

**Title:**

Identification of an epithelial biomarker signature in Idiopathic Pulmonary Fibrosis: an analysis from the prospective multicentre PROFILE study.

**Authors:**

Toby M Maher<sup>1,2</sup>, Eunice Oballa<sup>3</sup>, Juliet Simpson<sup>3</sup>, Joanne Porte<sup>4,8</sup>, Anthony Habgood<sup>4,8</sup>, William A Fahy<sup>3</sup>, Aiden Flynn<sup>3</sup>, Philip Molyneaux<sup>1,2</sup>, Rebecca Braybrooke<sup>5,9</sup>, Hrushikesh Divyateja<sup>6</sup>, Helen Parfrey<sup>7</sup>, Doris Rassl<sup>8</sup>, Anne-Marie Russell<sup>1,2</sup>, Gauri Saini<sup>4,9</sup>, Elisabetta A Renzoni<sup>1,2</sup>, Anne-Marie Duggan<sup>3</sup>, Richard Hubbard<sup>5,9</sup>, Athol U Wells<sup>1,2</sup>, Pauline T Lukey<sup>3</sup>, Richard P Marshall<sup>3</sup>, R Gisli Jenkins<sup>4,9</sup>

**Affiliations**

- 1 NIHR Respiratory Biomedical Research Unit, Royal Brompton Hospital, London, UK.
- 2 Fibrosis Research Group, National Heart and Lung Institute, Imperial College, London, UK.
- 3 Fibrosis Discovery Performance Unit, GlaxoSmithKline R&D, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage SG1 2NY, UK.
- 4 Respiratory Research Unit, Division of Respiratory Medicine, University of Nottingham, Nottingham, UK.
- 5 Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK.
- 6 Clinical Pathology, Nottingham University Hospitals, Nottingham, UK
- 7 Respiratory Medicine, Papworth Hospital, Cambridge, UK
- 8 Department of Pathology, Papworth Hospital, Cambridge, UK
- 9 National Institute for Health Research, Nottingham Biomedical Research Centre, Nottingham University Hospitals, Nottingham, UK

## **Corresponding Author**

R Gisli Jenkins

Respiratory Research Unit, Nottingham University Hospitals NHS Trust, Nottingham NG5 1PB

E-mail [gisli.jenkins@nottingham.ac.uk](mailto:gisli.jenkins@nottingham.ac.uk) Tel: +44 115 823 1711 Fax +44 115 823 1946

Abstract: 282 words

Body Text: 3844 words

5 Figures

**Keywords:** Interstitial lung disease, clinical trials, biomarker,

# Research in Context

## Evidence Before the Study

We searched PubMed for reports published up to Jul 20 2017, using the search terms “Idiopathic Pulmonary Fibrosis”, “biomarkers”, “surfactant protein D”, “matrix metalloproteinase”, “CA-125”, and “CA19-9”. No language restrictions were applied. At the time of the search there had been no biomarker studies that had undertaken a two-stage hypothesis free discovery and targeted validation design. Similarly only 1 study (our prior report from the PROFILE study) has previously assessed longitudinal change of biomarkers in a treatment naïve cohort of patients with IPF. Although tumour markers have been assessed in patients with IPF, no studies have identified the tumour marker CA-125 as a dynamic biomarker or CA19-9 as a biomarker of disease progression for IPF.

## Added value of this study

This is the largest study to prospectively and systematically assess the role of known and novel serum biomarkers in Idiopathic Pulmonary Fibrosis (IPF). Furthermore, the two stage discovery and validation design; the use of two distinct assays; prospective, systematic longitudinal sample measurement; assessment of over 300 patients with long term follow data; appropriately powered sample size for an a priori defined outcome; thorough evaluation of four epithelial-derived biomarkers; represent a definitive step-change over the currently available literature in terms of both the methodological approach and the resulting findings. Thus, this study clarifies the role of known biomarkers and highlights the importance of 2 previously unvalidated biomarkers for IPF.

## Interpretation

These data represent the largest prospective analysis of serum biomarkers in IPF. The two stage design permitted us to adopt an unbiased approach to identifying the most promising biomarkers based on their ability to 1) distinguish individuals with disease from age and gender matched healthy controls 2) identify individuals with IPF who have progressive disease and 3) identify changes in biomarkers over time to predict death as an outcome, that may therefore have potential as theranostic biomarkers. The most promising biomarkers in the discovery set were independently validated using singleplex ELISA's including those available in routine clinical practice. These results identified epithelial derived proteins as promising biomarkers in all three categories. Validation analysis confirmed that high levels of baseline Surfactant protein D and MMP7 distinguishes individuals with disease from controls and are predictive of outcome. In contrast high levels of CA19-

9 were able to predict 12-month disease progression whilst rising levels of CA-125 after three months were predictive of 12 month-disease progression, and overall mortality. Changes in these tumour markers were independent of baseline physiological parameters. These data identify biomarkers of IPF that could be used to streamline clinical trial design, identify individuals at high risk of progression and to assess clinical response to therapy.

## Abstract

**Background:** Idiopathic Pulmonary Fibrosis (IPF) is a progressive, fatal condition with a variable disease trajectory. The aim of this study was to evaluate potential biomarkers that predict outcome for people with IPF.

