

The basement membrane repair response biomarker PRO-C4 predicts progression in idiopathic pulmonary fibrosis: analysis of the PFBIO and PROFILE cohorts

Jannie M. B. Sand¹, Peder Frederiksen¹, Alison E. John², Filipa B. Simões¹, Nils Hoyer³, Thomas S. Prior⁴, Adnan Avdic-Belltheus², Philip L. Molyneaux², Iain D. Stewart², Hernan P. Fainberg², Simon R. Johnson⁵, Morten A. Karsdal¹, Diana J. Leeming¹, Elisabeth Bendstrup⁴, Saher B. Shaker³, Toby M. Maher^{2,6}, R. Gisli Jenkins²

¹ Nordic Bioscience, Herlev, Denmark

² Margaret Turner Warwick Centre for Fibrosing Lung Disease, National Heart and Lung Institute, Imperial College, London, UK

³ Department of Respiratory Medicine, Herlev and Gentofte University Hospital, Copenhagen, Denmark

⁴ Centre for Rare Lung Diseases, Department of Respiratory Disease and Allergy, Aarhus University Hospital, Aarhus, Denmark

⁵ Centre for Respiratory Research, NIHR Biomedical Research Centre and Biodiscovery Institute School of Medicine, University of Nottingham, UK

⁶ Keck School of Medicine, The University of Southern California, Los Angeles, CA, United States

Corresponding author: Jannie M. B. Sand, jsa@nordicbio.com

Keywords: Biomarkers, basement membrane, type IV collagen, collagen, extracellular matrix, idiopathic pulmonary fibrosis, cohort study, prognosis.

Abstract

Background: Idiopathic pulmonary fibrosis (IPF) is characterized by damage to the epithelial layer, closely associated with the alveolar basement membrane (BM). We aimed to investigate how type IV collagen (COL4) in the BM changes with progression of IPF.

Methods: COL4 synthesis (PRO-C4) was detected in blood by the nordicPRO-C4™ biomarker in IPF patients from the two prospective, multicentre, observational, longitudinal cohorts, PFBIO and PROFILE. PRO-C4 trajectories over 12 months were compared between progressors and non-progressors by linear mixed effects regression models. Rate of change in PRO-C4 and lung function were compared by Bayesian bivariate longitudinal models. Cox proportional hazards models analysed baseline PRO-C4 and three-years mortality. COL4 staining in IPF and non-IPF lungs were evaluated by immunohistochemistry.

Results: In PFBIO and PROFILE, 51/220 (23.2%) and 221/459 (48.1%) patients, respectively, had progressive disease at 12 months. Longitudinal PRO-C4 levels were higher in progressors vs. non-progressors (average differences: PFBIO 21.5% (95%CI 3.4-42.9, P=0.0184); PROFILE 10.9% (0.8-22.1; P=0.0340). Monthly rate of change in PRO-C4 was steeper in non-survivors vs. survivors (mean difference up to 3.12% [0.35-5.91]) and was inversely correlated with the change in lung function. High baseline PRO-C4 was associated with increased mortality risk in PFBIO (HR 2.55 [1.27-5.12], P=0.0083). COL4 staining was higher in IPF vs. non-IPF lung but was less obvious in end-stage tissue.

Conclusions: High and increasing serological PRO-C4 levels were prognostic for progression in two independent IPF cohorts. This study suggests that COL4 synthesis assessed by PRO-C4 is a pathologically relevant biomarker of alveolar BM repair in IPF.

What is already known about this topic

- Epithelial damage is a main driver of idiopathic pulmonary fibrosis (IPF) and the basement membrane is known to play a role in fibrosis development. Re-establishment of alveolar structures cannot occur on an injured basement membrane.
- Blood-based biomarkers of epithelial damage and interstitial collagen turnover have been associated with progression of IPF.

What this study adds

- This study investigated the basement membrane repair response associated with IPF progression using the biomarker PRO-C4 which quantifies type IV collagen synthesis in blood.
- High PRO-C4 levels were associated with IPF progression in two independent prospective longitudinal cohorts. Additionally, the monthly rate of increase in PRO-C4 was steeper in non-survivors compared with survivors and was inversely correlated to rate of change in lung function measures. Adding to the literature that a systemic measure of basement membrane repair and not only epithelial damage markers may identify progressive patients.
- Type IV collagen staining was higher in IPF compared with non-IPF lung tissue and became less obvious in end-stage IPF with the loss of alveolar structures. Such data have not been shown previously.
- These results suggest that type IV collagen synthesis is a basement membrane repair response that is elevated in progressive IPF to aid in the repair of the damaged epithelial layer.

How this study might affect research, practice, or policy

- This study suggests that PRO-C4 is a pathologically relevant prognostic biomarker for IPF that may be used as an enrichment tool for the progressive phenotype in clinical trials and potentially to monitor patients in clinical practice.

Introduction

Epithelial damage is a key feature of idiopathic pulmonary fibrosis (IPF)[1–3], and serological biomarkers reflecting this have been associated with progression and mortality[4,5]. Epithelial cells are closely associated with the basement membrane (BM), forming a fundamental scaffold for alveolar structures. Repetitive epithelial injury results in damage to the underlying BM, triggering a repair response to rebuild the tissue barrier[6]. The BM must be intact before it can be re-epithelialized[7], making its repair response a prerequisite for lung tissue repair. Loss of BM has been described as a cardinal feature of IPF[8].

Collagen fragments reflecting fibrotic changes in the lung interstitium have been identified as prognostic biomarkers in IPF[9–13]. Conversely, by assessing fragments of type IV collagen (COL4), a main constituent of BMs[14], we may elucidate the critical role of BM repair in IPF. Identifying and implementing biomarkers to guide prognosis and treatment is greatly needed in IPF[15]. A blood-based biomarker, reflecting a relevant disease mechanism, may be the ideal tool for precision medicine and clinical trial endpoints. Currently, the most rewarding use of such a biomarker may be as an enrichment tool for the progressive phenotype in trials, increasing the chance of detecting treatment effects in a shorter timeframe.

COL4 was measured with the serological biomarker PRO-C4, targeting the COL4 7S domain[16]. Serological COL4 7S levels have previously been found elevated in IPF [17] and in liver fibrosis patients, where levels correlated with the histological degree of fibrosis[16,18,19], indicating that circulating 7S may reflect COL4 synthesis. Given the importance of BM repair in lung injury and its dysregulation in IPF, we quantified PRO-C4 in two prospective cohorts to determine whether it was associated with IPF progression and mortality.

