PLOS ONE

RESEARCH ARTICLE

Dysregulation of club cell biology in idiopathic pulmonary fibrosis

Wu-Lin Zuo[1☯¤], **Mahboubeh R. Rostami**[1☯], **Michelle LeBlanc**[1], **Robert J. Kaner**[1,2], **Sarah L. O'Beirne**[1,2], **Jason G. Mezey**[1,3], **Philip L. Leopold**[1], **Karsten Quast**[4], **Sudha Visvanathan**[5], **Jay S. Fine**[5], **Matthew J. Thomas**[4], **Ronald G. Crystal**[1,2] *

1 Department of Genetic Medicine, Weill Cornell Medical College, New York, New York, United States of America, 2 Department of Medicine, Weill Cornell Medical College, New York, New York, United States of America, 3 Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, New York, United States of America, 4 Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach an der Riss, Germany, 5 Boehringer Ingelheim Pharmaceuticals, Ridgefield, Connecticut, United States of America

☯ These authors contributed equally to this work.
¤ Current address: The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, United States of America
* geneticmedicine@med.cornell.edu

Abstract

Idiopathic pulmonary fibrosis (IPF) is a progressive, chronic fibrotic lung disease with an irreversible decline of lung function. "Bronchiolization", characterized by ectopic appearance of airway epithelial cells in the alveolar regions, is one of the characteristic features in the IPF lung. Based on the knowledge that club cells are the major epithelial secretory cells in human small airways, and their major secretory product uteroglobin (SCGB1A1) is significantly increased in both serum and epithelial lining fluid of IPF lung, we hypothesize that human airway club cells contribute to the pathogenesis of IPF. By assessing the transcriptomes of the single cells from human lung of control donors and IPF patients, we identified two SCGB1A1[+] club cell subpopulations, highly expressing MUC5B, a significant genetic risk factor strongly associated with IPF, and SCGB3A2, a marker heterogeneously expressed in the club cells, respectively. Interestingly, the cellular proportion of SCGB1A1[+-]MUC5B[+] club cells was significantly increased in IPF patients, and this club cell subpopulation highly expressed genes related to mucous production and immune cell chemotaxis. In contrast, though the cellular proportion did not change, the molecular phenotype of the SCGB1A1[+]SCGB3A2[high] club cell subpopulation was significantly altered in IPF lung, with increased expression of mucins, cytokine and extracellular matrix genes. The single cell transcriptomic analysis reveals the cellular and molecular heterogeneity of club cells, and provide novel insights into the biological functions of club cells in the pathogenesis of IPF.

Introduction

Club cells, characterized by the apical dome shaped morphology with dense cytoplasmic granules and short microvilli, represent approximate 20% of the epithelial cells and are the major

secretory cells in human small airway epithelium (SAE) [1–3]. Secretoglobin family 1A member 1 (SCGB1A1), a secreted protein with anti-inflammatory properties, is the cell-specific marker and major secretory product for human SAE club cells [4–6]. Single cell RNA-sequencing of human SAE has identified the biological functions of club cells, including host defense, physical barriers and their potential roles in the pathogenesis of monogenetic and infectious lung disorders [3].

In asymptomatic smokers and chronic obstructive pulmonary disease (COPD), the number of club cells and the expression of club cell marker SCGB1A1 in the human SAE are decreased [1, 7], and the SCGB1A1 levels in lung epithelial lining fluid is decreased in asthma [8, 9], together suggesting that COPD and asthma are the club cell deficiency disorders. In contrast, SCGB1A1 expression is increased in both serum and epithelial lining fluid of idiopathic pulmonary fibrosis (IPF) [10]. Murine studies of club cells in the pulmonary fibrosis are mixed. Depletion of club cells by naphthalene suppress bleomycin-induced lung injury and fibrosis, while over-expression of another club cell marker SCGB3A2 in mouse lung exhibits an anti-fibrotic activity [11, 12]. Based on these observations, it is likely that the club cells play a unique role in the pathogenesis of IPF.

To assess this concept, we evaluated the club cell populations in the single-cell RNA-sequencing data from controls *vs* IPF described by Reyfman et al [13]. Analysis identified two unique club cell sub-populations, a SCGB1A1$^+$ club cell sub-population highly expressing SCGB3A2, another club cell marker [14, 15], and a second SCGB1A1$^+$ sub-population expressing MUC5B, a known genetic risk gene for IPF [16]. The proportion of SCGB1A1$^+$MUC5B$^+$ club cells was increased in IPF, with high expression of genes-related to mucins and immune cell chemoattractants. In contrast, the proportion of SCGB1A1$^+$SCGB3A2^{high} club cells was similar to the controls, but the transcriptome of the SCGB1A1$^+$SCGB3A2^{high} club cells was significantly dysregulated in IPF, with increased gene expression related to extracellular matrix formation, mucins and the growth factors relevant to pulmonary fibrosis. Together, these data provide novel insights into the molecular phenotypes and biological functions of club cells in the pathogenesis of IPF.

Methods

Source of single-cell RNA-sequencing data

The single-cell RNA-sequencing data described by Reyfman et al [13] was downloaded from a publically available database (Gene Expression Omnibus, series GSE122960). The data in the database did not contain protected health information or patient identifiers. The original publication noted that all procedures used to obtain tissue were reviewed by Institutional Review Boards as well as the funding agency and that patients involved in the study had provided written informed consent [13]. A total of 18,887 single cells from 4 control donors and 13,256 single cells from 4 IPF patient were analyzed. Detailed information of the 4 controls and the 4 IPF patients characterized by Reyfman et al [13] are summarized in S1 Table in S1 File.

Data analysis

Processing of the single-cell RNA-sequencing data was performed using Seurat package V2 and R 3.5 [17]. Initially, gene expression matrices from the 4 controls and 4 IPF patients were combined. Cells were filtered out that have unique gene counts over 10,000 or less than 200 and the percentage of mitochondrial genes for each cell were less than 0.25. Gene expression matrix of 32,143 cells for these eight samples and total number of 18,732 genes was normalized by the total number of unique molecular identifiers (UMIs) per cell, multiplying by a scale factor (10,000) and then log transformed. The resulting 1,069 variably expressed genes were

summarized by principal component analysis (PCA) and the 6 principle components further summarized using t-Distributed Stochastic Neighbor (t-SNE) dimensionally reduction method using RunTSNE function in Seurat package with perplexity parameter set to 30. The clustering algorithm, K-nearest neighbor graph based on the euclidean distance in PCA space was constructed by FindClusters function in Seurat to iteratively group cells together. Identification of clusters was guided by identifying marker genes by FindallMarkers function by Wilcox method in Seurat. Differential gene expression analysis was performed for each cluster between cells from controls *vs* IPF patients.

To evaluate the club cell populations, we re-analyzed cluster 1, which highly expressed club cell marker SCGB1A1. The most variable genes in cluster 1 cells were identified and applied dimensionally reduction PCA. The standard deviations of the PCs were plotted using a cutoff to identify which PCA is informative. The clustering algorithm, K-nearest neighbor graph based in PCA space approach implemented in FindClusters function was applied to find sub-clusters in cluster 1 cells. The resolution parameter was set as 0.5. This yielded 7 sub-clusters. The cell type of each sub-cluster of cluster 1 was identified by finding marker genes, and the sub-clusters visualized with t-SNE plots. Slingshot was applied for inferring cell lineage and pseudotimes from single cell RNA seq data [18]. The Slingshot algorithm uses prior knowledge for dimensional reduction and clustering and can capture multiple lineages in a dataset.