**Method:** The PROFILE study is a large prospective longitudinal cohort of treatment naïve IPF patients. We adopted a two-stage discovery and validation design using the PROFILE cohort. For the discovery analysis 106 individuals were examined alongside 50 age and gender matched healthy controls. We undertook an unbiased, multiplex evaluation of 123 biomarkers. Promising, novel, markers were further evaluated by immunohistochemical assessment of IPF lung tissue. The validation analysis examined samples from 206 IPF subjects, from the remaining 212 IPF patients recruited to PROFILE Central England, and were used for replication of the biomarkers identified from the discovery analysis using singleplex assays. This study addressed the predictive power of selected biomarkers to identify individuals with IPF at risk of: 1) progression and 2) death. The PROFILE studies are registered on clinicaltrials.gov (PROFILE Central England NCT01134822; PROFILE Royal Brompton Hospital NCT01110694).

**Findings:** The discovery analysis identified four serum biomarkers (Surfactant Protein D, Matrix Metalloproteinase 7, CA19-9 and CA-125) suitable for replication. Histological assessment of CA19-9 and CA-125 established these proteins as markers of epithelial damage. Replication analysis confirmed that baseline values of SP-D (46.6ng/ml vs 34.6 ng/ml;  $p = 0.002$ ) and CA19-9 (53.7 U/ml vs 22.2 U/ml  $p < 0.001$ ) were significantly higher in patients with progressive disease, and rising levels of CA-125 over 3 months were associated with increased risk of mortality (HR 2.542 CI 1.493-4.328  $p < 0.001$ ).

**Interpretation:** We have identified serum proteins secreted from metaplastic epithelium that predict disease progression and death in IPF.

**Funding:** GlaxoSmithKline R&D and the Medical Research Council

## Introduction

Idiopathic Pulmonary Fibrosis (IPF) is a progressive, fatal condition with a variable disease trajectory. The advent of effective therapy for IPF has generated an urgent need to identify biomarkers of multiple aspects of disease behaviour, particularly those that may be used to stratify therapy, enable the delivery of personalised medicine and which provide robust and reliable measures of response to anti-fibrotic therapy.

The current accepted pathogenic paradigm suggests that IPF occurs in genetically susceptible individuals following repeated or persistent epithelial injury. This in turn activates fibroblasts that display an invasive phenotype and secrete abundant extra-cellular matrix proteins resulting in the progressive parenchymal destruction characteristic of IPF<sup>1</sup>.

A number of serum biomarkers have been consistently described in case-control and point of diagnosis studies including markers reflective of epithelial dysfunction; matrix metalloproteinase (MMP)-7, surfactant protein-D (SP-D) and Krebs von den Lungen (KL)-6 (MUC1). Putative inflammatory molecules such osteopontin and CCL18 have also been identified as potential prognostic biomarkers<sup>2</sup>. More recently we described biomarkers of matrix turnover which, when measured longitudinally, identify individuals with increased risk of disease progression and death<sup>3</sup>.

In the current study, by utilising a two-stage discovery and validation analysis from the PROFILE study cohort, we sought to identify potential biomarkers of prognosis and disease progression in IPF. This is the first study to adopt such an approach to biomarker discovery in a prospectively enrolled, longitudinally sampled cohort of IPF patients. In so doing we report a novel epithelial biomarker signature predictive of prognosis, disease progression and risk of death in IPF.

# Methods

## Participants

The PROFILE study is a prospective, multi-centre, observational cohort study of incident cases of fibrotic interstitial lung disease<sup>3</sup>. Participants with IPF or idiopathic NSIP were identified through two co-ordinating centres: Nottingham, UK and, Royal Brompton Hospital (RBH), London, UK. The PROFILE study is registered on clinicaltrials.gov (NCT01134822 and NCT01110694). These independent cohorts were set up to reflect the different referral patterns throughout the UK, with RBH recruiting patients following tertiary referrals, and Nottingham co-ordinating predominantly primary care referrals to regional hospitals. This ensures that the findings are generalizable across a broad population and facilitates replication in two separate cohorts. The human biological samples were sourced ethically and their research use was in accord with the terms of the informed consents.

## Study design

A two-stage discovery analysis of 106 samples from multidisciplinary team (MDT)-confirmed IPF cases and 50 age and gender matched healthy control and validation with 206 samples taken from MDT-confirmed IPF cases was undertaken. (See on-line supplement for details)

## Discovery Analysis

Concentrations of 123 cytokines (Supplementary table 1), chemokines, growth factors, MMPs, extracellular matrix proteins and markers of epithelial injury and apoptosis were analysed using a range of Human Discovery multiplex bead-based immunoassays (Myriad Rules-Based Medicine, Austin, Texas, USA). Proteins were excluded from analyses if greater than 95% of values were outside the upper and/or lower limits of detection of the assay.

## **Validation Analysis**

Concentrations of four biomarkers, identified from the discovery analysis, were measured in serum using independent assays for each biomarker. For details see on-line supplement

## **Immunohistochemistry**

Lung tissue from patients with IPF or non-IPF controls was analysed by immunohistochemistry using standard methodology<sup>4</sup> (see online supplement for details).

