



ORIGINAL ARTICLE

Mucin 5B promoter polymorphism is associated with idiopathic pulmonary fibrosis but not with development of lung fibrosis in systemic sclerosis or sarcoidosis

Carmel J Stock,¹ Hiroe Sato,¹ Carmen Fonseca,² Winston A S Banya,³ Philip L Molyneaux,¹ Huzaifa Adamali,¹ Anne-Marie Russell,¹ Christopher P Denton,² David J Abraham,² David M Hansell,⁴ Andrew G Nicholson,⁵ Toby M Maher,¹ Athol U Wells,¹ Gisela E Lindahl,¹ Elisabetta A Renzoni¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/thoraxjnl-2012-201786>).

¹Interstitial Lung Disease Unit, Royal Brompton Hospital and National Heart and Lung Institute, Imperial College London, London, UK

²Centre for Rheumatology and Connective Tissue Diseases, University College London Medical School, London, UK

³Clinical Trials and Evaluation Unit, Royal Brompton and Harefield NHS Foundation Trust, London, UK

⁴Department of Radiology, Royal Brompton Hospital and National Heart and Lung Institute, Imperial College London, London, UK

⁵Histopathology Department, Royal Brompton Hospital and National Heart and Lung Institute, Imperial College London, London, UK

Correspondence to

Carmel Stock, Interstitial Lung Disease Unit, Royal Brompton Hospital and National Heart and Lung Institute, Imperial College London, Emmanuel Kaye Building, 1B Manresa Road, London SW3 6LR, UK; c.stock@imperial.ac.uk

Received 17 February 2012

Revised 21 November 2012

Accepted 20 December 2012

Published Online First

15 January 2013



► <http://dx.doi.org/10.1136/thoraxjnl-2012-202957>

To cite: Stock CJ, Sato H, Fonseca C, et al. *Thorax* 2013;**68**:436–441.

ABSTRACT

Background A polymorphism (rs35705950) 3 kb upstream of *MUC5B*, the gene encoding Mucin 5 subtype B, has been shown to be associated with familial and sporadic idiopathic pulmonary fibrosis (IPF). We set out to verify whether this variant is also a risk factor for fibrotic lung disease in other settings and to confirm the published findings in a UK Caucasian IPF population.

Methods Caucasian UK healthy controls (n=416) and patients with IPF (n=110), sarcoidosis (n=180) and systemic sclerosis (SSc) (n=440) were genotyped to test for association. The SSc and sarcoidosis cohorts were subdivided according to the presence or absence of fibrotic lung disease. To assess correlation with disease progression, time to decline in forced vital capacity and/or lung carbon monoxide transfer factor was used in the IPF and SSc groups, while a persistent decline at 4 years since baseline was evaluated in patients with sarcoidosis.

Results A significant association of the *MUC5B* promoter single nucleotide polymorphism with IPF ($p=2.04 \times 10^{-17}$; OR 4.90, 95% CI 3.42 to 7.03) was confirmed in this UK population. The *MUC5B* variant was not a risk factor for lung fibrosis in patients with SSc or sarcoidosis and did not predict more rapidly progressive lung disease in any of the groups. Rather, a trend for a longer time to decline in forced vital capacity was observed in patients with IPF.

Conclusions We confirm the *MUC5B* variant association with IPF. We did not observe an association with lung fibrosis in the context of SSc or sarcoidosis, potentially highlighting fundamental differences in genetic susceptibility, although the limited subgroup numbers do not allow a definitive exclusion of an association.

INTRODUCTION

A single nucleotide polymorphism (SNP) (rs35705950) located 3 kb upstream of the transcriptional start site of *MUC5B*, the gene encoding Mucin 5 subtype B, has been shown to be strongly associated with both familial interstitial pneumonia¹ and sporadic idiopathic pulmonary fibrosis (IPF) in US populations.^{1 2} *MUC5B* is a gel-forming mucin and a major component of mucus

Key messages

What is the key question?

