ORIGINAL ARTICLE

BAL Cell Gene Expression Is Indicative of Outcome and Airway Basal Cell Involvement in Idiopathic Pulmonary Fibrosis

Antje Prasse[1,2,3,4], Harald Binder[5], Jonas C. Schupp[2], Gian Kayser[6], Elena Bargagli[7], Benedikt Jaeger[3,4], Moritz Hess[5], Susanne Rittinghausen[3], Louis Vuga[8], Heather Lynn[9], Shelia Violette[10], Birgit Jung[11], Karsten Quast[11], Bart Vanaudenaerde[12], Yan Xu[13], Jens M. Hohlfeld[3,4], Norbert Krug[3,4], Jose D. Herazo-Maya[9], Paola Rottoli[7], Wim A. Wuyts[12], and Naftali Kaminski[9]

[1]Department of Pulmonology, Hannover Medical School, Hannover, Germany; [2]Department of Pneumology and [6]Institute of Surgical Pathology, University Medical Center, Freiburg, Germany; [3]Fraunhofer Institute for Toxicology and Experimental Medicine, Hannover, Germany; [4]German Center for Lung Research, BREATH, Hannover, Germany; [5]Institute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany; [7]Respiratory Diseases and Lung Transplantation Unit, AOUS and Siena University, Siena, Italy; [8]Division of Respiratory Medicine, Simmons Center, University of Pittsburgh, Pittsburgh, Pennsylvania; [9]Section of Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, Connecticut; [10]Biogen Idec Inc., Cambridge, Massachusetts; [11]Boehringer Ingelheim Pharma GmbH & Co. KG Research, Biberach, Germany; [12]Department of Respiratory Medicine, University Hospitals Leuven, Leuven, Belgium; and [13]Division of Neonatology, Perinatal and Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio

Abstract

Rationale: Idiopathic pulmonary fibrosis (IPF) is a fatal disease with a variable and unpredictable course.

Objectives: To determine whether BAL cell gene expression is predictive of survival in IPF.

Methods: This retrospective study analyzed the BAL transcriptome of three independent IPF cohorts: Freiburg (Germany), Siena (Italy), and Leuven (Belgium) including 212 patients. BAL cells from 20 healthy volunteers, 26 patients with sarcoidosis stage III and IV, and 29 patients with chronic obstructive pulmonary disease were used as control subjects. Survival analysis was performed by Cox models and component-wise boosting. Presence of airway basal cells was tested by immunohistochemistry and flow cytometry.

Measurements and Main Results: A total of 1,582 genes were predictive of mortality in the IPF derivation cohort in univariate analyses adjusted for age and sex at false discovery rate less than

0.05. A nine-gene signature, derived from the discovery cohort (Freiburg), performed well in both replication cohorts, Siena ($P <$ 0.0032) and Leuven ($P =$ 0.0033). nCounter expression analysis confirmed the array results ($P <$ 0.0001). The genes associated with mortality in BAL cells were significantly enriched for genes expressed in airway basal cells. Further analyses by gene expression, flow cytometry, and immunohistochemistry showed an increase in airway basal cells in BAL and tissues of IPF compared with control subjects, but not in chronic obstructive pulmonary disease or sarcoidosis.

Conclusions: Our results identify and validate a BAL signature that predicts mortality in IPF and improves the accuracy of outcome prediction based on clinical parameters. The BAL signature associated with mortality unmasks a potential role for airway basal cells in IPF.

Keywords: IPF; transcriptome; BAL; airway basal cells; biomarker

At a Glance Commentary

Scientific Knowledge on the Subject: Although multiple peripheral blood-derived biomarkers have been shown to predict mortality in idiopathic pulmonary fibrosis (IPF), there is no information about which alveolar molecular changes are indicative of disease progression and outcome. To test this, we studied the transcriptome of BAL cells. Airway basal cells are the progenitor cells of the airway epithelium, and recent murine studies suggested a pathogenic role of these cells in IPF. Until now, these cells were not known to be present in BAL or indicative of high mortality in IPF.

What This Study Adds to the Field: Using gene expression microarrays we identified transcripts of BAL cells that were associated with mortality in patients with IPF. A nine-gene expression signature that predicted mortality in patients with IPF from a discovery cohort performed well in two replication cohorts and could be used in the future as a new biomarker. The obtained gene expression data allow insights into which tissue-derived signal pathways are associated with mortality in IPF. Unexpectedly, the genes predictive of mortality were significantly enriched for genes expressed in airway basal cells. This led to the discovery that these cells are indeed increased in the BAL and tissue of patients with IPF.

Idiopathic pulmonary fibrosis (IPF) is a fatal disease with an estimated median survival time of 3 years and a variable course (1, 2). Although there has been significant progress in predicting outcome in IPF using clinical disease staging systems (2, 3), peripheral blood biomarkers (4–6), and gene variants (7), very little is known whether molecular events in the lung milieu are predictive of outcome in IPF.

Ideally, research aiming to identify molecular markers in patients with IPF would focus on lung biopsies obtained at diagnosis. However, lung biopsies are invasive and indicated in less than 30% of patients and thus are limited in their utility. BAL, a procedure in which saline is injected and then collected through a bronchoscope, samples the cells residing on the external layer of the alveolus (8). Although BAL is not absolutely required for the diagnosis of IPF, it can be used in the routine diagnostic work-up, most often to exclude infections and other inflammatory conditions commonly associated with an usual interstitial pneumonia (UIP) pattern on high-resolution computed tomography (9). Multiple studies suggested that the alveolar molecular environment is altered in IPF (10, 11).

Considering that the cells obtained in BAL reside in the alveolar compartment we hypothesized that changes in gene expression of BAL cells will be predictive of IPF outcome. Some of the results of these studies have been previously reported in the form of an abstract (12).

Methods

A comprehensive description of all methods is available in the online supplement.