## **Statistical analyses**

Disease progression was defined as all-cause mortality within the first year, or  $\geq 10\%$  decline in FVC at 12 months. In cases where 12 month FVC data were absent, subjects were considered to have progressed if a  $\geq 10\%$  decline in FVC was observed at any time within 6 months. Where no lung function data were available beyond baseline, cases were adjudicated, following case note review, by the local principal investigator blinded to biomarker results. Missing lung function data were not imputed. Where biomarker data were below the lower limit of detection, values were imputed to be half the lower limit of detection. Values above the upper limit were conservatively imputed as the upper limit of detection. Except where stated the percentage of imputed data was below 1% of all measurements for each biomarker. Sensitivity analyses were performed to ensure that data imputation did not alter the impact of the reported findings. Overall mortality was determined using censor dates of 31<sup>st</sup> October 2014 and 4<sup>th</sup> April 2016 for the discovery and validation cohorts respectively.

A power calculation was conducted using data obtained at three-months in the discovery cohort using a two sample t-test of the difference in means for levels of MMP7 in progressive versus stable disease. MMP7 was selected as the most conservative of the four biomarkers for replication, having the lowest threshold for biomarker change and least statistical significance, thus powering on MMP7 would ensure the other biomarkers would be analysed with adequate power. This demonstrated that a sample size of 100 patients per outcome group would be adequate to detect a statistically

significant difference, with 80% power at an alpha level of 5%. Therefore the validation analysis was conducted on 100 patients per group.

Prior to performing any association analysis, relevant variables were transformed to hold normality assumption using base 10 or base 2 logarithms.

When comparing baseline biomarker values between healthy controls and IPF groups adjusted estimates of group effect were obtained using a general linear model with age, gender and group as explanatory variables. Baseline characteristics including age, gender and baseline lung function were summarised by cohort. To control the false discovery rate that could result from multiple comparisons the Benjamini Hochberg procedure was applied to the discovery analysis. Biomarkers with >95% imputation were excluded from analysis. When there was between 50-95% imputation the biomarker result was treated as binary, based on whether the result was imputed or not and an alternative analysis model was used suitable for binary data as opposed to continuous data.

A mixed-effects model with repeated measures was used to evaluate association between progression status, visit, visit by progression status and the continuous biomarker value, and to provide an estimate of the effect of the progression status group, least square means adjusted for the effects of age, gender, site and smoking status and the significance of progression status by visit.

The impact of biomarker result on mortality was thoroughly investigated using univariate analyses to assess the correlation between biomarker baseline levels and overall survival and a proportional hazards model to evaluate association between time-to-death or censoring. The gradient of the biomarker over three months was first treated as a continuous explanatory variable, and a separate analysis used a binary version of gradient dichotomised by whether this was rising or falling.

Receiver Operating Characteristics (ROC) curve analysis was conducted to select a cut-point using Youden's J Statistic which maximises sensitivity and specificity. Hazard ratios were generated comparing patient groups assigned based on the optimal cut-point

## **Role of the funding source**

The PROFILE study was funded by the Medical Research Council (G0901226) and GlaxoSmithKline R&D (CRT114316), and was sponsored by Nottingham University and Royal Brompton and Harefield NHS Foundation Trust. Investigators received no financial incentives from the funding sources. GlaxoSmithKline R&D participated in the study design, coordination of biomarker analysis, and statistical analysis. TMM and RGJ were involved in all stages of study development and delivery, had full access to all data in the study, and had final responsibility for the decision to submit the publication.

# Results

## Patient cohorts

106 patients with an MDT-confirmed diagnosis of IPF were included in the discovery analyses. Of the remaining 252 eligible subjects recruited to PROFILE Central England 206 had a confirmed diagnosis of IPF and were included in the validation analysis (See Figure 1). Subjects in both the discovery and validation cohorts were predominantly male (78.3% discovery, 76.7% validation) with a mean ( $\pm$  S.D.) age of 70.8 $\pm$ 8.33 (discovery) and 72.5 $\pm$ 7.66 (validation) years. 103 participants in the discovery cohort had baseline FVC data and 76 had 12-month FVC measurements available. 28 of the remaining 30 subjects with missing 12-month FVC had sufficient clinical information available for a progression status to be assigned. Repeated measures of FVC were unavailable due to death in 15 (15%) cases and progression adjudicated in 2 cases (2%). Overall survival was determined at the date of censoring, when 51 (48.1%) of subjects had died. 202 participants in the validation cohort had baseline FVC data and 139 had available 12-month FVC measurements. 65 of the remaining 67 subjects missing 12-month FVC had sufficient clinical information available for a progression status to be assigned. Repeated measures of FVC were unavailable due to death in 35 cases and progression adjudicated in 2 cases (2%). 19 patients received pirfenidone in the first 12 months of the study; 9 with progressive and 10 with stable disease. At the date of censoring 92 (44.6%) of subjects had died. Demographics for the discovery and validation cohorts and the 50 gender-matched control subjects are given in Supplementary Table 2.