Materials and Methods

PFBIO cohort

The Pulmonary Fibrosis Biomarker (PFBIO) cohort, initiated in 2016, is a prospective, multicentre, observational cohort of incident IPF patients diagnosed by a multidisciplinary team according to international guidelines[20,21] (NCT02755441). Before 2018, patients with a possible usual interstitial pneumonia (UIP) pattern on HRCT, but without other known causes of interstitial lung disease (ILD), underwent bronchoalveolar lavage before diagnosis. Patients with a non-UIP or a possible UIP pattern underwent a lung biopsy before diagnosis, when other ILD causes could not be excluded. After 2018, diagnostic procedures followed updated guidelines[22]. The only exclusion criterion was inability to provide informed consent. Patients were recruited at the two largest ILD centres in Denmark (Herlev and Gentofte University Hospital and Aarhus University Hospital) and assessed at baseline, six and 12 months and thereafter annually for up to five years, where clinical, physiological, and biochemical data were collected. Antifibrotic therapy was initiated after baseline blood sampling, if applicable. Patients with three years follow-up were included in this study.

PROFILE cohort

The Prospective Observation of Fibrosis in the Lung Clinical Endpoints (PROFILE) cohort is a prospective, multicentre, observational study of incident cases (within six months of presentation) of idiopathic fibrotic interstitial lung disease from 2010 to 2018. Participants were included through two centres in the UK: Nottingham University Hospital, Nottingham (NCT01134822) and Royal Brompton Hospital, London (NCT01110694) and diagnosed by a multidisciplinary team according to international guidelines[20]. Diagnosis of non-specific interstitial pneumonia (iNSIP) was based on radiological features of HRCT and the lack of any secondary cause. Patients who could not give informed consent and participating in an interventional clinical trial were excluded. Most patients did not receive antifibrotic therapy because visits largely occurred prior to their approval within the National Health Service. Patients were assessed at baseline, one, three, and six months and annually for three years where clinical, physiological, and biochemical data were collected.

Biomarker analysis

Blood for biomarker analysis were collected at the indicated timepoints, following standard procedures. For PFBIO, blood was collected in Vacuette Z serum clot activator tubes (Greiner Bio-One, Austria), left at room temperature for 30 min, and centrifuged at 1300g for 10 min at 4°C. For PROFILE, blood was collected in lithium heparin coated tubes (Becton Dickinson, UK). Plasma was separated by centrifugation at 1500g for 10 min at 4°C. All samples were processed within 2 hrs of collection, aliquoted, and stored at -80°C until analysed.

PRO-C4 levels were assessed in one batch per cohort using the competitive ELISA nordicPRO-C4™ specific for the alpha-1 chain 7S domain (Nordic Bioscience, Denmark) as previously described[16]. PRO-C4 assessed in serum and plasma performs similarly albeit absolute levels differ[23]. NordicPRO-C3™ and nordicPRO-C6™ were measured as previously described[10,11], quantifying type III and VI collagen formation, respectively. Samples outside the detection limits were assigned the lower or upper limit of quantification.

COL4 immunohistochemical detection in lung tissue

Formalin-fixed paraffin embedded lung tissue sections from IPF (n=12) and non-IPF (n=3) were stained for COL4 by immunohistochemistry. IPF tissue was obtained from post-mortem or lung transplant patients following informed written consent and ethical review (ethical approval numbers: Nottingham Respiratory

Research Unit, 08/H0407/1; Papworth Hospital Research Tissue Bank (REC) 08/H0304/56, Brompton Clinical Research Facility Respiratory Biobank, 20/SC/0142). Non-fibrotic tissue was obtained from normal margins of lung cancer resections. Tissues were sectioned (5 μ m thickness) and immunohistochemical analysis was performed using the Novolink Max Polymer Detection system (Leica, UK). In brief, sections were deparaffinised with xylene and rehydrated through graded ethanol prior to heat-induced antigen retrieval in sodium citrate buffer (pH 6.0). Sections were incubated with recombinant monoclonal antibody targeting the alpha-1(IV) collagen chain (EPR20966, ab214417, Abcam, UK, 1:500), overnight at 4°C. 3-3' Diaminobenzidine tetrahydrochloride (Novolink DAB substrate buffer plus) was freshly prepared and used as chromogen. Tissue was counterstained with Meyer's haematoxylin for 6 min. Negative (omission of the primary antibody) controls were included. Parallel tissue sections were stained with a Trichrome Stain Kit (ab150686, Abcam). All slides were imaged using a Leica Aperio VERSA slide scanner and analysed using Aperio Imagescope software version 12.4.6. Presented images are representative of staining patterns for tissue with low level fibrosis, dense fibrosis, and honeycombing.

Statistical analysis

Baseline demographics are presented as mean (SD), geometric mean (geometric SD), or counts (%) and groups are compared by t-tests or Chi-square tests for continuous or categorical variables, respectively. Baseline PRO-C4 levels were associated with forced vital capacity (FVC) and diffusing capacity for carbon monoxide (DLCO) in percent of predicted values (FVC%; DLCO%) by partial correlation analyses adjusting for age and sex.

A combined endpoint of disease progression ($\geq 10\%$ absolute FVC% decline or death at 12 months) was applied. Patients with missing FVC were grouped as indeterminate for PFBIO and excluded from the progression analysis. For PROFILE, these patients were assigned a progression status following clinical adjudication. A linear mixed effects model with patient-specific random intercepts, adjusted for age and sex (and antifibrotic treatment for PFBIO) was used to estimate contrasts in average PRO-C4 across visits. PRO-C4 was log-transformed for normality and variance homogeneity, but back-transformed to the original scale and presented as least square (LS) geometric means or geometric means ratios with 95% confidence intervals (CI). Linear mixed models implicitly impute missing outcome data, providing valid inference in a longitudinal setting under various forms of missingness at random. Since the model imputes post-dropout observations due to death, we applied a pattern mixture model[24], conditioning on time of death to distinguish missing data due to death from nonresponse, and examined the sensitivity of the results. This model, like the linear mixed model but with an extra term for time of death, described PRO-C4 trajectories relative to disease progression and survival.

Within PROFILE, we explored associations between longitudinal PRO-C4 and FVC% (or DLCO%) from baseline to 36 months using bivariate linear mixed effects models. Two models were applied: one for FVC and PRO-C4, and another for DLCO and PRO-C4. Each model assumed a linear relationship between the average lung function or PRO-C4 and time since baseline, with separate associations conditional on time to death. Random effects were introduced, representing patient-specific deviations from average intercepts and slopes. To link lung function to PRO-C4, we explicitly included the covariances of random effects across both models[25]. The bivariate longitudinal models were specified within a Bayesian framework, and the joint posterior distribution of the parameters was estimated using the R package rstanarm[26] with default prior specifications: normal distributions for intercepts and coefficients, half-Cauchy for residual variances, and an LKJ distribution for the correlation structure. Priors are intended to be weakly informative, providing moderate regularization and stabilizing the MCMC sampling. Posterior distribution summaries were based on 20,000 samples. As these models condition on the time of death, they are descriptive and not predictive, and

no measures of prediction performance were estimated. Additionally, data were analysed using a joint model linking a longitudinal sub-model for PRO-C4 trajectories with a survival sub-model, by a parameter quantifying the effect of the underlying “true” but unobserved patient-specific PRO-C4 trajectories (without random within patient variation) on the hazard of mortality. The longitudinal sub-model was a random intercept and slope model (no covariates), while the survival model was a Cox proportional hazards model with age, sex and baseline FVC% as covariates, and current PRO-C4 value as a time-dependent covariate. The model was fitted using MCMC sampling[27], and estimates of prediction performance (AUC and Brier score) were obtained by cross-validation.