Results

Cellular and transcriptomic heterogeneity

To assess the cellular and transcriptomic heterogeneities of cell populations in the control donors and IPF patients, 18,887 single cells from 4 control donors and 13,256 single cells from 4 IPF patients were analyzed. The initial unsupervised t-SNE clustering of these single cells from human lung identified a total of 14 unique cell populations in both control donors (Fig 1A) and IPF patients (Fig 1C). Based on signature genes for each cell population and the expression of the canonical markers for the cell populations, the 14 cell populations were assigned as: 1 –airway and alveolar epithelial cells, highly expressing club cell markers SCGB1A1 and SCGB3A2, basal cell (BC) marker KRT5 and KRT15, and type I alveolar cell marker AGER and HOPX; 2 –ciliated cells, highly expressing FOXJ1 and DNAI1; 3~5 –type II alveolar cells, highly expressing SFTPC and ABCA3; 6 –immune cells, highly expressing T cell marker CD3D and CD8A, mast cell marker TPSB2 and KIT, and B cell marker CD19 and MS4A1; 7 –plasma cells, highly expressing IGHG4 and CD27; 8~9 –monocytes, highly expressing CD14 and ITGAM; 10~13 –alveolar macrophages, highly expressing CD68 and PPARG; 14 –mesenchymal cells, highly expressing endothelial cell marker VWF and PECAM1, and fibroblast marker COL1A1 and COL14A1 (Fig 1B and 1D; S1 Fig and S2 Table in S1 File).

Club cell sub-populations in human lung of controls *vs* IPF

The airway epithelium are associated with the pathogenesis of IPF [19–24]: (1) there is small airway pathology and dysfunction in IPF [19, 20] (2) the numbers of epithelial BC are increased in IPF [20, 22], and (3) BC genes are associated with the mortality in IPF [22]. To further study the role of airway epithelial cells in the IPF pathogenesis, the 569 single cells from control donors and 1,013 single cell from IPF patients in the airway and alveolar epithelial cell cluster (cluster 1 in Fig 1A and 1C) were re-analyzed. The unsupervised t-SNE clustering identified 6 and 7 cell populations in the control donors and IPF patients, respectively (Fig 2A, S2 Fig in S1 File). Using the signature genes for each cluster and the known markers for the epithelial cell populations, we assigned these cell populations as: 1 –BC, highly expressing
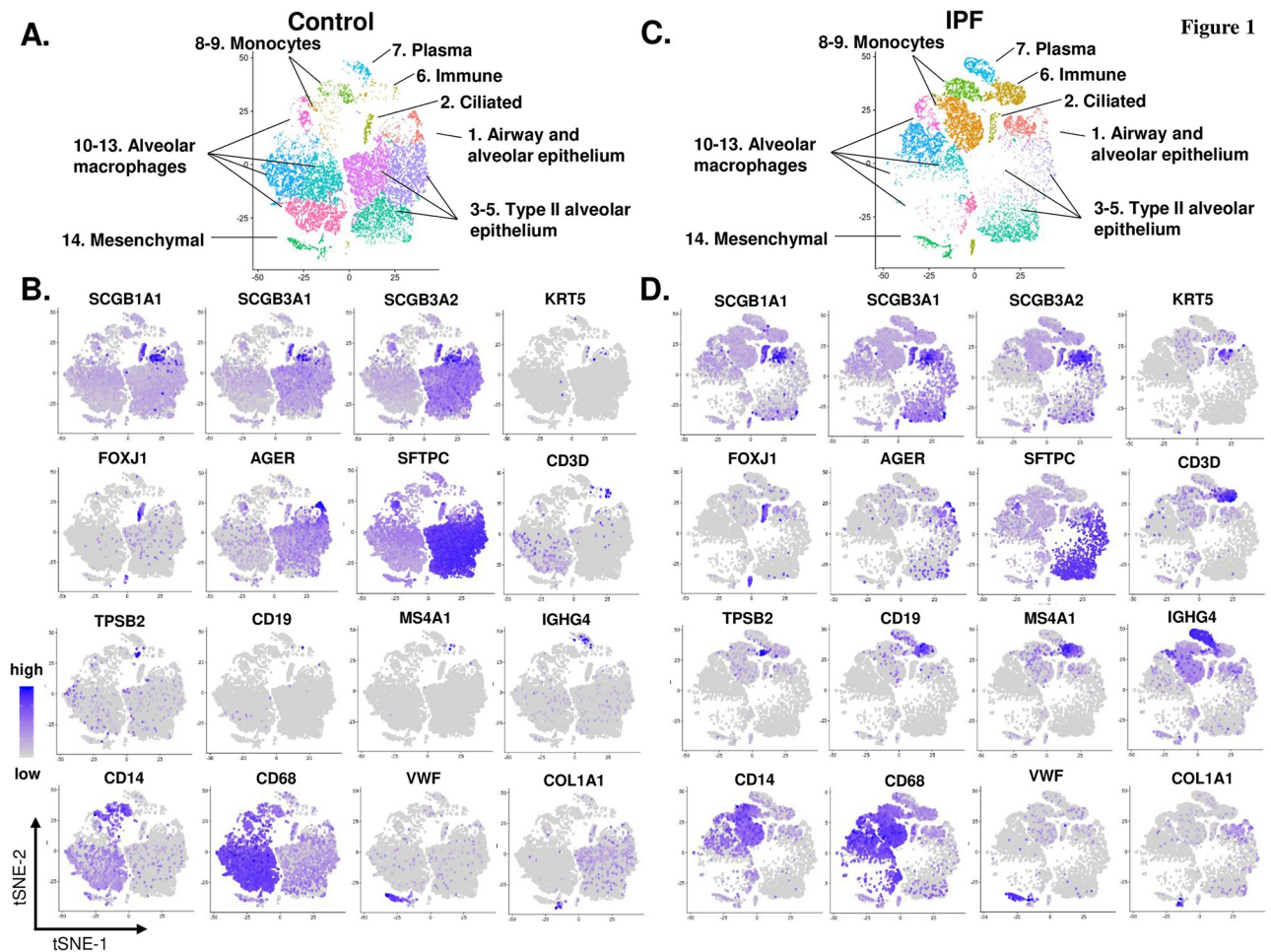
Fig 1. Identification of unique cell populations from the human lungs of controls *vs* IPF patients by single cell RNA-sequencing. Based on the data of Reyfman et al [13], 18,887 single cells from the lungs of 4 controls and 13,256 single cells from the lungs of 4 IPF patients (see S1 Table in S1 File for the detailed information of controls and IPF patients) were generated from the data in NCBI Gene Expression Omnibus (GEO accession number: GSE122960). **A.** Unsupervised t-distributed Stochastic Neighbor Embedding (t-SNE) clustering of the single cells from the lungs of 4 controls. A total of 14 unique cells populations were identified. **B.** Expression of the canonical markers of the different cell populations from control human lungs in the t-SNE plot shown in panel A. Club cell markers–SCGB1A1, SCGB3A1 and SCGB3A2; basal cell–KRT5; ciliated cell–FOXJ1; type I alveolar cells–AGER; type II alveolar cells–SFTPC; T cells–CD3D; mast cells–TPSB2; B cells–CD19 and MS4A1; plasma cells–IGHG; monocytes–CD14; alveolar macrophages–CD68; endothelial cells–VWF; and fibroblasts–COL1A1. **C.** Unsupervised t-SNE clustering of the single cells from the lung of 4 IPF patients. The same 14 cell populations were identified in IPF patients, as in the controls. **D.** Expression of the canonical markers of the different cell populations from in the IPF t-SNE plot shown in panel **C.** The markers are the same as in panel **B.** Based on signature genes (see S2 and S3 Tables in S1 File) and the expression of canonical markers of the different cell populations, the 14 cell populations identified in panels **A.** and **C.** were defined as: 1. airway (including basal and club cells) and alveolar (including type I and II alveolar cells) epithelial cells; 2. ciliated; 3–5. type II alveolar cells; 6. immune cells (including T cells, mast cells and B cells); 7. plasma cells; 8–9. monocytes; 10–13. alveolar macrophages; and 14. mesenchymal cells (including endothelial cells and fibroblasts).

https://doi.org/10.1371/journal.pone.0237529.g001

KRT5; 2~3 –club cells, highly expressing SCGB1A1; 4~6 –intermediate alveolar cells, expressing both type II (SFTPC, ABCA3) and I (AGER, HOPX) alveolar cell markers; 7 –type I alveolar cells, highly expressing AGER and HOPX (Fig 2B and 2C, S3A Fig and S3 Table in S1 File).