## Discovery analysis

For the initial discovery analysis 123 serum proteins were measured in 106 IPF subjects and 50 controls. Of these proteins, 23 had greater than 95% of measures imputed and were excluded from subsequent analyses, 16 had greater than 50% of measures imputed and were treated as categorical variables, 84 had less than 50% of measures imputed and were treated as continuous variables. The

proteins that most clearly distinguished IPF subjects from controls were surfactant protein D (SP-D), alpha2 macroglobulin, matrix metalloproteinase (MMP)-7, T-cadherin, AXL receptor tyrosine-kinase and MMP-9 (Figure 2A; supplementary table 3). When correcting for possible false discovery using the Benjamini-Hockberg 44 proteins continued to distinguish IPF from healthy controls (Figure 2B; supplementary table 3). To assess the relationship between baseline biomarker concentrations and subsequent outcome, IPF subjects were dichotomised into those with stable (n=54) or progressive (n=50) disease. The protein that most clearly distinguished progressive from stable disease was CA19-9 (see Figure 2C) and it was the only biomarker that remained significantly elevated following multiplicity correction (Figure 2D; Supplementary Table 4). To determine the relationship between changing levels of protein biomarkers with outcome, the rate of change in biomarker levels between baseline and month 3 were calculated, and 10 proteins including increasing CA-125, Macrophage migration inhibitory factor (MIF), Carcinoembryonic Antigen (CEA), free Prostate-Specific Antigen (PSAf) and MMP7 were associated with a significant increase in overall mortality (See figure 2E and 2F; supplementary tables 5).

### **Choice of biomarkers for validation**

Assessment of the discovery cohort analyses suggested that SP-D, CA19-9 and CA-125 were the most discriminatory biomarkers, based on having the greatest statistical significance for biomarkers reaching a threshold of 75% increase in disease versus control, progression versus stable disease and rate of change predicting mortality respectively. MMP7 appeared to have predictive values in two of these categories and, along with SP-D, has been one of the most frequently studied epithelial biomarker in IPF<sup>2</sup>. Although CA19-9 and CA-125 have been linked with IPF in a small number of studies<sup>5-7</sup> the cellular source of these markers in the IPF lung has not been fully evaluated. To ensure disease relevance of these markers we evaluated the immunohistochemical localization of CA19-9 and CA-125 in control and fibrotic lung tissue (See Figure 3). In normal lungs CA19-9 and CA-125 were only observed in the apical aspect of bronchial epithelium, without any expression in alveolar

epithelium (See Figure 3A and B). In fibrotic lung tissue there was an increase in CA19-9 and CA-125 throughout the metaplastic epithelium in fibrotic lesions. Furthermore, this was associated with mucous secretion that was particularly apparent within honeycomb cysts (Figure 3C and D). Rising levels of CEA were associated with increased mortality in the discovery cohort, and although not meeting the criteria for replication, the analysis platform for CA-125 and CA19-9 also included CEA (see Supplementary Figure 1).

## **Validation Cohort**

Levels of SP-D were significantly elevated in patients with progressive IPF compared with stable disease, however there was little change in levels over time (See Figure 4A). Levels of MMP7 and CEA did not significantly predict progressive disease, nor did they change over time (See Figure 4B; Supplementary Figure 1). Levels of CA19-9 were substantially elevated in patients with progressive disease, whereas levels were within the 'normal' range for stable disease (see Figure 4C). Levels of CA-125 similarly discriminated between stable and progressive disease at baseline and, in marked contrast with the other biomarkers, levels of CA-125 increased over three months in progressive disease (see Figure 4D). Reassuringly, findings in the validation cohort were consistent with those observed in the discovery cohort (Supplementary figure 2). When only paired samples for CA-125 were analysed, patients with stable disease had a mean of 20.16 U/mL that reduced to 18.6 U/mL at three months whereas patients with progressive disease had a baseline value of 21.46 U/mL that increased to 26.14 U/mL, and this mean change of 4.7 U/mL in patients with progressive disease was significantly greater than for patients with stable disease (-1.6 U/mL:  $p < 0.01$ ).

Baseline levels of all CA-125, CA19-9, and SP-D were associated with increased mortality in both discovery and validation cohorts, whereas baseline levels of all four biomarkers were associated with increased mortality in the validation cohort (Supplementary table 6). Receiver Operator Characteristic (ROC) curve defined thresholds, performed on the validation cohort, revealed that levels of CA-125 of 12 U/mL or above and CA19-9 22 U/mL or above were associated with a three-

fold increased risk of mortality, whereas MMP7 and SP-D were associated with a 2.4 and 2.7 fold increased risk of mortality (Supplementary Figure 3). However, to identify whether the four chosen epithelial biomarkers change dynamically in response to progression of fibrosis each marker was dichotomised to separate individuals into those with either rising or stable/falling levels over three months based on paired samples at 0 and 3 months. Rising levels of CA-125 were significantly associated with overall mortality (HR2.542 (CI 1.493-4.328;  $p < 0.001$ : see Figure 5A) even when adjusted for baseline FVC (HR2.390 (CI 1.400-4.078;  $p = 0.001$ ) or DLco (HR2.673 (CI 1.477-4.840;  $p = 0.001$ : Supplementary Figure 4 and 5 respectively). There was a trend to rising levels of CA19-9 associating with mortality (HR 1.707 (CI 0.985-2.959;  $p = 0.057$ : Figure 5B) although this reduced when adjusted for FVC or DLco (Supplementary Figure 4 and 5). However, there was no association between rising levels of MMP7 HR 1.184 (CI 0.706-1.985;  $p = 0.523$ ) or SP-D and overall mortality (HR 1.096 (CI 0.651-1.844;  $p = 0.731$ : see Figure 5C-D).