Cox multivariable proportional hazards models adjusted for age, sex, and baseline FVC% were used to describe the association between baseline PRO-C4, PRO-C3 and/or PRO-C6 and three-year mortality. Kaplan-Meier and restricted cubic splines (3 knots at percentiles 10, 50, and 90) analyses were initially applied for PRO-C4 measurements. The PRO-C4 effect on hazard was modelled linear in the multi-marker analyses, as no evidence for non-linearities was identified in the initial analysis. Prediction performance was estimated using bootstrapping.

All analyses were performed in R (R Core Team, 2023, version 4.3.0), with two-sided 95% CIs corresponding to a nominal two-sided 5% level of significance. No adjustments were made for multiple testing.

Results

Baseline characteristics

Patients with pulmonary fibrosis were included from PFBIO (n=220) and PROFILE (n=459) cohorts (**Figure 1**). Patients missing baseline FVC or PRO-C4 were excluded from the analysis. Baseline demographics are shown in **Tables 1, 2, and S1**. All PFBIO participants had IPF, while in PROFILE, 406 (88.5%) had IPF and 53 (11.5%) had iNSIP. PFBIO and PROFILE included 51 (23.2%) and 221 (48.1%) progressors, respectively. Progressors and non-progressors had similar age and sex, but progressors had lower baseline FVC% in PROFILE and DLCO% in both cohorts. Following baseline, 178 PFBIO participants (80.9%) started antifibrotic therapy, with non-progressors having the highest proportion of treated patients (88.2% vs. 66.7%). In PROFILE, 10.0% participants received antifibrotics one time during follow-up. These differences likely reflect prescribing rules at recruitment. At baseline, PRO-C4 levels inversely correlated with DLCO% (PFBIO $r=-0.17$, $P=0.0145$; PROFILE $r=-0.15$, $P=0.0019$) but not with FVC% in partial correlation analyses adjusted for age and sex.

PFBIO	Non-progressors (n=152)	Progressors (n=51)	Total (n=220)	P-value
Age, years	73.50 (7.79)	72.69 (8.51)	73.52 (7.84)	0.529
Sex				0.794
Male	116 (76.3%)	38 (74.5%)	168 (76.4%)	
Female	36 (23.7%)	13 (25.5%)	52 (23.6%)	
Smoking				0.951
Never	37 (24.3%)	12 (23.5%)	56 (25.5%)	
Former	105 (69.1%)	35 (68.6%)	149 (67.7%)	
Active	10 (6.6%)	4 (7.8%)	15 (6.8%)	
BMI	27.66 (4.31)	26.52 (4.37)	27.58 (4.52)	0.103
Antifibrotics at FU				0.001
None	18 (11.8%)	17 (33.3%)	42 (19.1%)	
Nintedanib	63 (41.1%)	13 (25.5%)	81 (36.8%)	
Pirfenidone	71 (46.7%)	21 (41.2%)	97 (44.1%)	
FVC%	90.24 (17.85)	88.14 (21.39)	89.10 (18.91)	0.491
DLCO%	54.15 (13.73)	49.00 (13.80)	52.43 (13.62)	0.022
GAP index				0.037
I	80 (52.6%)	21 (41.2%)	108 (49.1%)	
II	71 (46.7%)	27 (52.9%)	107 (48.6%)	
III	1 (0.7%)	3 (5.9%)	5 (2.3%)	
Serum PRO-C4, ng/mL	167.0 (1.63)	199.4 (1.88)	174.6 (1.70)	0.039
Change in FVC% 12m, absolute	2.47 (8.56)	-13.81 (8.89)	-0.65 (10.73)	< 0.001
Change in DLCO% 12m, absolute	-3.45 (7.26)	-10.50 (8.84)	-4.74 (8.03)	< 0.001
Death 12m				< 0.001
Alive	152 (100.0%)	30 (58.8%)	199 (90.5%)	
Dead	0 (0.0%)	21 (41.2%)	21 (9.5%)	
Death 36m				< 0.001
Alive	121 (79.6%)	21 (41.2%)	153 (69.5%)	
Dead	31 (20.4%)	30 (58.8%)	67 (30.5%)	

Table 1. PFBIO baseline demographics. Data are presented as mean (SD) or n (%), except PRO-C4 which is shown as geometric mean with geometric SD. Progression was defined as FVC decline $\geq 10\%$ or all-cause mortality at 12 months. For 17 patients, progression status could not be defined due to missing FVC data at follow-up. Data for these patients are shown in Table S1. Groups compared by t-tests or Chi-square tests. BMI, Body Mass Index; FVC, forced vital capacity; DLCO, diffusing capacity for carbon monoxide; FVC%/DLCO%, FVC/DLCO in percent of predicted value; GAP, gender-age-physiology; m, month.

PROFILE	Non-progressors (n=238)	Progressors (n=221)	Total (n=459)	P-value
Age, years	69.85 (8.50)	71.36 (8.44)	70.58 (8.49)	0.057
Sex				0.186
Male	188 (79.0%)	163 (73.8%)	351 (76.5%)	
Female	50 (21.0%)	58 (26.2%)	108 (23.5%)	
Smoking				0.098
Never	65 (27.4%)	76 (34.4%)	141 (30.8%)	
Former	155 (65.4%)	137 (62.0%)	292 (63.8%)	
Active	17 (7.2%)	8 (3.6%)	25 (5.5%)	
BMI	28.38 (4.27)	28.40 (4.70)	28.39 (4.48)	0.962
Diagnosis				0.013
IPF	202 (84.9%)	204 (92.3%)	406 (88.5%)	
iNSIP	36 (15.1%)	17 (7.7%)	53 (11.5%)	
Antifibrotics at FU				0.963
None	214 (89.9%)	199 (90.0%)	413 (90.0%)	
At any point	24 (10.1%)	22 (10.0%)	46 (10.0%)	
FVC%	82.64 (18.71)	76.47 (18.32)	79.67 (18.76)	< 0.001
DLCO%	50.69 (15.56)	41.12 (12.86)	46.19 (15.12)	< 0.001
GAP index				0.001
I	106 (48.2%)	66 (33.8%)	172 (41.4%)	
II	98 (44.5%)	97 (49.7%)	195 (47.0%)	
III	16 (7.3%)	32 (16.4%)	48 (11.6%)	
Plasma PRO-C4, ng/mL	120.7 (1.77)	129.9 (1.81)	125.1 (1.79)	0.179
Change in FVC% 12m, absolute	0.87 (7.16)	-15.92 (8.67)	-5.31 (11.21)	< 0.001
Change in DLCO% 12m, absolute	-2.26 (7.22)	-9.10 (7.17)	-4.49 (7.88)	< 0.001
Death 12m				< 0.001
Alive	238 (100.0%)	157 (71.0%)	395 (86.1%)	
Dead	0 (0.0%)	64 (29.0%)	64 (13.9%)	
Death 36m				< 0.001
Alive	194 (81.5%)	81 (36.7%)	275 (59.9%)	
Dead	44 (18.5%)	140 (63.3%)	184 (40.1%)	

Table 2. PROFILE baseline demographics. Data are presented as mean (SD) or n (%), except PRO-C4 which is shown as geometric mean with geometric SD. Progression was defined as FVC decline $\geq 10\%$ or all-cause mortality at 12 months, as well as clinical adjudication. Groups compared by t-tests or Chi-square test. BMI, Body Mass Index; FVC, forced vital capacity; DLCO, diffusing capacity for carbon monoxide; iNSIP, idiopathic non-specific interstitial pneumonia; FVC%/DLCO%, FVC/DLCO in percent of predicted value; GAP, gender-age-physiology; m, month.