- Is the *MUC5B* rs35705950 polymorphism, recently reported as strongly associated with idiopathic pulmonary fibrosis, a risk factor for lung fibrosis in the context of systemic sclerosis and sarcoidosis?

What is the bottom line?

- In our Caucasian UK based cohorts, we confirm a significant association of the *MUC5B* variant with idiopathic pulmonary fibrosis. By contrast, we found no significant association with lung fibrosis in the context of systemic sclerosis or sarcoidosis, although fibrotic subgroup numbers may have resulted in insufficient power to detect an association, especially if less pronounced than in IPF.

Why read on?

- The lack of an association with lung fibrosis in the context of systemic sclerosis and sarcoidosis suggests that the *MUC5B* variant is associated with an idiopathic pulmonary fibrosis-specific pathway, differing from SSc- and sarcoidosis-associated lung fibrosis which are likely to have distinct genetic and pathogenetic risk factors.

in the respiratory tract.³ The IPF-associated SNP is located within a region highly conserved across vertebrate species and, based on a prediction algorithm, is potentially involved in gene regulation.⁴

By playing important roles in airway mucus rheology⁵ and mucosal immune defence,⁶ *MUC5B* is essential in protecting the surface epithelium of the airways. *MUC5B* overexpression was observed in IPF lungs, with patchy staining of the epithelial cells lining honeycomb cysts.¹ Although the mechanisms through which *MUC5B* dysregulation plays a part in the development of IPF are currently unknown, the strength of the genetic association

suggests a pathogenetic role of mucins and/or mucin-producing cells which was hitherto unsuspected. However, whether the *MUC5B* variant is a prognostic marker of IPF is currently unknown.

Fibrotic lung disease occurs in a large proportion of patients with systemic sclerosis (SSc) and, together with pulmonary hypertension, represents the main cause of death in these patients.⁷ In sarcoidosis the development of irreversible lung fibrosis, seen in 10–20% of patients, is associated with poorer quality of life, the need for long-term treatment and a worse prognosis.⁸ A small number of genes associated with the development of lung fibrosis in the context of SSc⁹ or of sarcoidosis^{10–12} have been reported, highlighting a genetic component likely to involve several genes with small individual effects. Although SSc-associated interstitial lung disease (SSc-ILD) and pulmonary sarcoidosis differ from IPF in many respects, including better survival for equivalent disease severity,¹³ there are many similarities in the fibrotic process and alveolar epithelium abnormalities occur in all three diseases. In particular, the membrane tethered mucin MUC1 (also known as KL-6), a member of the mucin family expressed by type II alveolar epithelial cells, has been suggested as a biomarker for disease activity in IPF,¹⁴ pulmonary sarcoidosis¹⁵ and SSc-ILD,^{16 17} although large prospective studies are yet to be performed. We therefore investigated whether the *MUC5B* polymorphism increased the risk of lung fibrosis in the context of SSc and sarcoidosis and set out to confirm the reported association with IPF. We also assessed whether this polymorphism was associated with the severity of lung fibrosis and the likelihood of disease progression in the three fibrotic lung diseases.

MATERIALS AND METHODS

Study populations

DNA samples were collected from consecutive newly presenting patients with IPF (n=110) and sarcoidosis (n=180) attending the respective clinics at the Royal Brompton Hospital, London, and from patients with SSc (n=440) attending clinics at the Royal Brompton and Royal Free Hospitals, London. The diagnoses were made from well-defined criteria for IPF,^{18 19} sarcoidosis²⁰ and SSc.²¹ The majority of patients with IPF received combination immunosuppressive treatment and low-dose prednisolone.²² None were included in trials with novel agents which could have influenced functional decline. Patients with SSc-ILD and sarcoidosis were treated with intended 'best management' with introduction/changes in immunosuppression dictated by progression of disease and side effects. The control population (n=416) comprised healthy blood donors mainly collected from the south-east of the UK. Individuals were judged as healthy based on a self-administered questionnaire and by the routine laboratory investigations performed on blood donors. Only Caucasian individuals of Northern European descent were included.