## Discussion

This analysis from the PROFILE study systematically assessed a large number of serum proteins using a two-stage experimental approach in patients with incident IPF, initially undertaking a discovery analysis using a multiplex platform, followed by specific replication with individual ELISAs for the most promising candidate biomarkers. This robust approach to biomarker evaluation identified three biomarkers, SP-D, CA19-9 and CA-125, which have potential clinical utility based on their performance characteristics. These three biomarkers are all secreted at low levels by the pulmonary epithelium in health, with SP-D being secreted by alveolar type 2 cells and CA19-9 and CA-125 being secreted by the bronchial epithelium. However, all are secreted in abundance by the metaplastic epithelium associated with IPF.

Our data confirm previous studies demonstrating that MMP7, an epithelial secreted matrix metalloproteinase, is raised in IPF compared with healthy controls<sup>8-10</sup>. In contrast with other studies, baseline values, when considered as a continuous variable, failed to consistently predict progressive disease or increased mortality. However, using a ROC defined threshold to dichotomise subjects in to MMP7-high and –low groups produced findings consistent with previous reports that high levels of MMP7 are associated with poorer survival<sup>10,11</sup>, albeit with only moderate sensitivity and specificity. Nonetheless, the totality of our data would suggest MMP7 is inferior to baseline lung function or biomarkers with superior discrimination such as SP-D, CA-125 and CA19-9.

Surfactant Protein D is a protein that, in health, is secreted by alveolar type 2 cells, with only low levels normally detected in plasma<sup>6</sup>. Levels were substantially raised in IPF, with higher levels significantly associating with progressive disease and increased overall mortality consistent with previous reports<sup>12,13</sup>. However, levels of SP-D remained fairly constant over three months and rising levels did not alter risk of mortality. Furthermore, previous studies have suggested that SP-D levels may distinguish IPF from other interstitial lung diseases<sup>9</sup>. These data suggest that SP-D is a useful baseline marker of IPF and can identify individuals more likely to progress. Therefore, baseline

measurements of SP-D may be particularly useful in primary care to aid referral decisions where the diagnosis IPF may be uncertain.

Of particular interest was the observation that raised levels of the tumour marker CA19-9 were highly predictive of progressive fibrosis and rising levels of CA-125 predicted both disease progression and overall survival. In these studies patients with progressive disease had approximately a 25% increase in CA-125 levels at 3 months, over two times the lower limit of detection (2U/mL) and substantially greater than the 4% coefficient of variation of the assay. Previous studies have highlighted CA19-9 and CA-125 in patients with interstitial lung disease. Mean levels of CA19-9 and CA-125 were elevated in 61 patients with IPF compared with COPD<sup>6</sup>, and levels of CA19-9 above 37 U/mL have been associated with worse survival in 30 patients with ILD<sup>5</sup>. Interestingly, levels of CA19-9, although not CA-125, have been noted to reduce in patients with IPF following lung transplantation<sup>6</sup>, but no prior study has demonstrated a dynamic increase in tumour markers predicting disease progression.

Both these biomarkers are mucous associated antigens. CA19-9 is a tumour associated carbohydrate antigen secreted by mucin secreting tumours, particularly pancreatic cancer<sup>14,15</sup>, whereas CA-125, also known as MUC16, is a transmembrane mucin commonly used as an ovarian cancer biomarker<sup>16</sup>. IPF has traditionally been considered a disease of alveolar type 2 cells, however, an emerging role for bronchial epithelium has been described<sup>17,18</sup>. Additionally, the mucin gene MUC5B, expressed in bronchial epithelium, is the strongest identified genetic risk factor for IPF<sup>19</sup>. An intriguing possibility is that mucin related biomarkers are surrogates for MUC5B genotype and this is the subject of ongoing analysis. However, we believe it is unlikely that the biomarkers map precisely to MUC5B genotype because: a) MUC5B is a risk allele for IPF whereas CA-125 and Ca19-9 do not discriminate between disease and controls b) the MUC5B polymorphism has a small protective effect on prognosis whereas high levels of CA-125 and Ca19-9 imply a substantially worse prognosis. Furthermore, genotype does not change over time whereas CA-125 levels rise in progressive disease.

However, taken together these data suggest that mucin secreting epithelial cells play a greater role in the pathogenesis of IPF than previously realised.

Recent work has focused on targeting CA-125/MUC16 therapeutically<sup>20,21</sup>, and intriguingly MUC16 is known to bind galectins, including Galectin-3, on epithelial surfaces<sup>22</sup>. Galectin-3 is a recognised therapeutic target in IPF<sup>23</sup>, raising the prospect that CA-125 may potentially be a useful theranostic biomarker for Galectin inhibitors currently under development (NCT02257177). More speculatively, novel CA-125 targeting therapies could be considered for assessment in IPF.

Our observation that increased tumour markers are predictive of mortality in IPF could reflect occult, or undiagnosed tumours in our study cohort, and indeed the high levels of CA-125 and CEA have been associated with an increased risk of cancer in patients with ILD<sup>7</sup>. However, we believe that this is unlikely due to the large number of patients expressing high levels of CA-125 and CA19-9, and the high levels of immunostaining observed in the fibrotic lesions. Furthermore, Individuals with IPF are known to have shared risk factors and an increased risk of lung cancer<sup>24,25</sup>, largely explaining the excess cancer risk in IPF patients<sup>24</sup>. Furthermore, the tyrosine kinase inhibitor nintedanib has been approved for the treatment of both IPF and lung cancer implicating shared mechanisms of pathogenesis<sup>26-28</sup>. We, therefore hypothesise that the striking expression of tumour biomarkers by metaplastic epithelium contributes to the serum levels of these markers in progressive IPF and reflects shared mechanisms between fibrogenesis and oncogenesis<sup>29</sup>.