Longitudinal PRO-C4 levels were associated with disease progression

PRO-C4 levels were evaluated longitudinally to investigate associations with disease progression (**Figure 2**). In PFBIO, progressors had higher serum PRO-C4 at all timepoints compared with non-progressors, with an average difference between groups of 28% at 12 months in the implicit imputation model (**Table 3**). In PROFILE, plasma PRO-C4 was numerically higher in progressors at all timepoints (**Figure 2B**), reaching statistical significance at 12 months (18% difference, **Table 3**). The average monthly rate of change was higher among progressors compared with non-progressors (0.96% [95% CI 0.43-1.5] vs. 0.27% [95% CI -0.19-0.74]; $P=0.026$). Similar results were obtained using an FVC-based progression definition comparable to PFBIO, excluding patients lacking follow-up FVC data (**Figure S1**).

To evaluate the sensitivity of the results without implicit imputation of biomarker observations following death, a pattern mixture analysis was performed, evaluating PRO-C4 trajectories by time to death and 12-month disease progression status (**Figure S2, Table 3**). Longitudinal PRO-C4 levels were numerically higher in

progressors in the two cohorts. In both analyses, PFBIO had a greater difference in PRO-C4 levels between progression groups than PROFILE. Overall differences between analyses were small but probably reflect mortality patterns, as patients who died faster had higher PRO-C4 levels (**Figure S2**).

Disease progressors – non-progressors	Difference in PRO-C4 at 12 months % [95% CI]	P-value
PFBIO (Implicit imputation)	28% [6%, 54%]	0.011
PFBIO (Pattern mixture)	23% [0%, 51%]	0.054
PROFILE (Implicit imputation)	18% [6%, 31%]	0.003
PROFILE (Pattern mixture)	13% [0%, 26%]	0.047

Table 3. Difference in PRO-C4 levels at 12 months between disease progressors and non-progressors in PFBIO and PROFILE. Differences were estimated by two different models and are presented as % with 95% confidence intervals. The implicit imputation model imputes observations after dropout due to death, while the pattern mixture model conditions on time of death. CI, confidence interval.

Rate of change in plasma PRO-C4 was associated with rate of change in lung function

To visualise the three-year development in PRO-C4 and lung function in PROFILE, we stratified patients by survival time and visualized changes in FVC%, DLCO%, and PRO-C4 (**Figure S3**). Patients with shorter survival times demonstrated steeper decline in FVC% and DLCO% compared to survivors. PRO-C4 remained stable in survivors, whereas groups with shorter survival times demonstrated progressively steeper PRO-C4 trajectories. These observations agree with rate of change estimates from Bayesian bivariate models shown in **Table S2**. A shorter time to death was associated with a larger monthly reduction in FVC% and DLCO% and an increase in PRO-C4.

Figure 3 shows the joint posterior distribution contour of the monthly rate of change in plasma PRO-C4 vs. FVC% or DLCO%. The width of the plots reflects large variability of the slope estimates at the individual patient level, higher in patients with shorter survival time. The within-patient correlations between monthly rate of change in PRO-C4 and lung function were negative (FVC%: -0.16 [95 % CI -0.41 to 0.11]; DLCO%: -0.25 [95% CI -0.49 to -0.00]).

To further investigate the association between longitudinal changes in PRO-C4 and the three-year mortality risk, a joint model was applied, integrating a survival sub-model with a longitudinal sub-model of underlying PRO-C4 (**Tables S3-5, Figure S4**). Longitudinal PRO-C4 levels were significantly associated with mortality risk with an adjusted hazard ratio (HR) of 1.27 (P=0.025) (**Table 4, Figure S5**).

Term	HR [95% CI]	P-value
Current PRO-C4	1.27 [1.03, 1.58]	0.025
Age	1.05 [1.01, 1.06]	0.002
Female:Male	1.01 [0.68, 1.45]	0.952
FVC%	0.96 [0.96, 0.97]	< 0.001

Table 4. Association between mortality risk and PRO-C4 trajectories in PROFILE. Hazard ratios for three-year mortality risk estimated from a joint model that combines a longitudinal sub-model for PRO-C4 trajectories with a survival sub-model. Hazard ratios are reported per two-fold increase in current underlying PRO-C4, one-year increase in age, female versus male and 1% increase in %FVC. Data are presented as mean with 95% confidence intervals. HR, hazard ratio; CI, confidence interval; %FVC percent predicted forced vital capacity.

Baseline PRO-C4 levels were related to mortality in one cohort

The association of baseline PRO-C4 with three-years mortality was evaluated in both cohorts (**Figure 4**), with 67 (30.5%) deaths in PFBIO and 184 (40.1%) in PROFILE. PFBIO participants in the highest baseline PRO-C4 quartile had a lower survival rate compared to the lowest quartile, with an adjusted HR of 2.55 (95% CI 1.27-

5.12; P=0.0083). This association was not observed in PROFILE (HR=1.23 [95% 0.81-1.87], P=0.3367). Non-linear effects of PRO-C4 on the hazard were not identified in any cohort by Wald tests (1 degree of freedom).

To further explore this relationship, this analysis was expanded to include the fibrogenesis biomarkers, PRO-C3 and PRO-C6, previously shown to be associated with mortality in both cohorts[10,11]. To assess their combined value with PRO-C4, we applied multivariable Cox regression models for three-year mortality risk, adjusting baseline biomarkers for age, sex and FVC% (**Table 5**). In PFBIO, PRO-C4 was the strongest biomarker (AUC=0.65), while FVC% was not significantly associated with mortality (AUC=0.60). Conversely, in PROFILE, baseline FVC% (AUC=0.72) and fibrogenesis biomarkers (AUC=0.73) had a stronger prognostic value than PRO-C4.