Study Population

BAL was obtained from 212 patients with IPF in three independent cohorts obtained at three referral centers: Freiburg (Germany), derivation; Siena (Italy) and Leuven (Belgium), replication (Table 1; see Figure E1 in the online supplement). IPF diagnosis was established by a multidisciplinary board at each institution according to the American Thoracic Society/European Respiratory Society criteria (13) and later determined to be consistent with recent guidelines (14, 15). BAL cells from 20 healthy donors served as control for IPF (see Table E1). For active disease controls we used gene expression data of two additional studies testing BAL cells from 26 patients with sarcoidosis stage III and IV versus 20 healthy volunteers and BAL cells from 29 currently smoking patients with chronic obstructive pulmonary disease (COPD) versus 28 currently smoking control subjects (see Tables E2 and E3). All studies were approved by local ethics committees. In all three centers pulmonary function tests were routinely performed with a standard methodology, according to the American Thoracic Society/European Respiratory Society recommendations using a body-plethysmograph. Survival status was obtained from follow-up visits and telephone interviews. Patients who had not been seen within 3 months were called to confirm their vitality. Ten patients underwent a lung transplant. None of the patients received pirfenidone or nintedanib before BAL examination; however, during follow-up, patients were treated with varied treatment regimens including corticosteroids, azathioprine, *N*-acetylcysteine, or pirfenidone. All patients were white except one patient of the Leuven cohort. For further details, *see* Table 1 and the online supplement.

Microarray

For detailed description *see* the online supplement and our recent publication (5). Gene expression was detected using Whole Human Genome arrays. All MIAME compliant raw data have been deposited in the Gene Expression Omnibus with the accession GSE70867.

nCounter Expression Analysis

To validate the microarray data of IPF samples we applied multiplexed, color-coded probe pairs using the nCounter expression analysis system (Nanostring). Detailed information is given in the online supplement.

Determining the Gene Expression Profile of Airway Basal Cells

Airway basal cells (ABCs) were isolated from bronchial brushes of subsegmental bronchi of the right lower lobe as recently described (16). Detailed information regarding cell isolation and gene expression profiling is given in the online supplement.

Immunocytology and Immunohistochemistry

Cell smears of BALs from 20 patients with IPF, 20 patients with sarcoidosis, and 10 healthy volunteers were evaluated by immune-cytology. Immunohistochemistry of lung tissues from 15 patients with IPF (seven wedge biopsies and eight explants), three patients with sarcoidosis, and three healthy lung donors (transplants) was performed. A monoclonal mouse-antihuman cytokeratin-5/6 antibody (DAKO clone D5/16 B4) and polyclonal rabbit antihuman ΔNP63 antibody (Calbiochem PC373) were used.

Gender Age Physiology Index

The Gender Age Physiology (GAP) index, a staging system for patients with IPF, was calculated as recently described using age, sex, FVC, and $D_{L_{CO}}$ (2).

Table 1. Baseline Characteristics of Patients with IPF

Characteristic	All IPF (n = 176)	Freiburg (n = 62)	Leuven (n = 64)	Siena (n = 50)	P Value*
Age, yr	68.1 ± 9.5	67.4 ± 9.1	68.2 ± 8.5	68.7 ± 11.2	0.768
Male sex, %	82	85	80	80	0.648
FVC % predicted value, %	71 ± 21	66 ± 20	78 ± 18	67 ± 23	<0.001
$D_{L_{CO}}$					
Percent predicted value, %	43 ± 14	44 ± 16	45 ± 12	40 ± 15	0.396
Could not perform $D_{L_{CO}}$, n	20	7	1	12	
Deaths, n (%)	100 (57)	45 (73)	24 (38)	31 (62)	
Transplants, n (%)	10 (6)	3 (5)	3 (5)	4 (8)	0.704
Median observation time, mo		20	18	16	
Smoking status, %					0.156
Never smoked	33	42	23	34	
Former smoker	64	56	70	64	
Current smoker	3	2	6	2	
HRCT UIP, n (%)					0.208
Definite	134 (76)	43 (69)	53 (83)	38 (76)	
Possible	42 (24)	19 (31)	11 (17)	12 (24)	
HRCT emphysema present, n	24 (14)	5 (8)	16 (25)	3 (6)	
BAL					
Cell count, ×10^6 cells	13.0 ± 7.7	12.1 ± 7.2	13.3 ± 8.8	13.7 ± 6.9	0.518
Alveolar macrophages, %	74 ± 17	71 ± 17	80 ± 17	71 ± 16	0.006
Lymphocytes, %	10 ± 9	11 ± 9	9 ± 10	10 ± 9	0.358
Neutrophils, %	11 ± 14	12 ± 14	9 ± 14	13 ± 14	0.192
Eosinophils, %	4 ± 5	4 ± 5	3 ± 4	5 ± 6	0.128

Definition of abbreviations: HRCT = high-resolution computed tomography; IPF = idiopathic pulmonary fibrosis; UIP = usual interstitial pneumonia.
Data are mean ± SD unless otherwise indicated.
*Group comparison by ANOVA for continuous and chi-square test for categorical characteristics.

Statistical Analysis

In the IPF Freiburg (derivation) cohort, univariate Cox proportional hazards models were calculated for each gene separately adjusted for age and sex using false discovery rate less than 0.05. Median survival was calculated, censoring at the time of transplant.

We used componentwise likelihood-based boosting (17) for developing a multivariable risk prediction signature, considering all genes simultaneously in a Cox model, adjusted for age and sex. The number of boosting steps, which determines the number of genes to be selected for the signature, was chosen by 10-fold cross-validation. For identifying high- and low-risk groups, we considered for each patient and each signature gene whether the expression level of the gene for this patient was above the median expression level of all patients. Genes with risk-increasing effect (i.e., positive coefficient, being above the median) were counted as one point in the risk score, and reversely for genes with risk-decreasing effect. Patients were divided at the median of the resulting risk score, and differences were evaluated by log-rank tests.