The strengths of this study relate to the two-stage design, the longitudinal analysis of incident cases and the identification of a biomarker signal reflective of underlying disease pathology. Lung function test are often missed in prospective studies because patients feel unwell, choose not to do lung function measurements or they die. This occurs in all studies including prospective interventional clinical trials where approximately 20% of lung function data are missing. In the ASCEND<sup>30</sup> and CAPACITY<sup>31</sup> studies missing data were imputed, and in the INPULSIS<sup>26</sup> studies missing data were adjusted for measurement of an annualized change in spirometry. Our prospective study design, the

use of a composite for progression, and blinded assignment of progression status avoids the critical problem of survival bias that afflicts retrospective studies. The real potential for serum biomarkers in IPF management are as minimally invasive assessments that can be used to track change in disease behaviour over time. Furthermore, the dynamic relationship between CA-125 and currently accepted measures of disease progression suggest that CA-125 has potential as a theranostic marker of response to anti-fibrotic therapy. The two-stage discovery and validation design ensured that initial observations from the discovery platform could be independently validated in separate patients in appropriately powered studies using specific assays optimised for the protein of interest. Similarly, measuring biomarkers that reflect specific aspects of pathology<sup>3,31</sup> may identify specific endotypes of disease with the potential to direct personalised approaches to therapy<sup>32</sup>.

The primary limitations of this study relate to the lack of an independent external replication cohort. However, as far as we are aware this is the only longitudinal study of treatment naïve patients with multiple biological samples available and therefore external replication is not currently possible. A further limitation is possible confounding in the validation cohort due to a small number of patients receiving anti-fibrotic therapy during the course of the study. However, only three patients receiving pirfenidone had samples available for three-month analysis, and it is therefore unlikely that the biomarker values have been affected by therapy. Whilst it is possible that pirfenidone may have altered the progression status of patients following initial biomarker assessment the numbers were less than 10% of the total cohort and split evenly between stable and progressive disease. It is also reassuring that the data from the validation cohort and discovery cohorts generated similar results, suggesting minimal confounding due to pirfenidone therapy. However, the small numbers of patients receiving anti-fibrotic therapy does mean it is not possible to determine how useful these biomarkers will be at predicting response to therapy. We are reassured that the results will be generalizable not least because we have replicated previously observed point of diagnosis biomarker data.

In conclusion, this study identifies a number of epithelial derived biomarkers that distinguish both stable from progressive IPF and those individuals at increased risk of mortality. Furthermore, these markers show disease relevance, reflecting histopathological abnormalities associated with IPF. In addition, elevated levels of CA-125 are associated with severe disease and change dynamically with progression of fibrosis suggesting it may have potential as a theranostic biomarker. This is a particularly exciting observation given that CA-125 is a clinical grade assay that is widely and routinely available in biochemistry laboratories throughout the world. The identification of a panel of biomarkers reflecting epithelial injury, combined with our previous study identifying markers of matrix turnover, raises real hope for precision approaches to treatment of IPF in an era of multiple therapeutic options.

## **Contributors**

TMM, EO, JS, WAF, RH, PTL, RPM, RGJ. Participated in design and planning of the study. JP, AH, HD Data generation including performing immunohistochemistry and biochemical assays. RB, GS, PM, EAR, HP, DR, AMR, AUW Participated in recruitment of study patients and collection of their data. TMM, EO, JS, WAF AMD, AF, RGJ undertook the data analysis and prepared the report for publication.

## **Declarations of Interest**

TMM has received industry-academic funding from GlaxoSmithKline R&D, UCB and Novartis and has received consultancy or speakers fees from Astra Zeneca, Bayer, Biogen Idec, Boehringer Ingelheim, Cipla, GlaxoSmithKline R&D, Lanthio, InterMune, ProMetic, Roche, Sanofi-Aventis, Takeda and UCB. RGJ has received industry-academic funding from Galecto, GlaxoSmithKline R&D, MedImmune, Novartis and Biogen and has received consultancy or speakers fees from Biogen, Boehringer

Ingelheim, GlaxoSmithKline R&D, InterMune, MedImmune, PharmAkea, Pulmatrix, and Roche. PTL, RPM, EO, WF, JKS and AF are all employees and shareholders of GlaxoSmithKline R&D.

## **Acknowledgements**

The PROFILE study was funded by the Medical Research Council (G0901226) and GlaxoSmithKline R&D (CRT114316), and was sponsored by Nottingham University and Royal Brompton and Harefield NHS Foundation Trust. Additionally, Dr Toby Maher is supported by an NIHR Clinician Scientist Fellowship (NIHR Ref: CS-2013-13-017). Dr Gisli Jenkins has received support from the MRC (MICA grant G0901226). We are grateful to all patients for their participation in the PROFILE study and to the Brompton/Imperial NIHR BRU for infrastructure support in the collection and processing of samples. Tissue used in this study was obtained from the Papworth Hospital Research Tissue Bank and the Nottingham University Hospitals Biobank.

