^{high} and Lyso^{low} populations (average 15.2% and 28.5% respectively – Figure 7B-D, individual patient data in Supplemental Figure 7B, C). Given the basal/AEC2 profile of the Lyso^{low} CD24^{pos} population suggested by our transcriptomic data, our cell surface expression analysis allows us to speculate that the CD24^{pos} Lyso^{low} cells might represent a sub-population of IPF specific intermediate cell type.

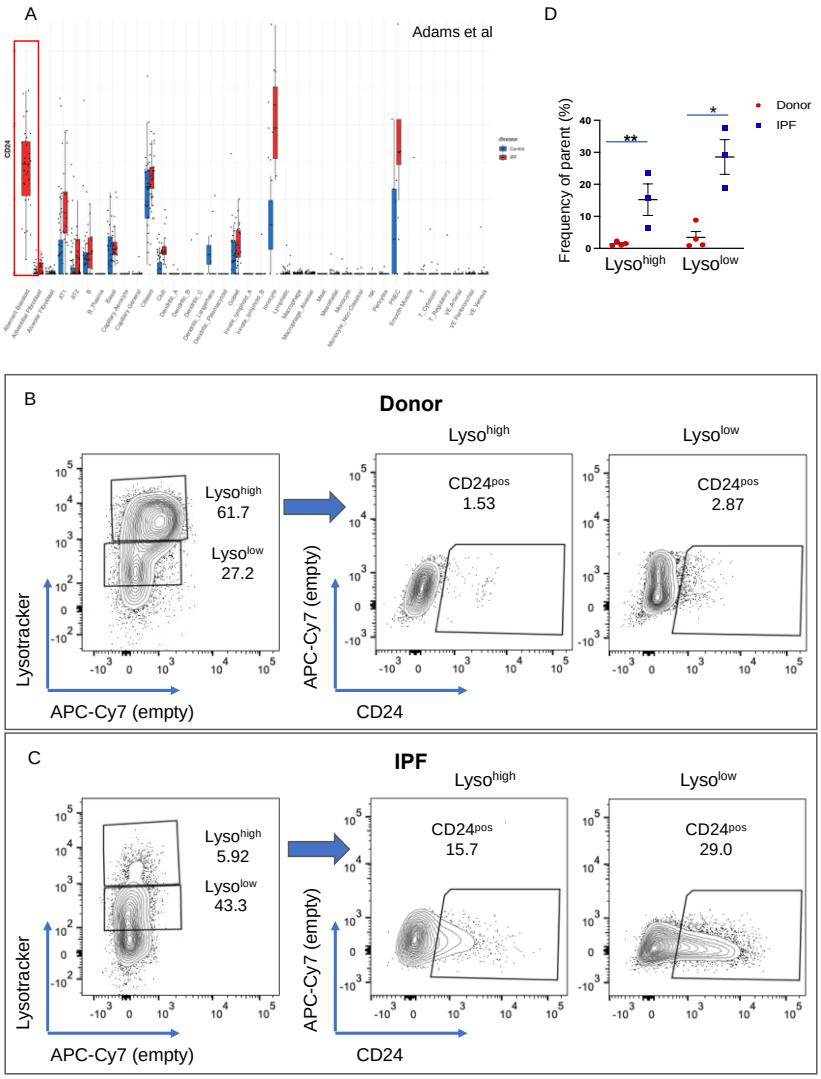


Figure7. CD24 expression in donor and IPF lung. (A) Differential expression of CD24 in donor and IPF scNGS data published by Adams et al. (B), (C) Flow cytometry analysis of n=4 donor (B) and n=3 IPF (C) of Lysotracker incorporation in the epithelial cell

compartment (left panel). Right panels show the expression of CD24 in the Lyso^{high} and Lyso^{low} populations shown in the panels on the left. (D) Quantification of the data in (B) and (C) showing the difference between donor and IPF Lyso^{high} and Lyso^{low} populations. Average donor Lyso^{high} CD24^{pos} 1.4%, donor Lyso^{low}CD24^{pos} 0.34%; IPF Lyso^{high} CD24^{pos} 15.2% and IPF Lyso^{low} CD24^{pos} 28.53%. Data are presented as the mean ± SEM of percentage cells from the parent population. Statistical analysis was performed on Log(10) values. *p < 0.05, **p < 0.01, ***p < 0.001, n.s. = not significant by ANOVA