Consistent with the literature that the airway and alveolar epithelial cells are reorganized in the IPF lung [19–26], the proportions of airway BC and alveolar cells were increased and decreased, respectively (S3B Fig in S1 File). Two club cell subpopulations, both highly expressing the club cell marker SCGB1A1, were identified in the single-cell data. One club cell
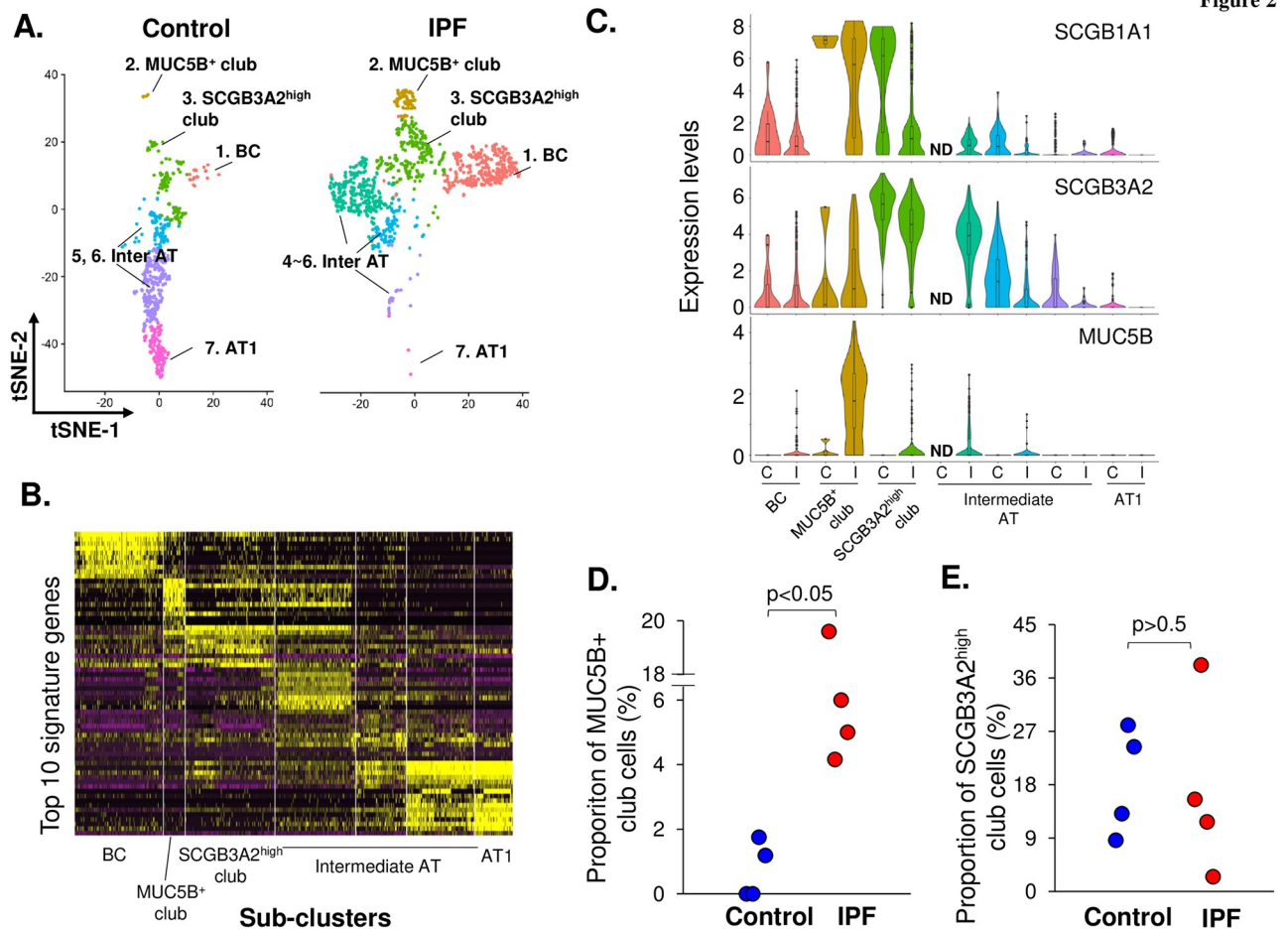
Fig 2. Identification of club cell sub-populations from the human lung of controls *vs* IPF patients by single cell RNA-sequencing. A. Unsupervised t-SNE clustering of the single cells from airway epithelial and alveolar cell cluster (cluster 1) in Fig 1A and 1C. A total of 6 unique cell populations were identified in controls (*left*) and 7 unique cell populations in IPF patients (*right*). Based on the signature genes for each cell population (see S3 Table and S3A Fig in S1 File), these cell populations were defined as: (1) basal cells–highly expressing KRT5; (2 and 3) club cells–highly expressing SCGB1A1; (4–6) intermediate alveolar cells–expressing type I alveolar cell markers AGER and HOPX, and type II alveolar cell marker SFTPC and ABCA3; (7) type I alveolar cells–highly expressing both AGER and HOPX. **B.** Heatmap plot in the expression of top 10 signature genes from the 7 unique cell populations identified in panel **A**. **C.** Violin plots showing the expression of club cell genes SCGB1A1 *(top)*, SCGB3A2 *(middle)*, and IPF genetic risk factor gene MUC5B *(bottom)* in the cell populations from control (C) and IPF (I) donors identified in panel **A**. Based on the violin plots, the sub-cluster 2 and 3 were defined as SCGB1A1+MUC5B+ and SCGB1A1+SCGB3A2high club cells, respectively. Note that sub-cluster 4 was not detected (ND) in control cell populations. **D.** Fraction of the SCGB1A1+MUC5B+ club cell sub-population in controls *vs* IPF patients. **E.** Fraction of the SCGB1A1+SCGB3A2high club cell sub-population in controls *vs* IPF. In panels **D-E.,** each dot represents one individual. *Blue dots*–controls, *red dots*–IPF patients. p values are shown.

subpopulation highly expressed MUC5B (MUC5B+ club cells), a known genetic risk gene for IPF [16], while the other club cell subpopulation highly expressed SCGB3A2 (SCGB3A2high club cells), a marker heterogeneously expressed in the club cells [14, 15] (Fig 2C). Both club cell populations were present in both control and IPF datasets (Fig 2C). While there were only a few MUC5B+ club cells in the control donors, the fraction of MUC5B+ club cells was significantly increased in the IPF patients (Fig 2D). In contrast, the proportion of SCGB3A2high club cells did not change in control donors *vs* IPF patients (Fig 2E).