## Figure Legends

**Figure 1:** PROFILE Epithelial Biomarker Study Consort Diagram

**Figure 2:** Multiplex Discovery Analysis. A) Fold Change of biomarker in IPF patients compared with Healthy control. B) False Discovery Rate of biomarker in IPF patients compared with healthy controls using Benjamini-Hochberg procedure. C) Ratio of biomarker in progressive IPF patients compared with stable IPF patients. D) False Discovery Rate of biomarker comparing stable IPF patients with IPF patients with progressive disease using Benjamini-Hochberg procedure. 2 biomarkers were significantly different with Brain-Derived Neurotrophic Factor marginally reduced. E) Risk of mortality in patients with rising levels of biomarkers after 3 months of compared with stable or falling levels of biomarkers. F) False Discovery Rate of biomarker comparing effect of rising levels of biomarkers of three months versus stable levels of biomarker on overall survival.

**Figure 3** Immunohistochemistry of lung tissue obtained from: A) control patients stained with anti-CA-125 antibody; B) IPF patients stained with anti-CA-125 antibody; C) control patients stained with anti-CA19-9 antibody and D) IPF patients stained with CA19-9 antibody.

**Figure 4:** Serum levels of four biomarkers assessed in the validation cohort in patients with progressive disease (red bars) compared with patients with stable disease (blue bars) assessed at baseline and three months A) SP-D B) MMP7 C) CA19-9 and D) CA-125. Mean plus standard deviation shown.

**Figure 5:** Kaplan Meier curves demonstrating risk of overall mortality in relation to rising biomarker levels at 3 months (Red line) compared with stable or falling levels of biomarkers at 3 months (Blue line). A) Patients with rising levels of CA-125 at 3 months had an increased risk of mortality compared with patients with stable or falling levels of CA-125 (HR 2.542 (CI 1.493-4.328;  $p < 0.001$ ); B) Patients with rising levels of CA19-9 at 3 months had a trend towards increased mortality compared with stable or falling levels of CA19-9 (HR 1.707 (CI 0.985-2.959;  $p = 0.057$ )). There was no difference in survival between patients with rising or stable levels of MMP7 C) HR 1.184 (CI 0.706-1.985;  $p = 0.523$ ) or SP-D D) HR 1.096 (CI 0.651-1.844;  $p = 0.731$ ). Overall survival was adjusted for age, gender and smoking status.

## References

- 1) Wolters PJ, Collard HR, Jones KD. Pathogenesis of idiopathic pulmonary fibrosis. *Annu Rev Pathol.* 2014;9:157-79. doi: 10.1146/annurev-pathol-012513-104706.
- 2) Ley B, Brown KK, Collard HR. Molecular biomarkers in idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol.* 2014 Nov 1;307(9):L681-91. doi: 10.1152/ajplung.00014.2014.
- 3) 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 Jun;3(6):462-72. doi: 10.1016/S2213-2600(15)00048-X.
- 4) Saini G, Porte J, Weinreb PH, et al.  $\alpha\beta 6$  integrin may be a potential prognostic biomarker in interstitial lung disease. *Eur Respir J.* 2015 Aug;46(2):486-94. doi:10.1183/09031936.00210414.
- 5) Kodama T, Satoh H, Ishikawa H, Ohtsuka M. Serum levels of CA19-9 in patients with nonmalignant respiratory diseases. *J Clin Lab Anal.* 2007;21(2):103-6.
- 6) Rusanov V, Kramer MR, Raviv Y, Medalion B, Guber A, Shitrit D. The significance of elevated tumor markers among patients with idiopathic pulmonary fibrosis before and after lung transplantation. *Chest.* 2012 Apr;141(4):1047-1054. doi: 10.1378/chest.11-0284.
- 7) Dai H, Liu J, Liang L, et al. Increased lung cancer risk in patients with interstitial lung disease and elevated CEA and CA125 serum tumour markers. *Respirology.* 2014 Jul;19(5):707-13. doi: 10.1111/resp.12317.
- 8) Rosas IO, Richards TJ, Konishi K, et al. MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Med.* 2008 Apr 29;5(4):e93. doi: 10.1371/journal.pmed.0050093.

- 9) White ES, Xia M, Murray S, et al. Plasma Surfactant Protein-D, Matrix Metalloproteinase-7, and Osteopontin Index Distinguishes Idiopathic Pulmonary Fibrosis from Other Idiopathic Interstitial Pneumonias. *Am J Respir Crit Care Med*. 2016 Nov 15;194(10):1242-1251.
- 10) Tzouvelekis A, Herazo-Maya JD, Slade M, et al. Validation of the prognostic value of MMP-7 in idiopathic pulmonary fibrosis. *Respirology*. 2016 Oct 19. doi: 10.1111/resp.12920.
- 11) Richards TJ, Kaminski N, Baribaud F, et al. Peripheral blood proteins predict mortality in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2012 Jan 1;185(1):67-76. doi: 10.1164/rccm.201101-0058OC. Erratum in: *Am J Respir Crit Care Med*. 2012 Feb 15;185(4):464.
- 12) Greene KE, King TE Jr, Kuroki Y, et al. Serum surfactant proteins-A and -D as biomarkers in idiopathic pulmonary fibrosis. *Eur Respir J*. 2002 Mar;19(3):439-46.
- 13) Collard HR, Calfee CS, Wolters PJ, et al. Plasma biomarker profiles in acute exacerbation of idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol*. 2010 Jul;299(1):L3-7. doi: 10.1152/ajplung.90637.2008.
- 14) Kim JE, Lee KT, Lee JK, Paik SW, Rhee JC, Choi KW. Clinical usefulness of carbohydrate antigen 19-9 as a screening test for pancreatic cancer in an asymptomatic population. *J Gastroenterol Hepatol*. 2004;19(2):182-6.
- 15) Poruk KE, Gay DZ, Brown K, et al. The clinical utility of CA 19-9 in pancreatic adenocarcinoma: diagnostic and prognostic updates. *Curr Mol Med*. 2013 Mar;13(3):340-51.
- 16) Das S, Batra SK. Understanding the Unique Attributes of MUC16 (CA-125): Potential Implications in Targeted Therapy. *Cancer Res*. 2015 Nov 15;75(22):4669-74. doi: 10.1158/0008-5472.CAN-15-1050.