4. Discussion

Here we provide an in-depth phenotypical analysis of human alveolar epithelial cells isolated from donor and end-stage IPF explanted lungs. In doing so, we identify two populations which differ markedly in their ability to process and secrete surfactant, the defining differentiated function of AEC2s. During bleomycin induced lung fibrosis the two populations vary in complementary directions, suggestive of correlative behavior. Comparing the transcriptome of the IPF and donor Lyso^{pos} populations in human lung, we determine that the IPF Lyso^{pos} population co-expresses markers of basal and AEC2 lineages. We further confirm by flow cytometry and immunofluorescence analysis the CK5^{pos} cell expansion in IPF, and show that CK5 and NGFR expression define two distinct basal cell populations with differential behavior in IPF. The increased expression in IPF of CD24, a cell surface marker known to be highly expressed in cancer cells[50] and aberrant epithelial cells in IPF[30,31], is also confirmed, offering a potential cell surface marker to sub-type together with Lysotracker incorporation various populations of epithelial cells in donor and IPF lung.

The emergence of single cell transcriptomics led to the in-depth profiling of already known populations of cells, and the identification of other novel ones in the mouse and human lung[27,29–31,35,40,41]. Recent landmark papers led to the identification of epithelial populations with an intermediate transcriptomic signature in the IPF lung. First, a population of cells with "intermediate phenotype" that resembled the transcriptomic profiles of both AEC2s and basal cells was identified by Xu et al[28]. Recently, a similar population of cells specific for IPF called aberrant basal cells was described by Reyfman et al and by Adams et al[30,31].

Similarly, we find that the Lyso^{pos} population in IPF expresses several markers defining these intermediate populations. Characteristic markers like KRT5, 15 and 17, ITGA2, Sox4, Sox9 and CD24 are expressed at high levels, together with AEC2 genes SFTPB and C. Genes involved in surfactant protein processing and secretion are also expressed in the IPF Lyso^{pos} population, although they do not reach the level of expression of the donor AEC2s, which correlates well with the low, intermediate levels of Lysotracker incorporation in this population. Interestingly, two genes defining the AEC2 cell fate, *ETV5* and *SOX9* are also expressed in the IPF Lyso^{pos} population, but while *SOX9* is expressed at levels exceeding that of bona-fide AEC2s, *ETV5* expression is much lower, suggesting that the they regulate different aspects of the IPF alveolar epithelial fate. Indeed, *SOX9* role is crucial for mouse and human distal lung epithelium specification[52–54], while *ETV5*, acting downstream of *SOX9* [58,57] and FGF signaling[58,59] is crucial for AEC2 fate maintenance, and is downregulated in transitional states as the AEC2 to AEC1 transition[60]. Interestingly, a recent paper identified a population with similar characteristics in the mouse lung[61]. Lineage tracing alveolar epithelial cells using the

SFTPCcreER / Rosa26^{TdTomato} double transgenic mice revealed a TdTomato^{low} population of cells which express low levels of AEC2 markers proSP-C, *etv5* and *Fgfr2b*, and has progenitor cell properties. This suggests that a similar population with intermediate AEC2 characteristics and progenitor properties exists in the mouse lung, which is supported by our data showing the maximum expansion of the Lyso^{low} population at the peak of epithelial proliferation (day 14) following bleomycin injury.

There are multiple circumstances that require the transition through an intermediate fate. First, it has been shown that multiple progenitors can participate in the repair of the alveolar epithelium[62]. It is thus conceivable that they converge on a common intermediate state on their way to becoming fully differentiated AEC2s. Second, aberrant progenitors, that in a normal state do not participate in alveolar repair can be recruited when the local progenitor pool is exhausted, as is the case of the IPF lung, converging towards the same intermediate fate. Third, AEC2 divergent interemediate fates could also emerge, such as (1) AEC2s differentiating into AEC1s[44], (2) AEC2s de-differentiating in order to assume progenitor function[22] or (3) AEC2s that temporarily limit their differentiated function to allow recovery from injury. Thus, although the Lyso^{low} population appears homogenous from a phenotypic perspective, as seen by the surfactant protein expression and Lysotracker incorporation, one cannot exclude that it might represent a lineage-diverse population. Based on our data we propose that the Lyso^{low} population represents a stable cellular state on the way to or from a mature AEC2, rather than a particular cell type. Our population level transcriptomic analysis does not allow us to draw conclusions about the lineage composition or transcriptomic heterogeneity of the Lyso^{low} population, but flow cytometry analysis offers a modality of isolating cells in this intermediate state for further analysis.

In conclusion we show that Lysotracker incorporation defines two cellular states in donor and IPF distal epithelial lung, with the Lyso^{high} state representing bona-fide AEC2s and the Lyso^{low} state characterizing an intermediate cell population displaying both basal and AEC2 characteristics.