Transcriptomic differences between MUC5B$^+$ and SCGB3A2^{high} club cells

To assess the differences between MUC5B$^+$ and SCGB3A2^{high} club cells, we first compared the signature genes of these 2 club cell subpopulations. There were total 124 and 135 signature genes for the MUC5B$^+$ and SCGB3A2^{high} club cells, respectively. Only 34 of these signature genes, including the club cell genes SCGB1A1, SLPI and WFDC2, were overlapping (Fig 3A), suggesting differing roles for these club cell populations in the pathogenesis of IPF. Compared to the SCGB3A2^{high} club cells, the IPF-induced MUC5B$^+$ club cells had a unique signature gene set, which included genes related to mucous cell differentiation (SPDEF, MUC5B, TFF3
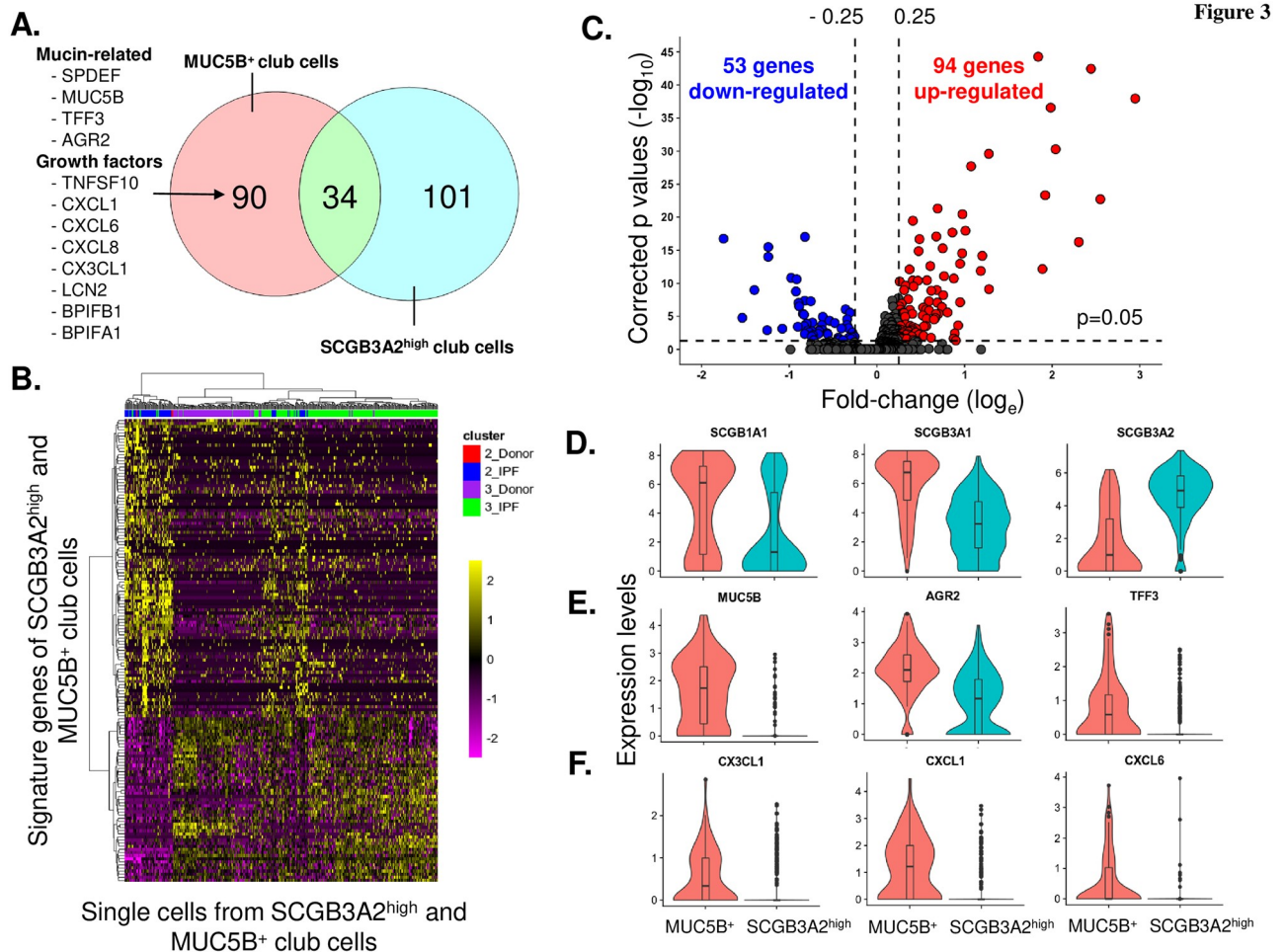

Fig 3. Transcriptomic differences between the club cell sub-populations. A. Venn diagram of the signature genes from club cell sub-populations. *Left*–signature genes of SCGB1A1$^+$MUC5B$^+$ club cells (n = 124), *right*–signature genes of SCGB1A1$^+$SCGB3A2^{high} club cells (n = 135), and overlap signature genes (n = 34). The SCGB1A1$^+$MUC5B$^+$ club cell unique signature genes related to mucin and growth factors are highlighted on the left of the Venn diagram. **B.** Heat map plot showing the expression of the signature genes in SCGB1A1$^+$SCGB3A2^{high} and SCGB1A1$^+$MUC5B$^+$ club cells. *X-axis*–signature genes, *y-axis*–single cells. *Y-axis*, *red*–control of SCGB1A1$^+$MUC5B$^+$ club cells; *blue*—IPF SCGB1A1$^+$MUC5B$^+$ club cells; *purple*–control SCGB1A1$^+$SCGB3A2^{high} club cells; and *green*—IPF SCGB1A1$^+$SCGB3A2^{high} club cells. **C.** Volcano plots of significantly down-regulated (53 genes, *blue*) and up-regulated (94 genes, *red*) genes in SCGB1A1$^+$MUC5B$^+$ club cells *vs* SCGB1A1$^+$SCGB3A2^{high} club cells. *Y-axis*–p values (-log$_{10}$), *x-axis*–the fold-change (log$_e$). Genes selected for analysis differed between populations by a fold-change (log$_e$) >0.25 as the up-regulated genes or <-0.25 as the down-regulated genes (vertical dotted lines), and p<0.05 with Bonferroni correction (horizontal line at y = 0.5). **D.** Violin plots of the expression of club cell genes (SCGB1A1, SCGB3A1 and SCGB3A2) in SCGB1A1$^+$MUC5B$^+$ and SCGB1A1$^+$SCGB3A2^{high} club cells. **E.** Violin plots in the expression of mucin-related genes (MUC5B, AGR2 and TFF3) in SCGB1A1$^+$MUC5B$^+$ and SCGB1A1$^+$SCGB3A2^{high} club cells. **F.** Violin plots in the expression of growth factor-related genes (CX3CL1, CXCL1, CXCL6 and CXCL8) in SCGB1A1$^+$MUC5B$^+$ and SCGB1A1$^+$SCGB3A2^{high} club cells. In **D.-F.**, *left*–SCGB1A1$^+$MUC5B$^+$, *right*–SCGB1A1$^+$SCGB3A2^{high} club cells.

and AGR2), growth factors related to host defense (LCN2, BPIFB1 and BPIFA1) and immune cell attractants (CXCL1, CXCL6, CXCL8 and CX3CL1) (Fig 3A). The heatmap plot of the signature genes also showed that MUC5B$^+$ and SCGB3A2^{high} club cells were clearly different, while a subset of the MUC5B$^+$ club cells in IPF patients co-expressed the SCGB3A2^{high} club cell signature genes (Fig 3B).

Comparison of the transcriptomes of the MUC5B$^+$ and SCGB3A2^{high} club cell subpopulations identified 53 genes significantly down-regulated and 94 genes significantly up-regulated in the MUC5B$^+$ club cells. In this MUC5B$^+$ subpopulation, the club cell markers SCGB1A1 and SCGB3A1 were up-regulated, while SCGB3A2 was down-regulated (Fig 3D). Consistent with the signature gene analysis, expression of the mucin-related genes (AGR2 and TFF3), and the immune cell chemoattractant-related genes were all significantly increased in the MUC5B$^+$ club cells (Fig 3E and 3F).