Stable selection of genes was performed using cross-validation and resampling. Genes from the signature obtained from the original

data were considered for further validation by nCounter if they were selected in at least 20% of these resamplings. Prediction error curves, which indicate the mean squared error in predicting patient survival in the course of time, were used to evaluate the signature alone, and the signature in addition to the GAP index (2). In addition to prediction accuracy (e.g., as evaluated by receiver operating characteristic curves), these prediction error curves also indicate calibration (18).

Results

Patients' Characteristics

The clinical characteristics of the patients with IPF in the three cohorts are provided in Table 1 and indicate that the cohorts differed slightly. The information about 20 old healthy volunteers and the cohorts of patients with sarcoidosis or COPD and their control subjects is available in Tables E1–E3.

BAL Cell Transcriptome Is Predictive of Mortality in IPF

After adjustment for age and sex, 1,582 genes were associated with mortality (false discovery rate <0.05) (*see* Table E4) in the derivation cohort of patients with IPF. A

multivariate prediction model consisting of nine genes was generated by componentwise likelihood-based boosting in the derivation cohort (Figure 1A; *see* Table E5). Prediction performance on the replication cohorts was evaluated by considering for these nine genes whether the expression level is above or below the median expression level, adding up a risk score accordingly. High- and low-risk groups were determined by splitting this score at the median. This multivariate prediction model performed well in the Leuven cohort (P = 0.0033; c-index, 0.66; CI, 0.55–0.76) (Figure 1B) and the Siena cohort (P = 0.0032; c-index, 0.63; confidence interval [CI], 0.54–0.72) (Figure 1C). Combining the information from the GAP index and the expression signature resulted in better prediction performance than using the GAP index alone (*see* Table E6).

A Stable Six-Gene Signature Confirms Microarray Results and Accurately Predicts Outcome in the Combined Cohort

Because it is well known that risk prediction signatures can be unstable (18), we used a resampling approach for identifying a stable subset of the signature for further validation. Six of the nine genes were selected stably
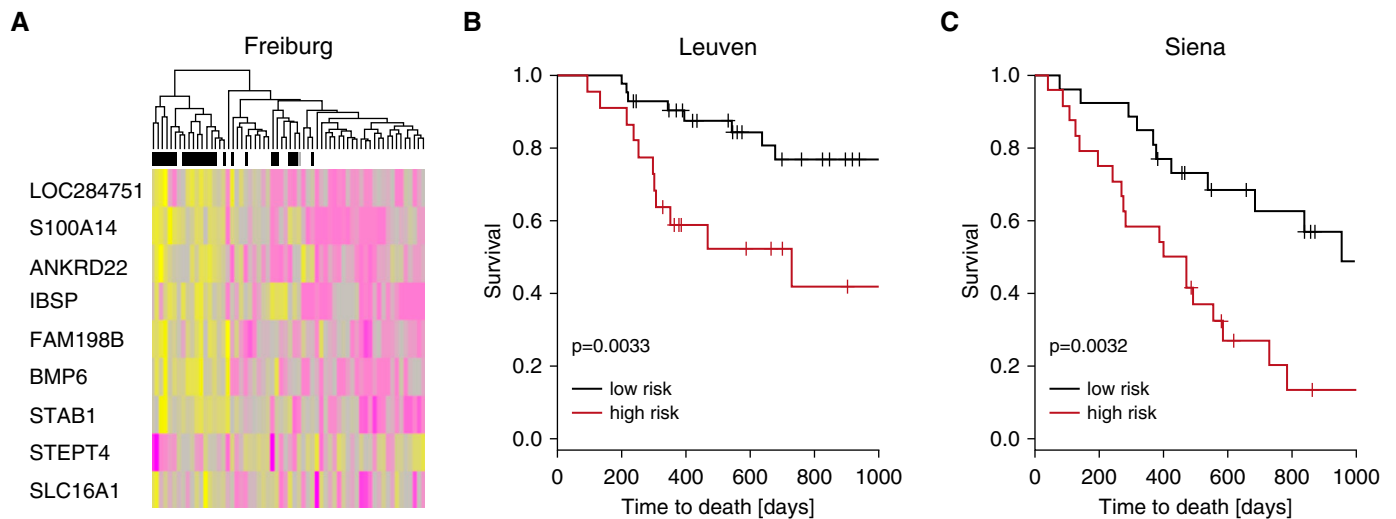
Figure 1. The BAL cell transcriptome is predictive of mortality in idiopathic pulmonary fibrosis. Based on the microarray data from the Freiburg cohort, we found 1,582 genes predictive of mortality. We developed a prediction signature consisting of nine genes by componentwise likelihood-based boosting. (*A*) Expression levels of the nine genes in the Freiburg cohort. Every row represents a gene, and every column, a patient. Yellow denotes increase over the mean of samples, and purple, decrease. Black denotes patients who died within 18 months, gray patients censored before 18 months, and white patients who survived 18 months. The prediction model based on the gene expression data of nine genes was validated in the Leuven cohort (*B*) and in the Siena cohort (*C*) and statistically significantly separated between mortality in both cohorts.

(using selection in at least 20% of the resampling data sets as a cutoff) and were considered for further validation (*see* Tables E4, E5, and E7). Using the nCounter digital expression system we confirmed that this signature is significantly predictive of mortality in the total cohort of 168 patients with IPF ($P < 0.00001$; hazard ratio, 3.951; CI, 2.132–7.323; c-index, 0.67; 95% CI, 0.62–0.71) (Figure 2A). Impressively, this stable six-gene signature alone performed better than the GAP index (Figure 2B) (c-index, 0.63; 95% CI, 0.58–0.69). More importantly, when the six-gene signature and the GAP index were combined, the outcome prediction error rate was reduced, indicating a significant added value (Figure 2B) (c-index, 0.72; 95% CI, 0.66–0.77).

Network Analysis

Network analysis using Ingenuity revealed that the processes associated with mortality in IPF are similar to cancer, organismal injury, and stem cells (reproductive system disease). Processes involved are important for cell proliferation and migration (*see* Table E8).