Author Contributions: Conceptualization, Roxana Wasnick and Andreas Guenther; Data curation, Jochen Wilhelm; Formal analysis, Roxana Wasnick, Irina Shalashova, Jochen Wilhelm, Ali Khadim, Nicolai Schmidt and Elie El Agha; Investigation, Roxana Wasnick, Irina Shalashova, Jochen Wilhelm and Elie El Agha; Methodology, Jochen Wilhelm, Ali Khadim, Nicolai Schmidt and Holger Hackstein; Project administration, Werner Seeger; Resources, Holger Hackstein, Andreas Hecker, Konrad Hoetzenecker, Werner Seeger, Clemens Ruppert and Andreas Guenther; Supervision, Andreas Guenther; Writing – original draft, Roxana Wasnick; Writing – review & editing, Roxana Wasnick, Saverio Bellusci, Elie El Agha, Clemens Ruppert and Andreas Guenther.

Funding: This study was supported in part by the RARE-ILD consortium (European Joint Program on Rare Diseases, EJP-RD), the Lung Fibrosis Stipend of the Lung Clinic Waldhof Elgershausen, the Clinical Research Group (KFO) 309 (Virus-induced Lung Injury, German Research Council, DFG).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the Justus-Liebig-University School of Medicine (No. 31/93, 29/01, and No. 111/08: European IPF Registry), and informed consent was obtained in written form from each subject. Animal studies were performed in accordance with the Helsinki convention for the use and care of animals and were approved by the local authorities at Regierungspräsidium Giessen V54-19 c 2015 (1) GI 20/10 Nr. 109/2011 (Bleomycin) or V54-19 c 20 15 h 02 GI 20/10 Nr. A53/2012 (untreated controls).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study prior to biospecimen collection.

Data Availability Statement: Transcriptomic data is publicly available at the following location: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE185691

Acknowledgments: The authors would like to acknowledge the technical assistance of Simone Becker, Gabriele Dahlem. We would like to thank Gabriela Michel and Dr. Neli Baal from the Zentrum für Transfusionsmedizin und Hämotherapie, Institut für Klinische Immunologie (UKGM Giessen) for their priceless assistance with flow cytometry. We also acknowledge the extremely valuable contributions of the Multiscale Imaging Platform and the Genomic and Bioinformatics Platform of the Institute for Lung Health (ILH - Giessen). We would also like to thank Dr. Walter Klepetko and the surgical team for collecting the patient material, Dr. Ludger Fink for the pathological assessment, and the patients themselves without who this work would not have been possible.

Conflicts of Interest: The authors declare no conflict of interest.

5. References

1. Günther, A. *et al.* Unravelling the progressive pathophysiology of idiopathic pulmonary fibrosis. *European Respiratory Review* **21**, 152–160 (2012).
2. Raghu, G. *et al.* AMERICAN THORACIC SOCIETY DOCUMENTS Diagnosis of Idiopathic Pulmonary Fibrosis An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *American Journal of Respiratory and Critical Care Medicine* **198**, 44–68 (2018).
3. Olson, A. L. & Swigris, J. J. Idiopathic Pulmonary Fibrosis: Diagnosis and Epidemiology. *Clinics in Chest Medicine* **33**, 41–50 (2012).
4. Shumar, J. N., Chandel, A., King, C. S. & Bendstrup, M. Clinical Medicine Antifibrotic Therapies and Progressive Fibrosing Interstitial Lung Disease (PF-ILD): Building on INBUILD. *J. Clin. Med* **10**, 2285 (2021).
5. Parimon, T., Hohmann, M. S., Yao, C., Marchioni, A. & Tonelli, R. Pathogenic Mechanisms in Lung Fibrosis. *Int. J. Mol. Sci* **2021**, 6214 (2021).
6. Noble, P. W., Barkauskas, C. E. & Jiang, D. Review series Pulmonary brosis : patterns and perpetrators. *Am J Physiol Lung Cell Mol Physiol* **122**, 4–10 (2012).