Model	Term	PFBIO				PROFILE			
		HR [95% CI]	P-value	AUC [95% CI]	Brier [95% CI]	HR [95% CI]	P-value	AUC [95% CI]	Brier [95% CI]
1	Age	1.02 [0.98, 1.06]	0.362	0.60 [0.46, 0.71]	0.21 [0.18, 0.26]	1.04 [1.02, 1.05]	< 0.001	0.72 [0.65, 0.79]	0.21 [0.19, 0.24]
	Female:Male	0.52 [0.23, 1.17]	0.115			1.03 [0.72, 1.47]	0.876		
	FVC%	0.99 [0.97, 1.00]	0.054			0.96 [0.96, 0.97]	0.000		
2	PRO-C4	2.07 [1.31, 3.27]	0.002	0.65 [0.53, 0.75]	0.20 [0.17, 0.26]	1.05 [0.87, 1.24]	0.688	0.72 [0.65, 0.77]	0.21 [0.19, 0.24]
	Age	1.03 [0.99, 1.07]	0.192			1.04 [1.02, 1.05]	< 0.001		
	Female:Male	0.49 [0.22, 1.09]	0.080			1.03 [0.72, 1.46]	0.891		
	FVC%	0.99 [0.97, 1.00]	0.052			0.96 [0.96, 0.97]	< 0.001		
3	PRO-C3	1.55 [1.10, 2.21]	0.014	0.62 [0.51, 0.76]	0.21 [0.17, 0.26]	1.38 [1.09, 1.74]	0.008	0.73 [0.67, 0.78]	0.21 [0.19, 0.23]
	Age	1.02 [0.98, 1.06]	0.364			1.03 [1.01, 1.05]	< 0.001		
	Female:Male	0.56 [0.25, 1.25]	0.157			1.10 [0.77, 1.57]	0.614		
	FVC%	0.99 [0.97, 1.00]	0.066			0.96 [0.96, 0.97]	< 0.001		
4	PRO-C6	1.82 [0.99, 3.35]	0.054	0.61 [0.49, 0.73]	0.21 [0.17, 0.26]	1.73 [1.26, 2.37]	< 0.001	0.73 [0.66, 0.79]	0.21 [0.19, 0.23]
	Age	1.01 [0.97, 1.05]	0.775			1.03 [1.01, 1.05]	0.005		
	Female:Male	0.48 [0.21, 1.09]	0.080			1.06 [0.74, 1.51]	0.746		
	FVC%	0.99 [0.97, 1.00]	0.055			0.97 [0.96, 0.97]	< 0.001		
5	PRO-C4	1.95 [1.23, 3.08]	0.005	0.65 [0.50, 0.76]	0.21 [0.16, 0.26]	0.97 [0.81, 1.15]	0.696	0.73 [0.66, 0.78]	0.21 [0.19, 0.24]
	PRO-C3	1.49 [0.99, 2.26]	0.065			1.16 [0.86, 1.56]	0.325		
	PRO-C6	1.05 [0.52, 2.10]	0.891			1.59 [1.09, 2.31]	0.015		
	Age	1.03 [0.98, 1.07]	0.234			1.03 [1.01, 1.05]	0.006		
	Female:Male	0.51 [0.23, 1.16]	0.110			1.09 [0.76, 1.56]	0.637		
	FVC%	0.99 [0.97, 1.00]	0.071			0.97 [0.96, 0.97]	< 0.001		

Table 5. Multivariable Cox regression analyses for mortality risk in PFBIO and PROFILE. Hazard ratios for three-years mortality risk adjusted for age, sex and baseline forced vital capacity. Models 2-4 included single biomarkers (PRO-C4, PRO-C3 or PRO-C6, respectively), while model 5 included all three biomarkers. Hazard ratios are reported per two-fold increase in baseline biomarker levels, one-year increase in age, female versus male and 1% increase in %FVC, along with estimated time-dependent area under ROC and Brier scores for a three-year prediction horizon. Data are presented as mean with 95% confidence intervals. HR, hazard ratio; CI, confidence interval; %FVC percent predicted forced vital capacity; AUC area under the ROC curve. AUC and Brier scores were estimated using bootstrap resampling (B=250) using the R package riskRegression[28].

Immunohistochemical localization of COL4 in lung tissue

COL4 expression in lung tissue from non-fibrotic (non-IPF) and IPF lungs were compared by immunohistochemistry (**Figure 5**). In non-IPF lungs, COL4 localised to the alveolar epithelium, in association with the BM, and in blood vessel walls. In IPF samples, COL4 immunostaining was variable. In regions with visible alveolar epithelium, airspaces, and less dense fibrosis (Pt1 and 2), COL4 was expressed along the alveolar BM and in the epithelial cells cytoplasm near developing fibrotic regions. In contrast, COL4 expression was reduced in regions with honeycombing and widespread dense fibrosis (Pt3), evidenced by extensive collagenous connective tissue deposition by Masson's trichrome staining (Pt3). COL4 was

undetected within epithelial cells lining honeycomb cysts and surrounding areas but remained detected around blood vessels.

Discussion

In this study we evaluated COL4 in IPF, with key findings from two independent prospective multicentre cohorts: (1) higher longitudinal PRO-C4 levels were associated with greater progression; (2) PRO-C4 rate of change was steeper in non-survivors compared with survivors and negatively correlated with change in lung function; and (3) COL4 staining was higher in IPF than non-IPF lung tissue. The association between rate of change in PRO-C4 and lung function suggests that this systemic biomarker may reflect changes in disease behaviour. An IPF phase 2 trial underlined this potential by demonstrating reduced serum PRO-C4 upon treatment with the lysophosphatidic acid receptor 1 antagonist BMS-986020[29]. Together, these data indicate that COL4 is important in IPF pathogenesis and that serological PRO-C4 may be a potential biomarker for disease progression.

Epithelial damage is a key factor in promoting and upholding IPF, with alveolar BM disruption occurring in early disease stages[30]. The BM importance in epithelial repair prompted us to assess a systemic biomarker of COL4 synthesis, PRO-C4. Epithelial damage induces a response to repair the BM and re-establish the protective epithelial barrier[6–8]. Thus, PRO-C4 could be a biomarker for the epithelium/BM repair response. An early study quantified serological COL4 7S by a radioimmunoassay, reporting elevated levels in IPF compared with healthy controls[17], with further increases during exacerbations. This aligns with our findings of elevated PRO-C4 in progressive IPF and suggests that injury induces COL4 synthesis.

The role of COL4, a key BM protein, in IPF pathogenesis is unclear but is suggested to affect epithelial organization and fibroblast activity[6,8,31–36]. COL4 consists of six α -chains (α 1– α 6) which self-assemble intracellularly into heterotrimeric protomers (α 1 α 1 α 2, α 3 α 4 α 5, α 5 α 5 α 6) and associate following secretion to form a lattice-shaped network[14]. Irregular thickening and gaps in the linear COL4 pattern in alveolar BMs were already observed in early IPF[30]. Our data demonstrate variable COL4 staining in IPF lung tissue, with high expression in sections with visible alveolar airspaces and evidence of alveolar thickening within and near fibrotic regions. In contrast to the non-IPF lung, alveolar epithelial cells had extensive cytoplasmic staining, suggesting that COL4 was being synthesised by these cells during fibrosis. This observation aligns with single-cell RNA sequencing studies identifying alveolar epithelial type I, aberrant basaloid, and endothelial cells as key sources of COL4 (α 1) in IPF lungs [37]. COL4 was less obvious in BM and epithelial cells near honeycomb cysts in regions with extensive extracellular matrix deposition, suggesting a loss of COL4 synthesis in end-stage IPF.