Enrichment for ABC-derived Genes in the BAL Signature Associated with Mortality

Surprisingly, although 98% of the cells seen in the BAL pellet were of hematopoietic origin, 10% of the genes in the BAL signature were known to be specifically upregulated in cultured ABC (16). We generated an empiric ABC signature by comparing gene expression of isolated ABCs with alveolar macrophages and bronchial epithelial cells (*see* Figure E3). The overlap between our signature and the one recently described by Hackett and colleagues (16) is 38%. We identified 921 genes that were increased in ABCs and listed in the BAL IPF microarray dataset (*see* Table E9). Impressively, 165 of these 921 ABC genes were among the 1,582 genes associated with mortality, indicating significant enrichment ($P < 0.0001$). ABC genes were overexpressed in patients with short survival. Because the signatures were all developed in cultured cells, we also analyzed whether our ABC signature was representative of ABCs *in vivo* using a recently published single cell RNAseq dataset (19). The genes we identified were significantly enriched in ABCs compared with alveolar type II cells ($P < 0.0001$; two tail Fisher exact test).

To investigate whether ABC genes could predict mortality, we derived a survival signature based only on ABC genes, again using componentwise likelihood-based boosting. The resulting model contains 16 ABC genes (*see* Table E10). This multivariate prediction model based on ABC genes performed well in all three cohorts: the Freiburg cohort, ($P < 0.00001$; c-index, 0.73; 95% CI, 0.68–0.77) (Figure 3A), the Leuven cohort ($P = 0.00508$; c-index, 0.67; 95% CI, 0.56–0.76) (Figure 3B), and the Siena cohort ($P = 0.000512$; c-index, 0.66; 95% CI, 0.57–0.74) (Figure 3C). Combining the information from the GAP index and the expression signature resulted in better prediction performance than using the GAP index alone (*see* Table E11).

Presence of ABCs in the BAL of Patients with IPF

To evaluate the presence of ABCs in BAL, we stained BAL cell smears for the expression of ΔNP63 and CK5/6 (cytokeratin 5/6) and CK5 alone (20). We found clusters of CK5/6$^+$ cells and CK5$^+$ cells within the BAL of patients with IPF, consistent with the presence of ABCs in IPF lavages (Figures 4A, 4B, and E4). Ciliated cells did not stain for CK5/6 (Figure 4C). ABCs were rarely observed in BAL of control subjects or patients with sarcoidosis (Figures 4C and 4D).

ABCs Are Increased in BAL of Patients with IPF Compared with Healthy Volunteers but Not in BAL from Patients with Sarcoidosis or COPD

ABC genes were significantly ($P = 0.0156$) enriched in BAL obtained from patients
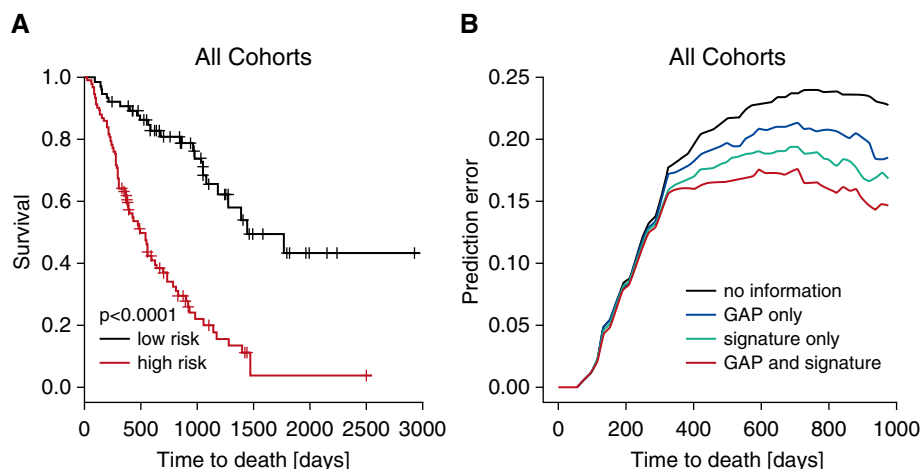
Figure 2. The BAL gene signature is predictive of mortality in idiopathic pulmonary fibrosis. The microarray data were validated by nCounter expression analysis. (*A*) Based on inclusion frequency data from the resampling, we further reduced the signature to six genes. The six-gene signature measured by nCounter expression analysis predicted mortality in all cohorts. (*B*) Prediction error curves (indicating mean squared error in predicting survival status) were calculated for the signature of six genes (green) and Gender Age Physiology (GAP) index (blue). The six-gene signature was superior to the GAP score in predicting mortality in idiopathic pulmonary fibrosis. Combining the six-gene signature with GAP score (red) reduced prediction errors and resulted in better prediction ($P < 0.01$).

with IPF compared with healthy volunteers but not in the BAL of patients with sarcoidosis ($P = 0.3527$) or COPD ($P = 0.1326$). Flow cytometry revealed that 0.26% of BAL cells were EPCAM$^+$ CK5/6$^+$ in IPF (Figures 4E and 4F), compared with 0.05% in age-matched control subjects ($P = 0.001$) and 0.07% in patients with sarcoidosis ($P = 0.008$).

ABCs Are Enriched in the Alveolar Compartment of IPF Tissues

To localize ABCs in the alveolar compartment we stained lung tissues from healthy donor patients with sarcoidosis and patients with IPF for CK5/6 and ΔNP63. In normal and sarcoid tissues, cells expressing CK5/6 and ΔNP63 were found only in the airways, but not in the alveolar compartments (Figures 4G and 4H).