As both the Royal Brompton Hospital and the Royal Free Hospital Units are tertiary referral centres, patient cohorts are likely to be a mixture of incident and prevalent cases. To indirectly assess whether there could have been a selection bias favouring prevalent cases in the *MUC5B* variant carriers in the IPF cohort, we compared duration of dyspnoea at presentation between carriers and non-carriers. Details on duration of dyspnoea were retrievable from the clinical notes of 98/110 patients with IPF. Duration of dyspnoea at first presentation to the Royal Brompton Hospital was not significantly different between *MUC5B* T allele carriers and non-carriers (32.1 months in carriers vs 28.3 months in non-carriers, p=0.5

Mann–Whitney U test), suggesting there was not a significant imbalance between prevalent/incident cases according to *MUC5B* allele carriage.

Clinical assessment

Pulmonary function tests and high-resolution CT (HRCT) were performed as previously reported.²³ All investigations were performed as part of a prospective routine clinical protocol. Pulmonary function tests (expressed as percent predicted) from the time of first presentation at the Royal Brompton Hospital were available for patients with IPF (n=107), SSc-ILD (n=206) and sarcoidosis (n=178). Among patients with SSc, ILD was defined as a forced vital capacity (FVC) <75% and/or the presence of fibrosis on chest imaging (chest x-ray or HRCT).

Patients with SSc were also classified into autoantibody subgroups by previously described standard methods.^{24 25} The patients with SSc for whom antibody status was known (n=417) were grouped according to status for the two most common antibodies found in the cohort, anti-DNA topoisomerase 1 antibodies (ATA), which is known to be tightly linked to the presence of ILD, and anti-centromere antibodies (ACA), which is associated with limited skin disease and pulmonary hypertension but protective for ILD.²⁶

Quantification of disease severity and progression in IPF and SSc-ILD

The composite physiological index (CPI), a functional index of disease severity, was calculated as: $CPI = 91.0 - (0.65 \times TLCO\% \text{ predicted}) - (0.53 \times FVC\% \text{ predicted}) + (0.34 \times FEV_1\% \text{ predicted})$, where TLCO is lung carbon monoxide transfer factor, and was used as a marker of lung disease severity in patients with IPF and SSc-ILD.²⁷

Time to decline was quantified using serial pulmonary functional indices starting from first presentation to the Royal Brompton Hospital. Significant functional deterioration was defined as a decline (quantified as percentage change from baseline) of $\geq 10\%$ in FVC and/or $\geq 15\%$ in TLCO, according to established criteria.^{18 19} In IPF, a relentlessly progressive disease, one subsequent time point from baseline was the minimum requirement for follow-up, and time to first observed decline was used. In SSc-ILD, to allow for possible response to treatment or spontaneous fluctuations, at least two follow-up time points from baseline were required. For SSc-ILD, time to irreversible decline was used, defined as time to the first significant change in FVC and/or TLCO (as above) observed on at least two consecutive occasions, as previously described.²⁸ When follow-up ended with functional decline on a single occasion, this was accepted as a significant decline provided there was symptomatic or radiological evidence of worsening. Data at a sufficient number of time points were available to calculate time to decline in 95 patients with IPF and 177 with SSc-ILD.

Quantification of disease severity and progression in sarcoidosis

Based on the chest x-ray, staging of pulmonary disease was classified into stage 0 (normal chest x-ray), stage I (bilateral hilar lymphadenopathy (BHL)), stage II (BHL with pulmonary infiltrates), stage III (pulmonary infiltrates without BHL) and stage IV (pulmonary fibrosis), according to established criteria.²⁰ Patients with sarcoidosis for whom lung function tests at presentation and 4 years were available (n=140) were also categorised into patients who showed a significant decline in lung function at the 4-year time point and those who did not. In order for the decline to be deemed clinically significant, it had to be

confirmed at the following clinic visit. A significant decline in lung function was defined as a reduction in one or more of: TLCO by $\geq 15\%$, forced expiratory volume in 1 s (FEV₁) by $\geq 10\%$ or FVC by $\geq 10\%$.