The loss of COL4 staining in advanced disease may reflect the loss of alveolar epithelial type I cells and a complete disruption of the BM and epithelial integrity. Thus, rising PRO-C4 levels likely reflect an ineffective response to repair the alveolar BM. In severe end-stage tissue, the BMs repair capacity is reduced, and this futile response is lost along with the loss of alveolar units, reaching a “point of no return” for re-epithelialization and COL4 synthesis, thus reducing PRO-C4 release. This hypothesis is consistent with our recent cluster analysis where patients with very low PRO-C4 in combination with high epithelial injury had a poor prognosis[38].

Differences in baseline disease severity and antifibrotic therapy across cohorts may explain the observed variations in the relationship between baseline PRO-C4/fibrogenesis biomarkers and mortality. In PROFILE, patients had more advanced disease (**Table S6**), likely resulting in significant alveolar damage and collagen accumulation, which may account for the stronger association between fibrogenesis biomarkers and mortality. Conversely, PFBIO included less severe patients who probably had more active BM remodeling. As a result, baseline PRO-C4 had a greater prognostic value in this group.

When investigating trajectories of COL4 synthesis in PROFILE, we found that the PRO-C4 monthly rate of change was steeper in non-survivors compared with survivors and inversely correlated with change in FVC% and DLCO%. Patients with shorter survival times experienced the largest PRO-C4 increase and lung function decline. These data indicate that COL4 synthesis increases with progression as a response to epithelial injury and intensifies with closeness to a fatal outcome until the loss of alveolar units hinders the capacity to repair the BM completely. IPF patients do not display homogeneous end-stage tissue in their entire lung, and locations with somewhat preserved alveolar BMs seem to over-produce COL4 to compensate for the damaged epithelium.

The major strength of our study is the replication of results in two prospective cohorts allowing longitudinal analysis of incident IPF and the identification of a biomarker that reflects disease pathology, as illustrated by COL4 immunostaining. One limitation is the missing longitudinal lung function and biomarker measurements (**Tables S7-11**), a common problem in longitudinal studies[39,40]. Here, we used a composite endpoint for progression and implicitly imputed missing biomarker data, but sensitivity analyses indicated that this did not affect results. Conditional on time to death, the longitudinal trajectories for patients who dropped-out resembled those of patients who completed the last visit. Data for patients with and without measurements at subsequent visits were also similar. No clear relationship was found between disease progression and the likelihood of missing values (**Figures S6-13**). Thus, missing at random appears to be a reasonable assumption when conditioning on time to death. Another limitation is the availability of different sample matrices across cohorts, resulting in different average PRO-C4 values. Due to this, we avoided pooling the measurements and applying a multilevel cohort-patient model, as this approach may not capture potential matrix-related heterogeneity in effect estimates.

In conclusion, we showed that COL4 levels in blood and lung tissue were associated with IPF progression. As COL4 is a key BM component, PRO-C4 is a potential biomarker for alveolar BM repair in IPF. High and increasing PRO-C4 levels correlated with lung function decline and reduced survival. Taken together, this study shows that PRO-C4 is a relevant biomarker for IPF, reflecting key pathological features, and may be used to identify patients at high risk of progression.

References

- 1 Camelo A, Dunmore R, Sleeman MA, *et al.* The epithelium in idiopathic pulmonary fibrosis: Breaking the barrier. *Front Pharmacol.* 2014;4 JAN. doi: 10.3389/fphar.2013.00173
- 2 Xu Y, Mizuno T, Sridharan A, *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight.* 2016;1:e90558. doi: 10.1172/jci.insight.90558
- 3 Chambers RC, Mercer PF. Mechanisms of alveolar epithelial injury, repair, and fibrosis. *Ann Am Thorac Soc.* 2015;12:S16–20. doi: 10.1513/AnnalsATS.201410-448MG
- 4 Maher TM, Oballa E, Simpson JK, *et al.* An epithelial biomarker signature for idiopathic pulmonary fibrosis: an analysis from the multicentre PROFILE cohort study. *Lancet Respir Med.* 2017;5:946–55. doi: 10.1016/S2213-2600(17)30430-7
- 5 Molyneaux PL, Fahy WA, Byrne AJ, *et al.* CYFRA 21-1 Predicts Progression in Idiopathic Pulmonary Fibrosis A Prospective Longitudinal Analysis of the PROFILE Cohort. *Am J Respir Crit Care Med.* 2022;205:1440–8. doi: 10.1164/rccm.202107-1769OC
- 6 Wilson SE. Fibrosis Is a Basement Membrane-Related Disease in the Cornea: Injury and Defective Regeneration of Basement Membranes May Underlie Fibrosis in Other Organs. *Cells.* 2022;11.
- 7 Strieter RM. What differentiates normal lung repair and fibrosis? Inflammation, resolution of repair, and fibrosis. *Proc Am Thorac Soc.* 2008;5:305–10. doi: 10.1513/pats.200710-160DR
- 8 Basset F, Ferrans VJ, Soler P, *et al.* Intraluminal fibrosis in interstitial lung disorders. *Am J Pathol.* 1986;122:443–61.
- 9 Jenkins RG, Simpson JK, Saini G, *et al.* Longitudinal change in collagen degradation biomarkers in idiopathic pulmonary fibrosis: An analysis from the prospective, multicentre PROFILE study. *Lancet Respir Med.* 2015;3:462–72. doi: 10.1016/S2213-2600(15)00048-X
- 10 Organ LA, Duggan AMR, Oballa E, *et al.* Biomarkers of collagen synthesis predict progression in the PROFILE idiopathic pulmonary fibrosis cohort. *Respir Res.* 2019;20:1–10. doi: 10.1186/s12931-019-1118-7
- 11 Hoyer N, Jessen H, Prior TS, *et al.* High turnover of types III and VI collagen in progressive idiopathic pulmonary fibrosis. *Respirology.* 2021;26:582–9. doi: 10.1111/resp.14056
- 12 Jessen H, Hoyer N, Prior TS, *et al.* Turnover of type I and III collagen predicts progression of idiopathic pulmonary fibrosis. *Respir Res.* 2021;22:1–8. doi: 10.1186/s12931-021-01801-0
- 13 Jessen H, Hoyer N, Prior TS, *et al.* Longitudinal serological assessment of type VI collagen turnover is related to progression in a real-world cohort of idiopathic pulmonary fibrosis. *BMC Pulm Med.* 2021;21:1–8. doi: 10.1186/s12890-021-01684-3
- 14 Jandl K, Mutgan AC, Eller K, *et al.* The basement membrane in the cross-roads between the lung and kidney. *Matrix Biol.* 2022;105:31–52. doi: 10.1016/j.matbio.2021.11.003
- 15 Maher TM, Nambiar AM, Wells AU. The role of precision medicine in interstitial lung disease. *Eur Respir J.* 2022;60. doi: 10.1183/13993003.02146-2021
- 16 Leeming DJ, Nielsen MJ, Dai Y, *et al.* Enzyme-linked immunosorbent serum assay specific for the 7S domain of collagen type IV (P4NP 7S): A marker related to the extracellular matrix remodeling during liver fibrogenesis. *Hepatol Res.* 2012;42:482–93. doi: 10.1111/j.1872-034X.2011.00946.x