In contrast in IPF, ABCs frequently covered fibroblast foci (Figures 4I–4L), were within fibrotic lesions, and occasionally formed hollow structures (Figure 4M). In some patients, alveolar epithelium was replaced by multiple layers of ABCs resembling ABC hyperplasia and squamous metaplasia (Figures 4L and 4M). In addition, we performed immunohistochemistry for S100A14, a gene that was shown to be highly expressed by ABCs and that was included in the described predictive nine-gene signature of the microarray experiment and the predictive six-gene signature of the nanostring experiment. We found S100A14 highly expressed by ABCs and epithelial cells covering honeycomb cysts and fibroblast foci, but not in normal lung tissues (*see* Figure E5). The pattern of S100A14 staining was very similar to the described CK5/6 and ΔNP63 staining (Figure 4).

Discussion

In this paper we identified BAL gene expression patterns associated with increased mortality in patients with IPF and then developed a gene expression signature predictive of mortality. This signature was derived in one cohort and validated in two additional independent cohorts. Although BAL cells are comprised nearly entirely of hematopoietic cells, the genes associated
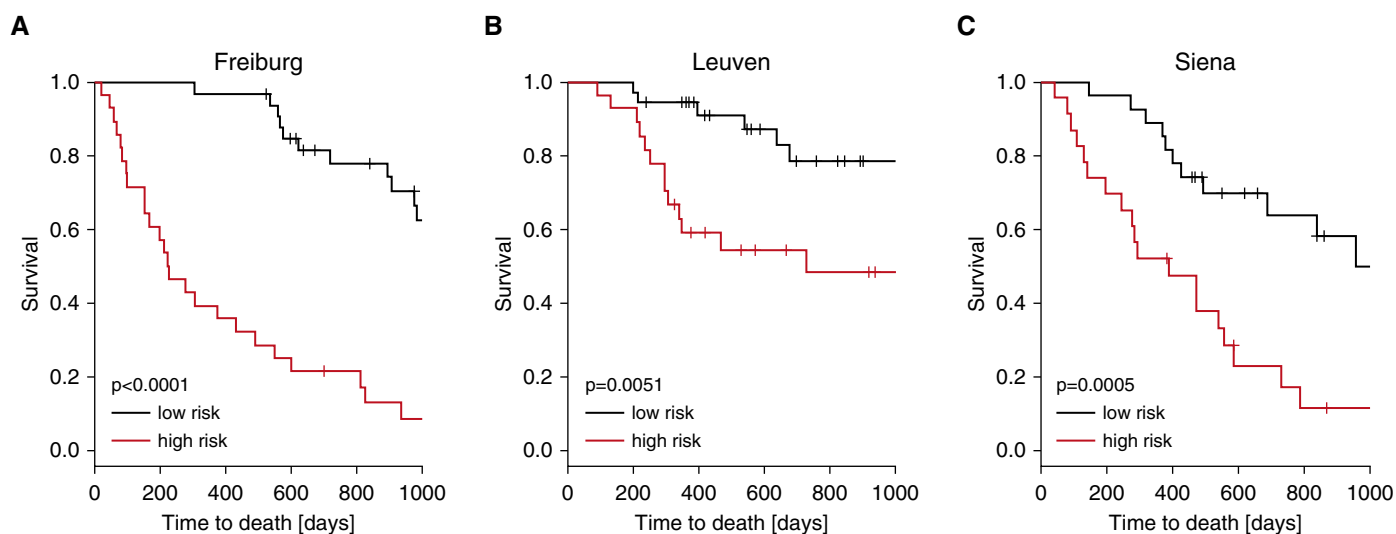