Transcriptomic dysregulation of the SCGB3A2^{high} club cells in IPF

Though the proportion of the SCGB3A2^{high} fraction club cells did not change, the transcriptome of the SCGB3A2^{high} club cells was significantly altered in IPF. A total of 125 genes were significantly down-regulated, while 110 genes were up-regulated in the IPF SCGB3A2^{high} club cells (Fig 4A). Gene ontology analysis identified that the down-regulated genes in SCGB3A2-^{high} club cells in IPF were mainly associated with translation, nuclear-transcribed mRNA catabolic process, viral transcription, response to hormone stimulus and cAMP (S4A Fig in S1 File). In contrast, the up-regulated genes in SCGB3A2^{high} club cells in IPF were related to
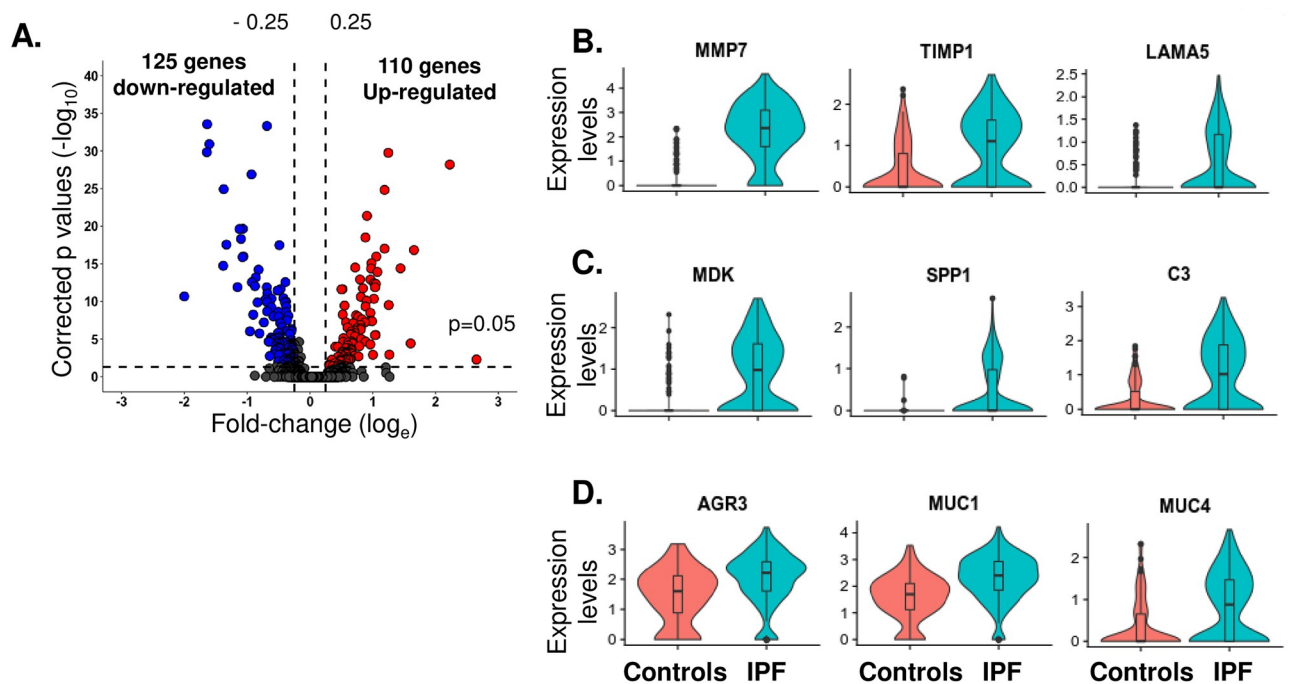

Fig 4. Transcriptomic dysregulation of SCGB3A2^{high} club cells in IPF. A. Volcano plots of significantly down-regulated (125 genes, *blue*) and up-regulated (110 genes, *red*) genes in SCGB1A1$^+$SCGB3A2^{high} club cells from IPF. *Y-axis*–p values (-log$_{10}$), *x-axis*–the fold-change (log$_e$). Cutoff: shown as dotted lines, fold-change (log$_e$) > 0.25 as the up-regulated genes or <-0.25 as the down-regulated genes, p value<0.05 with Bonferroni correction. **B.** Violin plots of the expression of extracellular matrix-related genes (MMP7, TIMP1 and LAMA5) in SCGB1A1$^+$SCGB3A2^{high} club cells of controls *vs* IPF. **C.** Violin plots in the expression of growth factor-related genes (MDK, SPP1 and C3) in SCGB1A1$^+$SCGB3A2^{high} club cells from controls *vs* IPF. **D.** Violin plots of the expression of mucin-related genes (AGR3, MUC1 and MUC4) in SCGB1A1$^+$SCGB3A2^{high} club cells of controls *vs* IPF. In **B.-D.**, *left*–controls, *right*–IPF.

https://doi.org/10.1371/journal.pone.0237529.g004

immune response, antigen processing and presentation, interferon-gamma signaling pathway, cellular protein metabolic process, and T cell co-stimulation (S4B Fig in S1 File), suggesting that the SCGB3A2^{high} club cells of IPF patients were highly associated with immune cell maturation. In addition, expression of many genes related to the functions of extracellular matrix formation (MMP7, TIMP1 and LAMA5) and growth factors (MDK, SPP1 and C3) were significantly up-regulated in the SCGB3A2^{high} club cells in IPF (Fig 4B and 4C). The SCGB3A2^{high} club cells in IPF also showed a mucous cell-like phenotype, which had increased expression of AGR3, MUC1 and MUC4 (Fig 4D).

SCGB3A2^{+} alveolar cells in IPF

Lastly, we identified a subpopulation of alveolar cells mainly contributed by one IPF individual, only presented in IPF and not in the control donors (Fig 2A, S2 Fig in S1 File). This alveolar cell population expressed mild level of type I alveolar cell markers AGER and HOPX and high level of type II alveolar cell markers SFTPC and ABCA3 (S3A Fig in S1 File), and also had some characteristics of club cells, by highly expressing SCGB3A2 (Fig 2B and 2C). Interestingly, the SCGB3A2^{+} alveolar cells highly expressed collagen (COL1A1) and fibronectin (FN1) genes (Fig 5A and 5B), suggesting that, in addition to myofibroblasts, the SCGB3A2^{+} hyperplastic alveolar cells also contributed to the extracellular matrix deposition in IPF.

Cell lineage of the club cells in IPF

To identify the ontology of the club cell populations in IPF patients, we applied a cell lineage pseudotime analysis on the IPF single cells in Fig 2. As BC are the stem/progenitor cells for human club cells [3], two different lineages were identified. A single lineage started from BC and gave rise to SCGB3A2^{high} club cells. At that point, the path bifurcated leading to either MUC5B+ club cells or SCGB3A2^{+} alveolar cells leading to intermediate alveolar cells and type I alveolar cells as the destination (S5 Fig in S1 File).

Discussion

IPF is a progressive, chronic fibrotic lung disorder of unknown cause and a median survival of 2 to 4 years since diagnosis [27, 28]. To determine the roles of club cell populations, the major secretory cells in human small airways, in the pathogenesis of IPF, we assessed the single cell transcriptome data from human lung of control donors and IPF patients. Two unique club cell subpopulations were identified in the human lung, highly expressing SCGB1A1 and MUC5B, and SCGB1A1 and SCGB3A2, respectively. The SCGB1A1^{+} MUC5B^{+} club cell is an IPF-dependent cell population, highly expresses genes related to mucin and chemoattractant cytokines to immune cells. In contrast, the SCGB1A1^{+} SCGB3A2^{high} club cell subpopulation presents in equal proportion in IPF and controls. The transcriptome of this club cell subpopulation is significantly altered in IPF patients, with increased expression of mucins, cytokines and extracellular matrix genes.