- 17 Kasuga I, Yonemaru M, Kiyokawa H, *et al.* Clinical evaluation of serum type IV collagen 7S in idiopathic pulmonary fibrosis. *Respirology*. 1996;1:277–81. doi: 10.1111/j.1440-1843.1996.tb00043.x
- 18 Murawaki Y, Ikuta Y, Koda M, *et al.* Comparison of serum 7S fragment of type IV collagen and serum central triple-helix of type IV collagen for assessment of liver fibrosis in patients with chronic viral liver disease. *J Hepatol*. 1996;24:148–54. doi: 10.1016/S0168-8278(96)80023-7
- 19 Jin SW, Kubo S, Tanaka H, *et al.* Type IV collagen 7s domain as a predictor of poor efficacy of portal vein embolization before major hepatectomy. *Surg Today*. 2005;35:41–6. doi: 10.1007/s00595-004-2882-6
- 20 Raghu G, Collard HR, Egan JJ, *et al.* 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. doi: 10.1164/rccm.2009-040GL
- 21 Hoyer N, Prior TS, Bendstrup E, *et al.* Risk factors for diagnostic delay in idiopathic pulmonary fibrosis. *Respir Res*. 2019;20. doi: 10.1186/s12931-019-1076-0
- 22 Raghu G, Remy-Jardin M, Myers JL, *et al.* Diagnosis of idiopathic pulmonary fibrosis An Official ATS/ERS/JRS/ALAT Clinical practice guideline. *Am J Respir Crit Care Med*. 2018;198:e44–68. doi: 10.1164/rccm.201807-1255ST
- 23 Sand JMB, Martinez G, Midjord AK, *et al.* Characterization of serological neo-epitope biomarkers reflecting collagen remodeling in clinically stable chronic obstructive pulmonary disease. *Clin Biochem*. 2016;49:1144–51. doi: 10.1016/j.clinbiochem.2016.09.003
- 24 Kurland BF, Johnson LL, Egleston BL, *et al.* Longitudinal data with follow-up truncated by death: Match the analysis method to research aims. *Stat Sci*. 2009;24:211–22. doi: 10.1214/09-STS293
- 25 Gao F, Philip Miller J, Xiong C, *et al.* Estimating correlation between multivariate longitudinal data in the presence of heterogeneity. *BMC Med Res Methodol*. 2017;17:1–11. doi: 10.1186/s12874-017-0398-1
- 26 Goodrich B, Gabry J, Ali I BS. Bayesian applied regression modeling via Stan. R package version 2.21.4. <https://mc-stan.org/rstanarm>
- 27 Rizopoulos D, Afonso PM, Papageorgiou G. JMBayes2: Extended Joint Models for Longitudinal and Time-to-Event Data. R Packag. version 0.5-0. 2024. <https://cran.r-project.org/package=JMBayes2>
- 28 Gerds TA, Johan Sebastian Ohlendorff, Paul Blanche, *et al.* riskRegression: Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks. R Packag. version 2023.12.21. 2023. <https://cran.r-project.org/package=riskRegression>
- 29 Decato BE, Leeming DJ, Sand JMB, *et al.* LPA1 antagonist BMS-986020 changes collagen dynamics and exerts antifibrotic effects in vitro and in patients with idiopathic pulmonary fibrosis. *Respir Res*. 2022;23:1–12. doi: 10.1186/s12931-022-01980-4
- 30 Raghu G, Striker LJ, Hudson LD, *et al.* Extracellular matrix in normal and fibrotic human lungs. *Am Rev Respir Dis*. 1985;131:281–9. doi: 10.1164/arrd.1985.131.2.281
- 31 Paralkar VM, Vukicevic S, Reddi AH. Transforming growth factor β type 1 binds to collagen IV of basement membrane matrix: Implications for development. *Dev Biol*. 1991;143:303–8. doi: 10.1016/0012-1606(91)90081-D
- 32 de Oliveira RC, Sampaio LP, Shiju TM, *et al.* Epithelial Basement Membrane Regeneration after PRK-

- Induced Epithelial-Stromal Injury in Rabbits: Fibrotic Versus Non-fibrotic Corneal Healing. *J Refract Surg*. 2022;38:50–60. doi: 10.3928/1081597X-20211007-02
- 33 Wilson MS, Wynn TA. Pulmonary fibrosis: Pathogenesis, etiology and regulation. *Mucosal Immunol*. 2009;2:103–21. doi: 10.1038/mi.2008.85
- 34 Parra ER, Pincelli MS, Teodoro WR, *et al*. Modeling pulmonary fibrosis by abnormal expression of telomerase/apoptosis/collagen V in experimental usual interstitial pneumonia. *Brazilian J Med Biol Res*. 2014;47:567–75. doi: 10.1590/1414-431X20143522
- 35 Vaccaro CA, Brody JS, Snider GL. Alveolar wall basement membranes in bleomycin-induced pulmonary fibrosis. *Am Rev Respir Dis*. 1985;132:905–12. doi: 10.1164/ARRD.1985.132.4.905
- 36 Loscertales M, Nicolaou F, Jeanne M, *et al*. Type IV collagen drives alveolar epithelial-endothelial association and the morphogenetic movements of septation. *BMC Biol*. 2016;14:1–21. doi: 10.1186/s12915-016-0281-2
- 37 Kaminski N. The IPF Cell Atlas. <http://www.ipfcellatlas.com/>
- 38 Fainberg HP, Moodley Y, Triguero I, *et al*. Cluster analysis of blood biomarkers to identify molecular patterns in pulmonary fibrosis: assessment of a multicentre, prospective, observational cohort with independent validation. *Lancet Respir Med*. 2024;12:681–92. doi: 10.1016/S2213-2600(24)00147-4
- 39 Richeldi L, du Bois RM, Raghu G, *et al*. Efficacy and Safety of Nintedanib in Idiopathic Pulmonary Fibrosis. *N Engl J Med*. 2014;370:2071–82. doi: 10.1056/nejmoa1402584
- 40 Fainberg HP, Oldham JM, Molyneau PL, *et al*. Forced vital capacity trajectories in patients with idiopathic pulmonary fibrosis: a secondary analysis of a multicentre, prospective, observational cohort. *Lancet Digit Heal*. 2022;4:e862–72. doi: 10.1016/S2589-7500(22)00173-X