Figure 3. Genes derived from airway basal cells are associated with high mortality in idiopathic pulmonary fibrosis. Based on the gene expression data from the Freiburg cohort, we found 165 genes highly expressed by airway basal cells predictive of mortality. (*A*) By componentwise likelihood-based boosting, we developed a prediction model in the Freiburg cohort. The prediction model based on the gene expression data of airway basal cell genes was validated in the Leuven cohort (*B*) and in the Siena cohort (*C*) and statistically significantly separated between mortality in both cohorts.
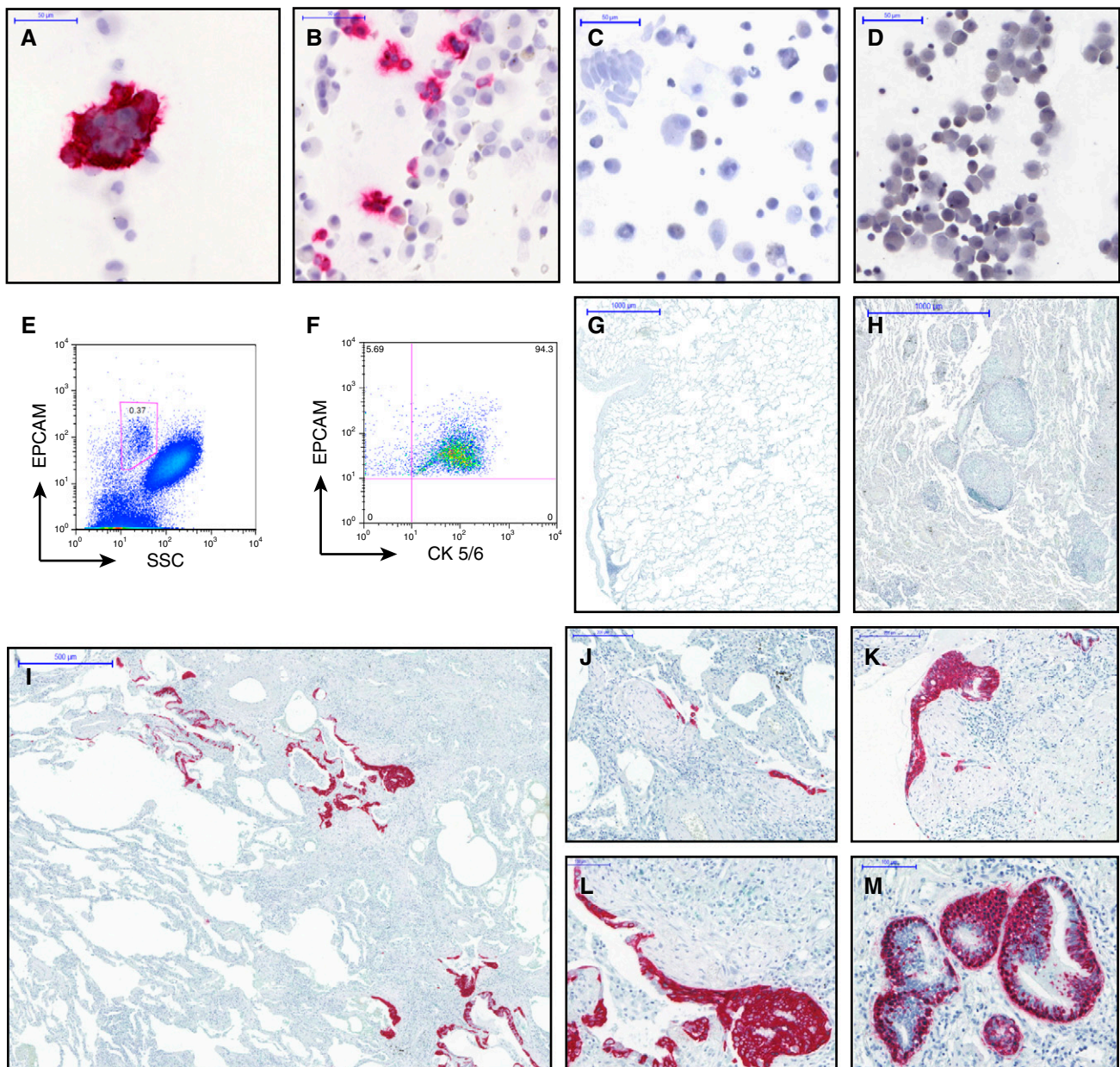
Figure 4. Increase in CK5/6$^+$ ΔNP63$^+$ airway basal cells (ABCs) in BAL and alveolar tissue of patients with idiopathic pulmonary fibrosis (IPF). (*A–D*) CK5/6$^+$ ABC (stained in red) were found frequently in cell smears of BAL from patients with IPF (*A* and *B*) and often formed cell clusters. In contrast, CK5/6$^+$ ABC were found only rarely in the BAL of old healthy volunteers (*C*) and patients with chronic obstructive pulmonary disease (*D*). Flow cytometry revealed presence of EPCAM$^+$ epithelial cells in the BAL of patients with IPF, and most of them coexpressed CK5/6, identifying these cells as ABC (*E* and *F*). Immunohistochemistry of lung tissues stained for CK5/6 (red) and ΔNp63 (nuclear turquois) showed ABCs in the basal layer of airway epithelium but not in the alveolar compartment of normal lung tissue (*G*) and sarcoid tissue (*H*). In contrast, in IPF tissues we observed an enrichment of ABCs within the alveolar compartment (*I–M*). ABCs frequently covered fibroblast foci (*J–L*), and occasionally formed hollow structures (*M*). In some patients, alveolar epithelium was replaced by multiple layers of ABCs consistent with basal cell hyperplasia and squamous metaplasia (*L* and *M*). Scale bars: *A–D*, 50 μm; *G* and *H*, 1,000 μm; *I*, 500 μm; *J* and *K*, 200 μm; *L* and *M*, 100 μm. CK5/6 = cytokeratin 5/6; SSC = side scatter.

with mortality were enriched for genes highly expressed in ABCs. Immunocytology, flow cytometry, and immunohistochemistry confirmed enrichment of ABCs within the BAL and tissues of patients with IPF.

In recent years there has been an increased recognition that molecular and genetic changes may be informative about distinct outcomes in IPF. Most of these studies focused on the peripheral blood and identified proteins or changes in gene expression in peripheral blood mononuclear cells that predict mortality (4–7) but did not directly investigate the alveolar compartment. Gene expression of lung tissues from patients with IPF has been extensively studied (21, 22). However, these studies rarely attempted to correlate gene expression with disease severity, were all hampered by missing mortality data, and did not develop outcome prediction

models. Although BAL is reflective of the alveolar milieu and relatively easy to obtain, very few studies tried to identify molecular biomarkers in BAL. Most recently high BAL neutrophil counts (23) and changes in BAL bacterial burden (24) have been shown to be associated with changed mortality. We used the BAL transcriptome data of patients with IPF from Freiburg as a derivation cohort and found 1,582 genes to be significantly associated with mortality. Based on these data we generated a multivariate prediction model consisting of nine genes that performed well in the independent replication cohorts, despite differences in lavage procedure and patient characteristics. For further validation we tested the six most stably selected genes from the signature by nCounter expression analysis in the total cohort. Using this second method we confirmed the signature derived from the microarray data. Impressively, our gene expression signatures improved the performance of accepted clinical predictors, such as the GAP index (2).

Thus, our study, the first comprehensive study of BAL gene expression patterns and their relevance to outcome in IPF, differs from previous studies not only by the focus on BAL and the novelty of the results, but also by the size of the cohort, the incorporation of a derivation and two replication cohorts, and the comparison to accepted clinical predictors of outcome, making it a significant advance over previous biomarker studies in IPF. This is important, because improvement of outcome prediction over what is currently available may have significant implications on timing and prioritization of lung transplantation. The nCounter expression analysis system is a robust technology that is used for molecular diagnostics and even has Food and Drug Administration approval for some assays and thus would be ideal for use in future prospective studies.