- 17) Seibold MA, Smith RW, Urbanek C, et al. The idiopathic pulmonary fibrosis honeycomb cyst contains a mucociliary pseudostratified epithelium. *PLoS One*. 2013;8(3):e58658. doi: 10.1371/journal.pone.0058658.
- 18) Evans CM, Fingerlin TE, Schwarz MI, et al. Idiopathic Pulmonary Fibrosis: A Genetic Disease That Involves Mucociliary Dysfunction of the Peripheral Airways. *Physiol Rev*. 2016 Oct;96(4):1567-91. doi: 10.1152/physrev.00004.2016.
- 19) Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet*. 2013 Jun;45(6):613-20. doi: 10.1038/ng.2609. Erratum in: *Nat Genet*. 2013 Nov;45(11):1409.
- 20) Chekmasova AA, Rao TD, Nikhamin Y, et al. Successful eradication of established peritoneal ovarian tumors in SCID-beige mice following adoptive transfer of T cells genetically targeted to the MUC16 antigen. *Clin Cancer Res* 2010;16:3594–606.
- 21) Xiang X, Feng M, Felder M, et al. HN125: A novel immunoadhesin targeting MUC16 with potential for cancer therapy. *J Cancer* 2011;2:280–91.
- 22) Argueso P, Guzman-Aranguez A, Mantelli F, Cao Z, Ricciuto J, Panjwani N. Association of cell surface mucins with galectin-3 contributes to the ocular surface epithelial barrier. *J Biol Chem* 2009;284:23037–45.
- 23) Mackinnon AC, Gibbons MA, Farnworth SL, et al. Regulation of transforming growth factor- $\beta$ 1-driven lung fibrosis by galectin-3. *Am J Respir Crit Care Med*. 2012 Mar 1;185(5):537-46. doi: 10.1164/rccm.201106-0965OC.
- 24) Le Jeune I, Gribbin J, West J, Smith C, Cullinan P, Hubbard R. The incidence of cancer in patients with idiopathic pulmonary fibrosis and sarcoidosis in the UK. *Respir Med*. 2007 Dec;101(12):2534-40.

- 25) Antoniou KM, Tomassetti S, Tsitoura E, Vancheri C. Idiopathic pulmonary fibrosis and lung cancer: a clinical and pathogenesis update. *Curr Opin Pulm Med*. 2015 Nov;21(6):626-33. doi: 10.1097/MCP.000000000000217. Review.
- 26) Richeldi L, du Bois RM, Raghu G, et al. INPULSIS Trial Investigators. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med*. 2014 May 29;370(22):2071-82. doi: 10.1056/NEJMoa1402584
- 27) Reck M, Kaiser R, Mellemegaard A, et al. LUME-Lung 1 Study Group. Docetaxel plus nintedanib versus docetaxel plus placebo in patients with previously treated non-small-cell lung cancer (LUME-Lung 1): a phase 3, double-blind, randomised controlled trial *Lancet Oncol*. 2014 Feb;15(2):143-55. doi: 10.1016/S1470-2045(13)70586-2.
- 28) McCormack PL. Nintedanib: first global approval. *Drugs*. 2015 Jan;75(1):129-39. doi: 10.1007/s40265-014-0335-0.
- 29) Vancheri C. Common pathways in idiopathic pulmonary fibrosis and cancer. *Eur Respir Rev*. 2013 Sep 1;22(129):265-72. doi: 10.1183/09059180.00003613
- 30) Noble PW, Albera C, Bradford WZ, ET AL ; CAPACITY Study Group. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. *Lancet*. 2011 May 21;377(9779):1760-9. doi: 10.1016/S0140-6736(11)60405-4.
- 31) King TE Jr, Bradford WZ, Castro-Bernardini S, et al. ASCEND Study Group. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med*. 2014 May 29;370(22):2083-92. doi: 10.1056/NEJMoa1402582.
- 32) Yang IV, Coldren CD, Leach SM, et al. Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax*. 2013 Dec;68(12):1114-21. doi: 10.1136/thoraxjnl-2012-202943.

- 33) Maher TM. Precision medicine in idiopathic pulmonary fibrosis. QJM. 2016 Sep;109(9):585-7.  
doi: 10.1093/qjmed/hcw117. Epub 2016 Aug 10.