Competing interests

JMBS, PF, FBS, MAK, and DJL are employees and shareholders of Nordic Bioscience. NH received honoraria from Boehringer Ingelheim for presentations. TSP received support from Boehringer Ingelheim, TrygFonden, and Aarhus University and payment or honoraria from Boehringer Ingelheim. PLM received grant funding to institution from AstraZeneca, advisory board fees from Hoffman-La Roche, Boehringer Ingelheim, AstraZeneca, Trevi, and Qureight, and speaker fees from Hoffman-La Roche and Boehringer Ingelheim. SRJ received grant funding to institution from Nottingham NIHR Biomedical Research Centre, Medical Research Council, LAM Action, and Morato Foundation and is a trustee for LAM action. EB received consulting fees from Boehringer Ingelheim, payment or honoraria from Boehringer Ingelheim, Daiichi-Sankyo, and AstraZeneca, and support for attending meetings from Boehringer Ingelheim. SBS received payment or honoraria from Novartis and Boehringer Ingelheim and support for attending meetings from Boehringer Ingelheim. TMM received consulting fees from Boehringer Ingelheim, Roche/Genentech, AstraZeneca, Bayer, Blade Therapeutics, Bristol-Myers Squibb, CSL Behring, Galapagos, Galecto, GlaxoSmithKline, IQVIA, Pfizer, Pliant Therapeutics, Respivot, Sanofi, Theravance, Trevi, Veracyte, and Vicore and participated on a data safety monitoring or advisory board for Fibrogen, Blade Therapeutics, and Nerre. RGJ received grant funding to institution from AstraZeneca, Biogen, Galecto, GlaxoSmithKline, Nordic Bioscience, RedX, and Pliant Therapeutics, consulting fees from AstraZeneca, Brainomix, Bristol Myers Squibb, Chiesi, Cohbar, Daewoong, GlaxoSmithKline, Veracyte, Resolution Therapeutics, and Pliant Therapeutics, payments or honoraria from Boehringer Ingelheim, Chiesi, Roche, PatientMPower, and AstraZeneca, payment for expert testimony from Pinsent Masons LLP, is the president of Action for Pulmonary Fibrosis and in the leadership of NuMedii. AEJ, AAB, IDS, and HPF declare no competing interests.

Author contributions

JMBS, NH, MAK, DJL, SBS, TMM, and RGJ designed the study. NH, TSP, PLM, IDS, HPF, SRJ, EB, SBS, TMM, and RGJ recruited study subjects and acquired the clinical data. JMBS, MAK, and DJL planned and undertook the biomarker assessments. PF undertook the statistical analyses. JMBS, PF, FBS, NH, MAK, DJL, SBS, TMM, and RGJ planned and interpreted the statistical analyses. AEJ and AAB undertook and interpreted the immunohistochemistry data. JMBS drafted the manuscript which was revised and approved by all authors. JMBS is the guarantor of this study.

Funding

The PROFILE study was funded by the UK Medical Research Council (grant/award number: G0901226) and GlaxoSmithKline R&D (grant/award number: CRT114316). AEJ is supported by the UK Medical Research Council (grant/award number: MR/V00235X/1). PLM is supported by an Action for Pulmonary Fibrosis Mike Bray fellowship and the Asthma+Lung UK Chair in Respiratory Research. TMM is supported by a National Institute for Health Research Clinician Scientist Fellowship (grant/award number: CS-2013-13-017) and a British Lung Foundation Chair in Respiratory Research (grant/award number: C17-3).

Acknowledgements

This study was supported by the Danish Research Foundation, the UK Medical Research Council, GlaxoSmithKline R&D, and Nottingham University and Royal Brompton and Harefield NHS Foundation Trust. The authors thank all patients and study coordinators for their participation in the PFBIO and PROFILE studies.

Data availability statement

Data are available from the corresponding author on reasonable request.

Figure Legends

Figure 1. Study populations. Patients with pulmonary fibrosis were included from the PFBIO (A) and PROFILE (B) cohorts. Patients were excluded from analysis due to missing PRO-C4 biomarker measurements or forced vital capacity (FVC) data at baseline. Included patients were stratified according to progression status at 12 months. Indeterminate patients in the PFBIO cohort were alive at 12 months but with unknown FVC status.

Figure 2. PRO-C4 trajectories for progressive and non-progressive disease. Serum PRO-C4 measured at baseline, six and 12 months in the PFBIO cohort (A) and plasma PRO-C4 measured at baseline, one, three, six, and 12 months in the PROFILE cohort (B). Data are shown as least square means with 95% confidence intervals adjusted for age and sex (and anti-fibrotic treatment in PFBIO). The differences between progressors and non-progressors averaged over all visits were 21.5% (95% CI 3.4-42.9%; $P=0.0184$) in PFBIO and 10.9% (95% CI 0.8-22.1%; $P=0.0340$) in PROFILE.

Figure 3. Contour plots of monthly rate of change in FVC%, DLCO% and PRO-C4. Contour plot of the joint posterior distribution of the percentage monthly rate of change in plasma PRO-C4 levels and FVC% (A) or DLCO% (B). Posterior means and 95% credible intervals of average monthly rate of change by survival is indicated by coloured dots and error bars. The contours (blue lines) indicate variability of slopes at the patient level. FVC%, forced vital capacity in percent of predicted value; DLCO%, diffusing capacity for carbon monoxide in percent of predicted value.

Figure 4. Risk of mortality by baseline PRO-C4 levels. Kapan-Meier survival curves are shown for baseline PRO-C4 quartiles in PFBIO (A) and PROFILE (B). Hazard ratios were estimated in Cox regression models using restricted cubic splines (3 knots) to illustrate the relationship between PRO-C4 levels and three-years mortality risk (adjusted for age, sex, and baseline forced vital capacity) in PFBIO (C) and PROFILE (D). Non-linearity effects of PRO-C4 on hazard ratios were examined by Wald tests (1 degree of freedom) and were not identified in any of the cohorts (PFBIO: $P=0.7730$, PROFILE: $P=0.1353$).

Figure 5. Type IV collagen staining in IPF and non-IPF lung tissue. Type IV collagen was stained by immunohistochemistry in the lungs of IPF patients ($n=12$) and in non-IPF lung tissue ($n=3$). (A) FFPE tissue samples were stained with antibody against type IV collagen (brown). (B) Parallel sections were subjected to Masson's trichrome staining in which fibrillar collagen/connective tissues were stained in bright blue. Representative sections from IPF (Pt1-3) and non-IPF lung are shown at x2 magnification (A and B) and at x20 magnification (C). Boxes denote magnified region shown in row C. Alveolar epithelium is indicated by black arrows. Blood vessels are denoted with black arrowheads. COL4, type IV collagen; IPF, idiopathic pulmonary fibrosis; FFPE, formalin-fixed paraffin embedded.