One of the most surprising aspects of our study was that many of the genes associated with mortality were epithelial genes described by Hackett and colleagues (16) to be expressed in ABCs, an airway progenitor population that gives rise to all types of airway epithelial cells and is capable of proliferation and self-renewal (20). The specific ABC gene expression signature we obtained was similar to that previously described by Hackett and colleagues (16) and significantly enriched

among 1,582 genes associated with mortality. Both signatures were derived from ABCs isolated from an outgrow of bronchial epithelial cells and thereby cultured for 21 days. The genes highly expressed by cultured ABCs were overexpressed in BAL of patients with poor survival. Importantly, a similar enrichment was also found with genes identified in ABCs by single-cell RNAseq (19). indicating that this is not solely an *in vitro* signature. Of course, many of the top ABC genes associated with early mortality, such as S100A14, stratifin are not only expressed by ABCs but also by metaplastic epithelial lesions and non–small cell lung and other cancers, especially the so-called basal-like breast cancer. Indeed, our network analysis revealed similarities with mechanisms present in cancer. Similarly, some of the ABC genes associated with mortality indicate a phenotype to squamous differentiation which may *in vivo* be promoted by transforming growth factor-β signaling (25).

The possibility that this finding is caused by bronchoscopy technique or a patient population peculiarity is ruled out because it was found in three independent cohorts, in IPF BAL samples obtained by different investigators, but not in samples obtained from patients with COPD or sarcoidosis. Applying immunocytology and flow cytometry we demonstrated that these findings were not limited to changes in gene expression. CK5/6 double and CK5+ cells were present in BAL of patients with IPF but very rarely in BAL of old healthy volunteers or patients with sarcoidosis. Immunohistochemical analysis of IPF lungs demonstrated abundance of ABCs expressing CK5/6 and ΔNP63 around fibroblast foci, within fibrotic lesions, and occasionally forming hollow structures similar to early honeycomb cysts. In normal histology controls such cells were observed in the airways, but never in the alveolar compartment.

Our results are consistent with previous findings by Chilosi and colleagues (26) that detected ΔNP63 expressing cells in the abnormal bronchiolization that characterizes the IPF lung. This is of particular interest; IPF is characterized by fibroblast foci, honeycomb cyst formation, and bronchiolization of the alveolar space (13, 14, 26–29). In the current model of disease pathogenesis, the bronchiolization and honeycomb cysts formation occur after

the formation of myofibroblast foci (30). However, recent data from animal models suggest that CK5/6+ ΔNP63+ cells may have important roles in the early response to fibrosis. In a model of influenza virus–induced pulmonary fibrosis, Kumar and colleagues (31) showed that a bronchoalveolar subpopulation of ABCs accumulates in fibrotic lesions and that this response may be part of regeneration in healthy mice (31). More recently, Vaughan and colleagues (32) demonstrated that migration and proliferation of a CK5/6+ ΔNP63+ progenitor cell population is an early and key event in the evolution of pulmonary fibrosis. Together with the work by Kumar and coworkers (31) and Vaughan and coworkers (32), our results suggest that ABCs are recruited potentially in response to alveolar epithelial cell injury and their proliferation and invasion determine the distortion of the alveolar structure that is typically associated with IPF.

Very recently, single-cell RNA sequencing of epithelial cells in IPF revealed that normal alveolar type II cells were very rare in the IPF lung (19). Instead, epithelial cells derived from IPF tissues consisted of three major populations: ABCs, goblet cells, and indeterminate cells. This study also demonstrated an abnormal differentiation program in the tissue microenvironment of IPF in which the proximal-peripheral patterns of cell differentiation are disrupted, with many respiratory epithelial cells acquiring aberrant, multilineage-like states (19). Impressively, many of the genes that characterized IPF epithelial genes in this study are in our BAL signature (19). Thus, our results concur and enhance previous observations, but adding to previous observations we demonstrate that the ABC is clinically relevant that may suggest an involvement of this cell type in the pathophysiology of this disease, but definitely need further evaluation.

Our study has several limitations. First, we studied the gene expression profile of a cellular admixture and therefore cannot clarify the exact cellular sources of each change in gene expression. While this is a significant limitation, it is also the foundation for our unexpected finding. Based on our knowledge we had no reason to look for epithelial cells, thus a more targeted experiment at lymphocytes or macrophages would probably have resulted in us missing this most novel finding.

However, this limitation is still an important consideration. Airway epithelial progenitors are a heterogeneous population of cells (32) and our study was not aimed to identify the exact regional source and subtype of the CK5/6[+] cells in the BAL of patients with IPF, which needs to be defined in future single cell–based studies.

A further limitation is that we used cultured ABCs to retrieve their cell-specific signature. We compared cultured ABCs, alveolar macrophages, and bronchial epithelial cells. Our ABC gene list may therefore contain genes highly expressed by other cell types not studied because we found them enriched in cultured ABCs compared with macrophages and bronchial epithelial cells. The described signature of ABCs is influenced by the culture conditions of the cells in our experiment, but the fact that it was also enriched applying a signature retrieved from naive ABCs of IPF lung tissues by single-cell RNAseq reduces the likelihood of culture conditions being a major confounder. Moreover, the described list of ABC genes is not exclusively expressed by ABCs, but rather also expressed by metaplastic lesions and cancers derived from ABCs and basal cells from other organs. However, what is important is that we found using unbiased approaches a significant enrichment of these genes in uncultured BAL of patients with IPF and an association of this signature with early mortality. We believe that our findings warrant a detailed characterization of CK5/6[+] ΔNP63[+] basal-like epithelial cells in IPF and follow-up studies to dissect their role in the pathogenesis of fibrosis.