Genotyping

Genotyping was carried out according to the manufacturer's instructions using a commercially available TaqMan assay and TaqMan universal PCR master mix, no AmpErase UNG (Applied Biosystems, Carlsbad, California, USA) on a Rotor-Gene 6000 real-time PCR machine (Qiagen Inc, Valencia, California, USA). Quality control and genotype determination were performed using Rotor Gene 6000 Series Software 1.7 (Corbett Research, Sydney, Australia).

Statistical analysis

To test for deviation from Hardy–Weinberg equilibrium (HWE), genotype frequencies were determined by direct counting and the χ^2 statistic or Fisher exact test used as appropriate. χ^2 analyses for association were carried out in Unphased V3.1²⁹ and the non-parametric Mann–Whitney test was used in GraphPad Prism V4 to test for differences in median lung function measures according to carriage of the minor T allele. p Values < 0.05 were considered significant. The current study had 96% power to detect an association with an OR of at least 3 in the IPF cohort, well below the published OR for the association between the *MUC5B* SNP and IPF in two separate populations.^{1 2} The study had 80% power to detect an association with an OR of at least 2.0 with SSc-ILD and of at least 2.6 with sarcoidosis with lung fibrosis. The prognostic value of *MUC5B* genotype and allele carriage was assessed in relation to disease progression. Based on the number of patients, rate of decline in FVC and TLCO and length of follow-up, we estimated that this study had a power of 80% to detect a HR of at least 2 (or 0.5) according to *MUC5B* allele carriage in both the IPF and SSc-ILD cohorts. Proportional hazards analysis was used to evaluate time to decline in FVC and/or time to decline in TLCO, using the non-zero slope test developed by Therneau and Grambsch as implemented in Stata (Computing Resource Centre, Santa Monica, California, USA). In both IPF and SSc-ILD univariate and multivariate Cox analysis, the slope was not significantly different from zero, thereby indicating non-violation of the proportional hazard assumption. Stepwise multivariate analysis was used to adjust for potential confounding factors including age, gender, smoking status and disease severity (CPI at first presentation), with retention of significant variables in the final regression equations.

RESULTS

Demographic and clinical data of the IPF, SSc and sarcoidosis cohorts are presented in table 1. Genotyping was successful for all samples except for one control sample. This ungenotyped sample is not included in the reported cohort size (n=416). The genotype distribution in all of the analysed cohorts, including all patient subgroups, conformed to HWE. The minor *MUC5B* T allele previously associated with IPF^{1 2} was present at a significantly higher frequency in the IPF patient cohort than in the healthy controls (36% vs 10%, p=2.04 $\times 10^{-17}$, table 2). The genotype p value was 9.11 $\times 10^{-17}$ (see online supplementary table S1), and the OR for IPF in individuals heterozygous or homozygous for the T allele were 6.62 (95% CI 4.10 to 10.67) and 11.81 (95% CI 4.26 to 33.72), respectively.

By contrast, there was no significant difference in the *MUC5B* allele or genotype frequencies between the healthy controls and patients with SSc as a whole, or according to presence or

Table 1 Patient demographics and clinical characteristics at presentation

	IPF (n=110)	SSc (n=440)	Sarcoidosis (n=180)
Mean (range) age, years	64.6 (45–85)	52.8 (15–83)	39.6 (15–75)
Gender (F/M)	31/79	356/84	93/87
Median (range) FVC, % predicted	69.3 (41–98.9)	91 (40–158)	96.9 (42.4–146.5)
Median (range) FEV ₁ , % predicted	74.2 (40–98.7)	87 (29.2–136)	91.9 (14.1–136)
Median (range) TLCO, % predicted	40 (18.1–76.8)	63 (20–124)	75.1 (25.3–125.4)
Median (range) CPI	52.8 (0.27–73.2)	30.7 (–9.9–68.3)	20.3 (–14.7–63.2)

CPI, composite physiological index; FEV₁, forced expiratory volume in 1 s; FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis; TLCO, lung carbon monoxide transfer factor.

absence of ILD (table 2 and online supplementary table S1). Similarly, no difference was found between the control cohort and patients with sarcoidosis as a whole or according to the presence of non-fibrotic (stages 0–III) or fibrotic (stage IV) pulmonary involvement (table 2 and online supplementary table S1).