The pathogenesis of IPF is associated with loss of alveolar architecture, airway epithelial remodeling, activation of fibroblasts, accumulation of extracellular matrix and chronic inflammation [19–28]. Originally observed 4 decades ago [19], there is increased evidence suggesting that "bronchiolization", characterized by ectopic appearance of airway epithelial cells in the alveolar space, is a characteristic feature in the IPF lung, likely playing a crucial role in IPF progression [19–24]. Basal cells (BC), the stem/progenitor cells in the human airway epithelium, are normally absent in the normal human alveoli. In contrast, an increased number of BC are present in the remodeled alveolar epithelium of IPF patients [20, 22]. A group of BC expressed genes is associated with the clinical outcome and mortality of IPF patients, consistent with an
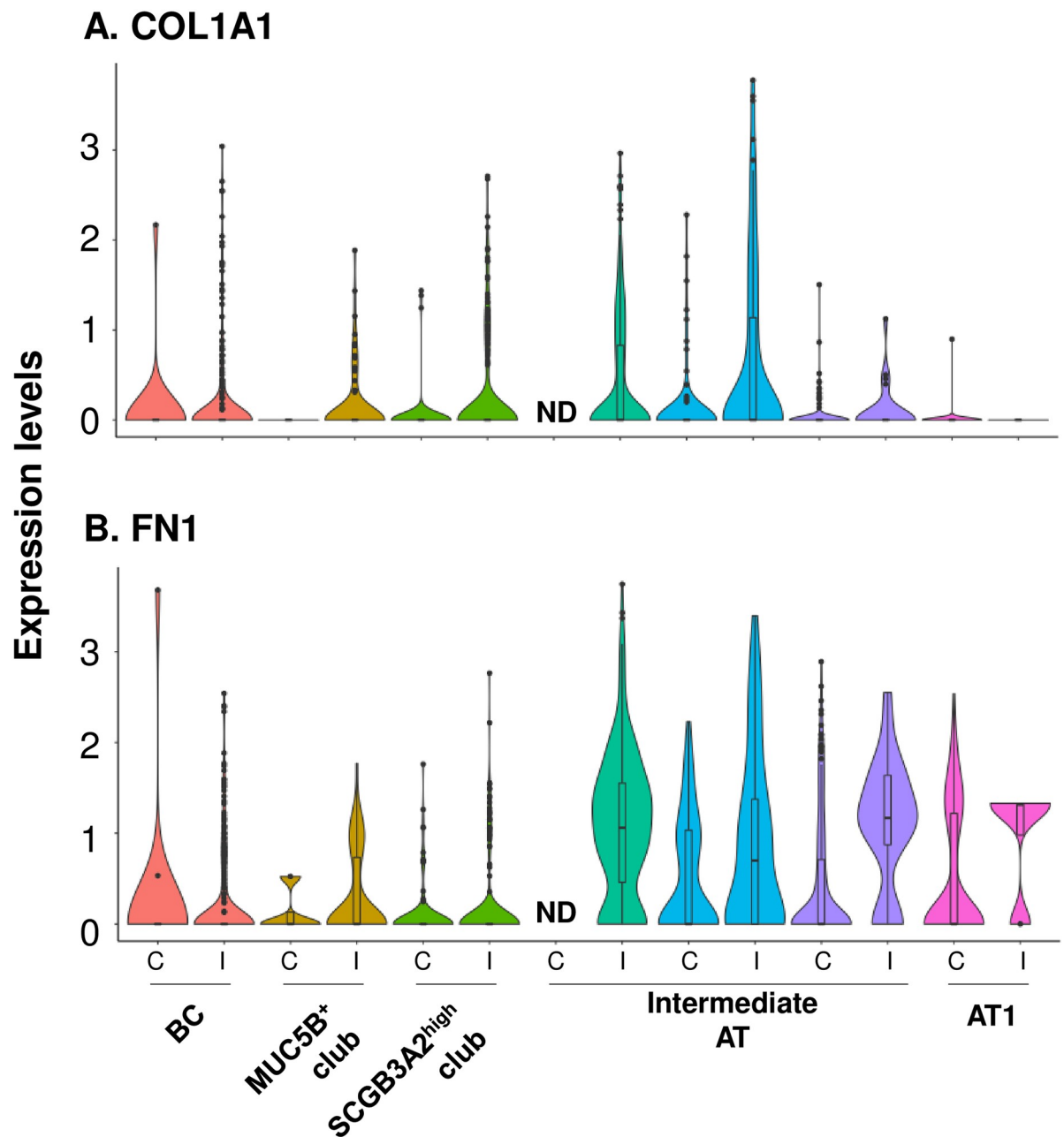
A. COL1A1

B. FN1

Fig 5. Expression of extracellular matrix genes in SCGB3A2⁺ alveolar cells in populations of lung epithelial cells. Violin plots showing the expression of extracellular matrix genes SCGB1A1 *(top)*, SCGB3A2 *(middle)*, and IPF genetic risk factor gene MUC5B *(bottom)* in the cell populations from control (C) and IPF (I) donors identified in Fig 2A. **A**. SCGB1A1. **B**. FN1. Note that sub-cluster 4 was not detected (ND) in control cell populations.

important role of airway basal cells in IPF [22]. Metaplasia of goblet cells, a differentiated product of BC, also plays a role in "bronchiolization". Immuno-stainings and RNA-sequencing have identified extensive goblet cell metaplasia in IPF, and many of the goblet cells co-express alveolar type I and II cell markers [21, 23, 24].

There are several lines of evidence that indicate a potential role of club cells in the pathogenesis of IPF. Club cells, another differentiation product of BC, are the major secretory and host defense cell population in human small airway epithelium. In contrast to the decrease of club cell numbers in COPD and asthma, SCGB1A1, the major secretory product of club cells, is increased in both serum and bronchoalveolar lavage (BAL) fluid of IPF patients [10]. In mouse, depletion of club cells by naphthalene attenuates bleomycin-induced lung injury and fibrosis [11, 12]. In contrast, over-expression of SCGB3A2, a club cell marker with heterogeneous expression [14, 15], in mouse lung accelerates resolution of bleomycin-induced pulmonary fibrosis [11, 12].

By assessing the transcriptomes of the single cells from human lung, we were able to evaluate the cellular and transcriptomic heterogeneities of club cells in IPF and controls from the study of Reyfman et al [13]. First, we identified a SCGB1A1[+] club cell sub-population highly expressing MUC5B. This is of significant interest in the context that MUC5B promoter polymorphism is a significant genetic risk factor strongly associated with IPF [16]. Overexpression of MUC5B induces impaired mucociliary clearance and enhances lung fibrosis in mice [16]. The single cell data analysis showed that the proportions of SCGB1A1[+] MUC5B[+] club cells was significantly increased in the IPF lung, consistent with observation that the MUC5B expression is increased in IPF [29]. The single cell analysis permitted identification of the molecular phenotype and function of the SCGB1A1[+] MUC5B[+] club cells in IPF. We found that these MUC5B[+] club cells expressed high levels of mucin-related genes such as SPDEF, TFF3 and AGR2, suggesting the MUC5B[+] club cells are a goblet cell-like subpopulation [30, 31]. Moreover, the SCGB1A1[+] MUC5B[+] club cells expressed a number of chemoattractant cytokines, including the neutrophil chemoattractant CXCL1, 6 and 8 [32–34], and CX3CL1, a potent chemoattractant for T cells and monocytes [35], suggesting that this club cell subpopulation could interact with immune cells.

In regard to the ontology of the MUC5B[+] club cells migrating to the alveoli in IPF, our previous study demonstrated that MUC5B is highly expressed in normal human small airway club cells [3], i.e., SCGB1A1[+] MUC5B[+] club cells in broncholarization of the IPF lung might be the result of migration of club cells from human small airway epithelium [20, 36]. Another possibility might be the small airway BC, since: (1) human small airway BC can differentiate into SCGB1A1[+] MUC5B[+] club cells [3]; and (2) there are increased numbers of BC in the lower respiratory tract of the IPF lung [20, 22]. The cell lineage analysis support this hypothesis that the BC generate the SCGB3A2[high] club cells first, to further differentiated into MUC5B + club cells.