In conclusion, our foray into the BAL transcriptome was very productive. We identified a gene expression signature that predicts mortality in IPF. This signature was validated in three cohorts and has been shown to improve the accuracy of outcome prediction based on clinical parameters suggesting that it should be considered for transplant prioritization and clinical trial design. Our unexpected finding that genes from ABCs were highly enriched in the BAL of patients likely to progress adds to recently published murine data and may suggest an unexpected role of ABCs in the pathogenesis of IPF. Our results should have significant impact on reconsideration of BAL as part of the evaluation of patients with IPF, and on further studies addressing ABCs as potential therapeutic targets in this devastating disease. ∎

References

1. Martinez FJ, Safrin S, Weycker D, Starko KM, Bradford WZ, King TE Jr, et al.; IPF Study Group. The clinical course of patients with idiopathic pulmonary fibrosis. Ann Intern Med 2005;142:963–967.
2. Ley B, Ryerson CJ, Vittinghoff E, Ryu JH, Tomassetti S, Lee JS, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med 2012;156:684–691.
3. du Bois RM, Weycker D, Albera C, Bradford WZ, Costabel U, Kartashov A, et al. Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. Am J Respir Crit Care Med 2011;184:1382–1389.
4. Prasse A, Probst C, Bargagli E, Zissel G, Toews GB, Flaherty KR, et al. Serum CC-chemokine ligand 18 concentration predicts outcome in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2009;179: 717–723.
5. Herazo-Maya JD, Noth I, Duncan SR, Kim S, Ma SF, Tseng GC, et al. Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. Sci Transl Med 2013;5: 205ra136.
6. Richards TJ, Kaminski N, Gibson KF. Plasma proteins for risk prediction in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2012;185: 1329–1330.
7. Peljto AL, Zhang Y, Fingerlin TE, Ma SF, Garcia JG, Richards TJ, et al. Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. JAMA 2013; 309:2232–2239.
8. Meyer KC, Raghu G, Baughman RP, Brown KK, Costabel U, du Bois RM, et al.; American Thoracic Society Committee on BAL in Interstitial Lung Disease. An official American Thoracic Society clinical practice guideline: the clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. Am J Respir Crit Care Med 2012;185:1004–1014.
9. Ohshimo S, Bonella F, Cui A, Beume M, Kohno N, Guzman J, et al. Significance of bronchoalveolar lavage for the diagnosis of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2009;179:1043–1047.
10. Prasse A, Pechkovsky DV, Toews GB, Jungraithmayr W, Kollert F, Goldmann T, et al. A vicious circle of alveolar macrophages and fibroblasts perpetuates pulmonary fibrosis via CCL18. Am J Respir Crit Care Med 2006;173:781–792.
11. Dickson RP, Huffnagle GB. The lung microbiome: new principles for respiratory bacteriology in health and disease. PLoS Pathog 2015; 11:e1004923.
12. Prasse A, Binder H, Vuga L, Jaeger B, Schupp J, Herazo-Maya JD, et al. BAL gene expression profiling unmask an unexpected role of airway epithelial progenitor cells in IPF [abstract]. Am J Respir Crit Care Med 2015;191:A6358.
13. American Thoracic Society. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). Am J Respir Crit Care Med 2000;161:646–664.
14. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al.; ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med 2011;183:788–824.
15. Raghu G, Rochwerg B, Zhang Y, Garcia CA, Azuma A, Behr J, et al.; American Thoracic Society; European Respiratory society; Japanese Respiratory Society; Latin American Thoracic Association. An official ATS/ERS/JRS/ALAT Clinical Practice Guideline: treatment of idiopathic pulmonary fibrosis. An Update of the 2011 Clinical Practice Guideline. Am J Respir Crit Care Med 2015;192:e3–e19. [Published erratum appears in Am J Respir Crit Care Med 192: 644.]
16. Hackett NR, Shaykhiev R, Walters MS, Wang R, Zwick RK, Ferris B, et al. The human airway epithelial basal cell transcriptome. PLoS One 2011;6:e18378.
17. Binder H, Schumacher M. Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models. BMC Bioinformatics 2008;9:14.
18. Ein-Dor L, Zuk O, Domany E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. Proc Natl Acad Sci USA 2006;103:5923–5928.
19. 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.
20. Rock JR, Randell SH, Hogan BL. Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. Dis Model Mech 2010;3:545–556.
21. Kaminski N. Microarray analysis of idiopathic pulmonary fibrosis. Am J Respir Cell Mol Biol 2003; 29(3, Suppl)S32–S36.

22. Yang IV, Coldren CD, Leach SM, Seibold MA, Murphy E, Lin J, et al. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. Thorax 2013;68: 1114–1121.

23. Kinder BW, Brown KK, Schwarz MI, Ix JH, Kervitsky A, King TE Jr. Baseline BAL neutrophilia predicts early mortality in idiopathic pulmonary fibrosis. Chest 2008;133:226–232.

24. Molyneaux PL, Cox MJ, Willis-Owen SA, Mallia P, Russell KE, Russell AM, et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2014;190: 906–913.

25. Hu L, Liu J, Li Z, Wang C, Nawshad A. Transforming growth factor-β1 activates ΔNp63/c-Myc to promote oral squamous cell carcinoma. Oral Surg Oral Med Oral Pathol Oral Radiol 2016;122: 460–482.e4.

26. Chilosi M, Poletti V, Murer B, Lestani M, Cancellieri A, Montagna L, et al. Abnormal re-epithelialization and lung remodeling in idiopathic pulmonary fibrosis: the role of deltaN-p63. Lab Invest 2002;82: 1335–1345.

27. Travis WD, Costabel U, Hansell DM, King TE Jr, Lynch DA, Nicholson AG, et al.; ATS/ERS Committee on Idiopathic Interstitial Pneumonias. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med 2013;188:733–748.

28. 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.

29. Davies D, MacFarlane A, Darke CS, Dodge OG. Muscular hyperplasia ("cirrhosis") of the lung and bronchial dilatations as features of chronic diffuse fibrosing alveolitis. Thorax 1966;21: 272–289.

30. King TE Jr, Pardo A, Selman M. Idiopathic pulmonary fibrosis. Lancet 2011;378:1949–1961.

31. Kumar PA, Hu Y, Yamamoto Y, Hoe NB, Wei TS, Mu D, et al. Distal airway stem cells yield alveoli in vitro and during lung regeneration following H1N1 influenza infection. Cell 2011;147: 525–538.

32. Vaughan AE, Brumwell AN, Xi Y, Gotts JE, Brownfield DG, Treutlein B, et al. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. Nature 2015;517: 621–625.