7. Sisson, T. H. *et al.* Targeted injury of type II alveolar epithelial cells induces pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine* **181**, 254–263 (2010).
8. C, Y. *et al.* Senescence of Alveolar Type 2 Cells Drives Progressive Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **203**, 707–717 (2021).
9. Camelo, A., Dunmore, R., Sleeman, M. A. & Clarke, D. L. The epithelium in idiopathic pulmonary fibrosis: Breaking the barrier. *Frontiers in Pharmacology* **4 JAN**, 1–11 (2014).
10. Desai, T. J., Brownfield, D. G. & Krasnow, M. A. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* **507**, 190–194 (2014).
11. Zacharias, W. J. *et al.* Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* **555**, 251–255 (2018).
12. T, H., J, O., F, F., H, V. & P, D. Dynamics of surfactant release in alveolar type II cells. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1579–1584 (1998).
13. Weaver, T. E. Synthesis, processing and secretion of surfactant proteins B and C. *Biochimica et Biophysica Acta - Molecular Basis of Disease* **1408**, 173–179 (1998).
14. Korimilli, A., Gonzales, L. W. & Guttentag, S. H. Intracellular localization of processing events in human surfactant protein B biosynthesis. *Journal of Biological Chemistry* **275**, 8672–8679 (2000).
15. Sever, N., Miličić, G., Bodnar, N. O., Wu, X. & Rapoport, T. A. Mechanism of Lamellar Body Formation by Lung Surfactant Protein B. *Molecular Cell* **81**, 49-66.e8 (2021).
16. Strunz, M. *et al.* Longitudinal single cell transcriptomics reveals Krt8+ alveolar epithelial progenitors in lung regeneration. *bioRxiv* (2019) doi:10.1101/705244.
17. Vaughan, A. E. *et al.* Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature* **517**, 621–625 (2015).
18. Liang, J. *et al.* Hyaluronan and TLR4 promote surfactant-protein-C-positive alveolar progenitor cell renewal and prevent severe pulmonary fibrosis in mice. *Nature Medicine* **22**, 1285–1293 (2016).
19. Lopez-Rodriguez, E. *et al.* Surfactant dysfunction during overexpression of TGF-β1 precedes profibrotic lung remodeling in vivo. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **310**, L1260–L1271 (2016).
20. Beike, L. *et al.* Surfactant dysfunction and alveolar collapse are linked with fibrotic septal wall remodeling in the TGF-β1-induced mouse model of pulmonary fibrosis. *Laboratory Investigation* **99**, 830–852 (2019).
21. Mulugeta, S., Nureki, S. I. & Beers, M. F. Lost after translation: Insights from pulmonary surfactant for understanding the role of alveolar epithelial dysfunction and cellular quality control in fibrotic lung disease. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **309**, L507–L525 (2015).
22. Wasnick, R. M. *et al.* Restored alveolar epithelial differentiation and reversed human lung fibrosis upon Notch inhibition. *bioRxiv* 580498 (2019) doi:10.1101/580498.
23. Watanabe, S. *et al.* Resetting proteostasis with ISRIB promotes epithelial differentiation to attenuate pulmonary fibrosis. *Proceedings of the National Academy of Sciences of the United States of America* **118**, (2021).
24. Chapman, H. A. *et al.* Integrin α6β4 identifies an adult distal lung epithelial population with regenerative potential in mice. *Journal of Clinical Investigation* **121**, 2855–2862 (2011).
25. Watson, J. K. *et al.* Distal lung epithelial progenitor cell function declines with age. *Scientific Reports* **10**, (2020).

26. Yee, M. *et al.* Alternative progenitor lineages regenerate the adult lung depleted of alveolar epithelial type 2 cells. *American Journal of Respiratory Cell and Molecular Biology* **56**, 453–464 (2017).
27. Smirnova, N. F. *et al.* Detection and quantification of epithelial progenitor cell populations in human healthy and IPF lungs. (2016) doi:10.1186/s12931-016-0404-x.
28. Xu, Y. *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* **1**, 1–18 (2016).
29. Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
30. Reyfman, P. A. *et al.* Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* **199**, 1517–1536 (2019).
31. Adams, T. S. *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Science Advances* **6**, (2020).
32. Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Science Advances* **6**, (2020).
33. McDonough, J. E. *et al.* Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight* **4**, 131597 (2019).
34. American Thoracic Society Idiopathic Pulmonary Fibrosis: Diagnosis and Treatment International Consensus Statement.
35. Teisanu, R. M., Lagasse, E., Whitesides, J. F. & Stripp, B. R. Prospective Isolation of Bronchiolar Stem Cells Based Upon Immunophenotypic and Autofluorescence Characteristics. *STEM CELLS* **27**, 612–622 (2009).
36. Teisanu, R. M. *et al.* Functional Analysis of Two Distinct Bronchiolar Progenitors during Lung Injury and Repair. doi:10.1165/rcmb.2010-0098OC.
37. Tighe, R. M. *et al.* AMERICAN THORACIC SOCIETY DOCUMENTS Improving the Quality and Reproducibility of Flow Cytometry in the Lung An Official American Thoracic Society Workshop Report. (2019) doi:10.1165/rcmb.2019-0191ST.
38. Smyth, G. K. Limma: Linear Models for Microarray Data.
39. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology 2004 5:10* **5**, 1–16 (2004).
40. Smyth, G. K. & Speed, T. Normalization of cDNA microarray data. *Methods* **31**, 265–273 (2003).
41. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3**, (2004).
42. van der Velden, J. L., Bertoncello, I. & McQualter, J. L. LysoTracker is a marker of differentiated alveolar type II cells. *Respiratory Research* **14**, (2013).
43. Haller, T., Ortmayr, J., Friedrich, F., Völkl, H. & Dietl, P. Dynamics of surfactant release in alveolar type II cells. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1579–1584 (1998).
44. Strunz, M. *et al.* Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nature Communications* **11**, (2020).
45. Williamson, J. D., Sadofsky, L. R. & Hart, S. P. The pathogenesis of bleomycin-induced lung injury in animals and its applicability to human idiopathic pulmonary fibrosis. *Experimental Lung Research* **41**, 57–73 (2015).