SCGB3A2 is a another marker with heterogeneous expression in club cells [14, 15]. There is a SCGB3A2[high] club cell subpopulation in the murine airways that highly express UPK3A, but reduced expression of SCGB1A1 [14, 15]. This cell population contributes to the generation of club and ciliated cells in the adult lung [14, 15]. We also identified a club cell sub-population in human lung that highly express SCGB3A2, with less expression of SCGB1A1 compared to the MUC5B[+] club cells. Although the proportion of the SCGB1A1[+] SCGB3A2[high] club cells is not different, the molecular functions of this club cell population is significantly altered in the IPF lung. In IPF, the SCGB1A1[+] SCGB3A2[high] club cells may directly contribute to the extracellular matrix (ECM) accumulation by increased expression of matrix metalloproteinase (MMP) and laminin. The high expression of inhibitor of MMP in the SCGB3A2[high] club cells may contribute to the complex regulation of ECM in IPF. In addition, SCGB3A2[high] club cells in IPF expressed many fibrotic growth factors, such as C3, MDK and SPP1 [37–39]. Interestingly, like the MUC5B[+] club cells, the SCGB3A2[high] club cells in the IPF lung also express many mucin-related genes, but with a different repertoire, suggesting the heterogeneity of goblet cell-like club cells in the IPF.

In a recent preprint paper from Haberman et al. [40], single-cell data from nonfibrotic controls and pulmonary fibrosis samples, including IPF, chronic hypersensitivity pneumonitis, nonspecific interstitial pneumonia, sarcoidosis and unclassifiable interstitial lung disease, were evaluated to identify four secretory cell phenotypes: $SCGB1A1^+MUC5B^+MUC5AC^+$ cells; $SCGB1A1^+MUC5B^+$ cells; $SCGB1A1^+SCGB3A2^+$ cells and $SCGB3A2^+$ only cells. The proportion of $SCGB3A2^+$ only cells was increased and widespread in the cystic and fibrotic lesions of pulmonary fibrosis tissue sections. In our current study which only focused on IPF, we found that the $SCGB1A1^+SCGB3A2^+$ cells were presented in both control and IPF, and $SCGB1A1^{+-}MUC5B^+$ cells were mostly contributed by the IPF patients. However, the $SCGB1A1^{+-}MUC5B^+MUC5AC^+$ and $SCGB3A2^+$ only cells were not detected in our study, suggesting these two secretory cell populations might occur in other types of non-IPF pulmonary fibrosis than IPF. In addition, consistent with the findings by Haberman et al [40], we also identified an alveolar cell population highly expressing SCGB3A2. These cells likely contribute to the pathogenesis of IPF by high expression of extracellular matrix-related genes. In conclusion, the single cell transcriptomic analysis reveals the cellular and molecular heterogeneity of club cells, and provide novel insights into the biological functions of club cells in the pathogenesis of IPF.

Supporting information

S1 File.
(PDF)

Acknowledgments

Author Contributions

Conceptualization: Wu-Lin Zuo, Jason G. Mezey, Sudha Visvanathan, Jay S. Fine, Matthew J. Thomas, Ronald G. Crystal.

Data curation: Wu-Lin Zuo, Mahboubeh R. Rostami, Michelle LeBlanc, Robert J. Kaner, Sarah L. O'Beirne.

Formal analysis: Wu-Lin Zuo, Mahboubeh R. Rostami, Michelle LeBlanc, Jason G. Mezey, Karsten Quast, Sudha Visvanathan, Jay S. Fine, Matthew J. Thomas, Ronald G. Crystal.

Funding acquisition: Ronald G. Crystal.

Investigation: Wu-Lin Zuo, Robert J. Kaner, Sarah L. O'Beirne, Sudha Visvanathan, Ronald G. Crystal.

Methodology: Mahboubeh R. Rostami, Jason G. Mezey, Matthew J. Thomas, Ronald G. Crystal.

Project administration: Philip L. Leopold, Ronald G. Crystal.

Resources: Robert J. Kaner, Sarah L. O'Beirne, Ronald G. Crystal.

Software: Mahboubeh R. Rostami, Jason G. Mezey.

Supervision: Robert J. Kaner, Sudha Visvanathan, Jay S. Fine, Matthew J. Thomas, Ronald G. Crystal.

Validation: Mahboubeh R. Rostami.

Visualization: Philip L. Leopold.

Writing – original draft: Wu-Lin Zuo, Mahboubeh R. Rostami, Ronald G. Crystal.

Writing – review & editing: Wu-Lin Zuo, Mahboubeh R. Rostami, Michelle LeBlanc, Robert J. Kaner, Sarah L. O'Beirne, Jason G. Mezey, Philip L. Leopold, Karsten Quast, Sudha Visvanathan, Jay S. Fine, Matthew J. Thomas, Ronald G. Crystal.

References

1. Lumsden AB, McLean A, Lamb D. Goblet and Clara cells of human distal airways: evidence for smoking induced changes in their numbers. Thorax. 1984; 39: 844–849 https://doi.org/10.1136/thx.39.11.844 PMID: 6505991

2. Boers JE, Ambergen AW, Thunnissen FB. Number and proliferation of clara cells in normal human airway epithelium. Am J Respir Crit Care Med. 1999; 159: 1585–1591 https://doi.org/10.1164/ajrccm.159.5.9806044 PMID: 10228131

3. Zuo WL, Shenoy SA, Li S, O'Beirne SL, Strulovici-Barel Y, Leopold PL, et al. Ontogeny and Biology of Human Small Airway Epithelial Club Cells. Am J Respir Crit Care Med. 2018; 198: 1375–1388 https://doi.org/10.1164/rccm.201710-2107OC PMID: 29874100

4. Singh G, Katyal SL. Clara cells and Clara cell 10 kD protein (CC10). Am J Respir Cell Mol Biol. 1997; 17: 141–143 https://doi.org/10.1165/ajrcmb.17.2.f138 PMID: 9271300

5. Tokita E, Tanabe T, Asano K, Suzaki H, Rubin BK. Club cell 10-kDa protein attenuates airway mucus hypersecretion and inflammation. Eur Respir J. 2014; 44: 1002–1010 https://doi.org/10.1183/09031936.00080913 PMID: 24833761

6. Gamez AS, Gras D, Petit A, Knabe L, Molinari N, Vachier I, et al. Supplementing defect in club cell secretory protein attenuates airway inflammation in COPD. Chest. 2015; 147: 1467–1476 https://doi.org/10.1378/chest.14-1174 PMID: 25474370

7. Laucho-Contreras ME, Polverino F, Gupta K, Taylor KL, Kelly E, Pinto-Plata V, et al. Protective role for club cell secretory protein-16 (CC16) in the development of COPD. Eur Respir J. 2015; 45: 1544–1556 https://doi.org/10.1183/09031936.00134214 PMID: 25700379

8. Van Vyve T, Chanez P, Bernard A, Bousquet J, Godard P, Lauwerijs R, et al. Protein content in bronchoalveolar lavage fluid of patients with asthma and control subjects. J Allergy Clin Immunol. 1995; 95: 60–68 https://doi.org/10.1016/s0091-6749(95)70153-2 PMID: 7822665

9. Guerra S, Vasquez MM, Spangenberg A, Halonen M, Martin RJ. Club cell secretory protein in serum and bronchoalveolar lavage of patients with asthma. J Allergy Clin Immunol. 2016; 138: 932–934 e931 https://doi.org/10.1016/j.jaci.2016.03.047 PMID: 27315766

10. Buendia-Roldan I, Ruiz V, Sierra P, Montes E, Ramirez R, Vega A, et al. Increased Expression of CC16 in Patients with Idiopathic Pulmonary Fibrosis. PLoS One. 2016; 11: e0168552 https://doi.org/10.1371/journal.pone.0168552 PMID: 27977812

11. Cai Y, Yoneda M, Tomita T, Kurotani R, Okamoto M, Kido T, et al. Transgenically-expressed secretoglobin 3A2 accelerates resolution of bleomycin-induced pulmonary fibrosis in mice. BMC Pulm Med. 2015; 15: 72 https://doi.org/10.1186/s12890-015-0065-4 PMID: 26178733

12. Yokoyama T, Yanagihara T, Suzuki K, Hamada N, Tsubouchi K, Ogata-Suetsugu S, et al. Depletion of club cells attenuates bleomycin-induced lung injury and fibrosis in mice. J Inflamm (Lond). 2017; 14: 20

13. Reyfman PA, Walter JM, Joshi N, Anekalla KR, McQuattie-Pimentel AC, Chiu S, et al. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. Am J Respir Crit Care Med. 2019; 199: 1517–1536 https://doi.org/10.1164/rccm.201712-2410OC PMID: 30554520

14. Guha A, Vasconcelos M, Cai Y, Yoneda M, Hinds A, Qian J, et al. Neuroepithelial body microenvironment is a niche for a distinct subset of Clara-like precursors in the developing airways. Proc Natl Acad Sci U S A. 2012; 109: 12592–12597 https://doi.org/10.1073/pnas.1204710109 PMID: 22797898

15. Kotton DN, Morrisey EE. Lung regeneration: mechanisms, applications and emerging stem cell populations. Nat Med. 2014; 20: 822–832 https://doi.org/10.1038/nm.3642 PMID: 25100528

16. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. N Engl J Med. 2011; 364: 1503–1512 https://doi.org/10.1056/NEJMoa1013660 PMID: 21506741

17. 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 https://doi.org/10.1038/nbt.4096 PMID: 29608179

18. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics. 2018; 19: 477 https://doi.org/10.1186/s12864-018-4772-0 PMID: 29914354

19. Fulmer JD, Roberts WC, von Gal ER, Crystal RG. Small airways in idiopathic pulmonary fibrosis. Comparison of morphologic and physiologic observations. J Clin Invest. 1977; 60: 595–610 https://doi.org/10.1172/JCI108811 PMID: 893665

20. Kawanami O, Ferrans VJ, Crystal RG. Structure of alveolar epithelial cells in patients with fibrotic lung disorders. Lab Invest. 1982; 46: 39–53 PMID: 7054589

21. Plantier L, Crestani B, Wert SE, Dehoux M, Zweytick B, Guenther A, et al. Ectopic respiratory epithelial cell differentiation in bronchiolised distal airspaces in idiopathic pulmonary fibrosis. Thorax. 2011; 66: 651–657 https://doi.org/10.1136/thx.2010.151555 PMID: 21422041

22. Prasse A, Binder H, Schupp JC, Kayser G, Bargagli E, Jaeger B, et al. BAL Cell gene expression is indicative of outcome and airway basal cell involvement in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2019; 199: 622–630 https://doi.org/10.1164/rccm.201712-2551OC PMID: 30141961

23. Whitsett JA. Airway Epithelial Differentiation and Mucociliary Clearance. Ann Am Thorac Soc. 2018; 15: S143–S148 https://doi.org/10.1513/AnnalsATS.201802-128AW PMID: 30431340

24. Xu Y, Mizuno T, Sridharan A, Du Y, Guo M, Tang J, et al. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. JCI Insight. 2016; 1: e90558 https://doi.org/10.1172/jci.insight.90558 PMID: 27942595

25. Kasper M, Barth K. Potential contribution of alveolar epithelial type I cells to pulmonary fibrosis. Biosci Rep. 2017;37:

26. Knudsen L, Ruppert C, Ochs M. Tissue remodelling in pulmonary fibrosis. Cell Tissue Res. 2017; 367: 607–626 https://doi.org/10.1007/s00441-016-2543-2 PMID: 27981380

27. Crystal RG, Bitterman PB, Rennard SI, Hance AJ, Keogh BA. Interstitial lung diseases of unknown cause. Disorders characterized by chronic inflammation of the lower respiratory tract (first of two parts). N Engl J Med. 1984; 310: 154–166 https://doi.org/10.1056/NEJM198401193100304 PMID: 6361560

28. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. Lancet. 2017; 389: 1941–1952 https://doi.org/10.1016/S0140-6736(17)30866-8 PMID: 28365056

29. Hancock LA, Hennessy CE, Solomon GM, Dobrinskikh E, Estrella A, Hara N, et al. Muc5b overexpression causes mucociliary dysfunction and enhances lung fibrosis in mice. Nat Commun. 2018; 9: 5363 https://doi.org/10.1038/s41467-018-07768-9 PMID: 30560893

30. Rajavelu P, Chen G, Xu Y, Kitzmiller JA, Korfhagen TR, Whitsett JA. Airway epithelial SPDEF integrates goblet cell differentiation and pulmonary Th2 inflammation. J Clin Invest. 2015; 125: 2021–2031 https://doi.org/10.1172/JCI79422 PMID: 25866971

31. Wiede A, Jagla W, Welte T, Kohnlein T, Busk H, Hoffmann W. Localization of TFF3, a new mucus-associated peptide of the human respiratory tract. Am J Respir Crit Care Med. 1999; 159: 1330–1335 https://doi.org/10.1164/ajrccm.159.4.9804149 PMID: 10194185

32. De Filippo K, Dudeck A, Hasenberg M, Nye E, van Rooijen N, Hartmann K, et al. Mast cell and macrophage chemokines CXCL1/CXCL2 control the early stage of neutrophil recruitment during tissue inflammation. Blood. 2013; 121: 4930–4937 https://doi.org/10.1182/blood-2013-02-486217 PMID: 23645836

33. de Oliveira S, Reyes-Aldasoro CC, Candel S, Renshaw SA, Mulero V, Calado A. Cxcl8 (IL-8) mediates neutrophil recruitment and behavior in the zebrafish inflammatory response. J Immunol. 2013; 190: 4349–4359 https://doi.org/10.4049/jimmunol.1203266 PMID: 23509368

34. Jovic S, Linge HM, Shikhagaie MM, Olin AI, Lannefors L, Erjefalt JS, et al. The neutrophil-recruiting chemokine GCP-2/CXCL6 is expressed in cystic fibrosis airways and retains its functional properties after binding to extracellular DNA. Mucosal Immunol. 2016; 9: 112–123 https://doi.org/10.1038/mi.2015.43 PMID: 25993443

35. Gevrey JC, Isaac BM, Cox D. Syk is required for monocyte/macrophage chemotaxis to CX3CL1 (Fractalkine). J Immunol. 2005; 175: 3737–3745 https://doi.org/10.4049/jimmunol.175.6.3737 PMID: 16148119

36. Fukumoto J, Soundararajan R, Leung J, Cox R, Mahendrasah S, Muthavarapu N, et al. The role of club cell phenoconversion and migration in idiopathic pulmonary fibrosis. Aging (Albany NY). 2016; 8: 3091–3109

37. Misa K, Tanino Y, Wang X, Nikaido T, Kikuchi M, Sato Y, et al. Involvement of midkine in the development of pulmonary fibrosis. Physiol Rep. 2017; 5:

38. Okamoto T, Mathai SK, Hennessy CE, Hancock LA, Walts AD, Stefanski AL, et al. The relationship between complement C3 expression and the MUC5B genotype in pulmonary fibrosis. Am J Physiol Lung Cell Mol Physiol. 2018; 315: L1–L10 https://doi.org/10.1152/ajplung.00395.2017 PMID: 29565179

39. Pardo A, Gibson K, Cisneros J, Richards TJ, Yang Y, Becerril C, et al. Up-regulation and profibrotic role of osteopontin in human idiopathic pulmonary fibrosis. PLoS Med. 2005; 2: e251 https://doi.org/10.1371/journal.pmed.0020251 PMID: 16128620

40. Habermann AC, Gutierrez AJ, Bui LT, Yahn SL, Winters NI, Calvi CL, et al. Single-cell RNA-sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. bioRxiv. 2019: 753806