46. Gonzalez, R. F., Allen, L., Gonzales, L., Ballard, P. L. & Dobbs, L. G. HTII-280, a Biomarker Specific to the Apical Plasma Membrane of Human Lung Alveolar Type II Cells. *Journal of Histochemistry & Cytochemistry* **58**, 891–901 (2010).
47. Schiller, H. B. *et al.* The Human Lung Cell Atlas: A High-Resolution Reference Map of the Human Lung in Health and Disease. *Am J Respir Cell Mol Biol* **61**, 31–41 (2019).
48. Konda, B. *et al.* Isolation and enrichment of human lung epithelial progenitor cells for organoid culture. *Journal of Visualized Experiments* **2020**, 1–17 (2020).
49. Rock, J. R. *et al.* Basal cells as stem cells of the mouse trachea and human airway epithelium.
50. Sinjab, A. *et al.* Resolving the spatial and cellular architecture of lung adenocarcinoma by multiregion single-cell sequencing. *Cancer Discovery* candisc.1285.2020 (2021) doi:10.1158/2159-8290.CD-20-1285.
51. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
52. Perl, A.-K. T., Kist, R., Shan, Z., Scherer, G. & Whitsett, J. A. Normal lung development and function after Sox9 inactivation in the respiratory epithelium. *genesis* **41**, 23–32 (2005).
53. Rockich, B. E. *et al.* Sox9 plays multiple roles in the lung epithelium during branching morphogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **110**, (2013).
54. Danopoulos, S. *et al.* Human lung branching morphogenesis is orchestrated by the spatiotemporal distribution of ACTA2, SOX2, and SOX9. *American Journal of Physiology - Lung Cellular and Molecular Physiology* **314**, L144–L149 (2018).
55. Ludbrook, L., Alankarage, D., Bagheri-Fam, S. & Harley, V. Dataset of differentially expressed genes from SOX9 over-expressing NT2/D1 cells. *Data in Brief* **9**, 194–198 (2016).
56. Alankarage, D. *et al.* SOX9 regulates expression of the male fertility gene Ets variant factor 5 (ETV5) during mammalian sex development. *International Journal of Biochemistry and Cell Biology* **79**, 41–51 (2016).
57. Reginensi, A. *et al.* SOX9 controls epithelial branching by activating RET effector genes during kidney development. doi:10.1093/hmg/ddq558.
58. Jones, M. R. *et al.* Characterization of Tg(Etv4-GFP)and Etv5RFP reporter lines in the context of fibroblast growth factor 10 signaling during mouse embryonic lung development. *Frontiers in Genetics* **10**, 1–12 (2019).
59. Liu, Y., Jiang, H., Crawford, H. C. & Hogan, B. L. M. Role for ETS domain transcription factors Pea3/Erm in mouse lung development. *Developmental Biology* **261**, 10–24 (2003).
60. Zhang, Z. *et al.* Transcription factor Etv5 is essential for the maintenance of alveolar type II cells. *Proceedings of the National Academy of Sciences of the United States of America* **114**, 3903–3908 (2017).
61. Ahmadvand, N. *et al.* Identification of a novel subset of alveolar type 2 cells enriched in PD-L1 and expanded following pneumonectomy. *European Respiratory Journal* **58**, 2004168 (2021).
62. Leach, J. P. & Morrisey, E. E. Repairing the lungs one breath at a time: How dedicated or facultative are you? *Genes and Development* vol. 32 1461–1471 (2018).