Association with disease severity and rate of decline

Idiopathic pulmonary fibrosis

Median FVC% predicted, TLCO% predicted and CPI did not differ according to carriage or non-carriage of the minor T allele (table 3). With a median follow-up time of 13.7 months, 61 of 95 patients (64.2%) had a decline in FVC (median time to decline 10.7 months) and 60 of 95 patients (63.3%) had a decline in TLCO (median time to decline 10.8 months). No significant difference in time to decline in FVC (HR 0.78, 95% CI 0.47 to 1.30, p=0.3) or TLCO (HR 1.03, 95% CI 0.61 to 1.73, p=0.9) was seen according to carriage of the disease-associated T allele on univariate analysis (see online supplementary figures S1A,B). However, after multivariate stepwise regression leaving only significant covariates in the equation (CPI), the minor T allele was associated with a longer time to decline in FVC, just reaching the limits of statistical significance (HR 0.59, 95% CI

Table 2 Allele frequency in control and ILD patient cohorts

	G allele	T allele	OR* (95% CI)	p Value
Control (n=416)	0.90	0.10		
IPF (n=110)	0.64	0.36	4.90 (3.42 to 7.03)	2.04 $\times 10^{-17}$
SSc (n=440)	0.88	0.12	1.16 (0.85 to 1.57)	0.36
No ILD (n=211)	0.89	0.11	1.06 (0.73 to 1.56)	0.76
SSc-ILD (n=229)	0.88	0.12	1.24 (0.86 to 1.79)	0.24
ATA +ve (n=110)	0.88	0.12	1.25 (0.79 to 1.98)	0.35
ACA +ve (n=109)	0.88	0.12	1.26 (0.79 to 1.99)	0.33
Non-ATA/non-ACA (n=208)	0.90	0.10	1.03 (0.69 to 1.51)	0.89
Sarcoidosis (n=180)	0.89	0.11	1.11 (0.75 to 1.65)	0.60
Stage 0–III (n=98)	0.87	0.13	1.30 (0.82 to 2.08)	0.29
Stage IV (n=73)	0.91	0.09	0.90 (0.49 to 1.64)	0.71

Data are presented as allele frequency; p value is compared with the control cohort.

*Of the T allele.

ACA, anti-centromere antibody; ATA, anti-DNA topoisomerase 1 antibody; ILD, interstitial lung disease; IPF, idiopathic pulmonary fibrosis; SSc, systemic sclerosis.

Table 3 Lung function impairment in IPF and SSc-ILD according to carriage of the T allele

	Non-carrier	Carrier	p Value
IPF	n=41	n=66	
CPI	54.0 (0.27–65.9)	50.0 (18–73.2)	0.43
FVC, % predicted	65.9 (41–132)	72.0 (28.6–143)	0.78
Tl _{CO} , % predicted	38.9 (23.4–101)	40.0 (18.1–76.8)	0.92
SSc-ILD	n=156	n=50	
CPI	39.1 (–0.65–68.1)	41.2 (3.9–68.3)	0.39
FVC, % predicted	80.3 (40–145)	82.3 (52.6–119)	0.99
Tl _{CO} , % predicted	55.4 (20–107.7)	54.6 (22.3–91.3)	0.30

Data are presented as median at presentation (range).

CPI, composite physiological index; FVC, forced vital capacity; IPF, idiopathic pulmonary fibrosis; SSc-ILD, systemic sclerosis associated interstitial lung disease; Tl_{CO}, lung carbon monoxide transfer factor.

0.35 to 1.005, $p=0.052$), while no association was seen with time to decline in Tl_{CO} (HR 0.96, 95% CI 0.56 to 1.62, $p=0.9$), even after adjustment for CPI.

Systemic sclerosis-associated interstitial lung disease

Median FVC% predicted, Tl_{CO}% predicted and CPI did not differ according to carriage or non-carriage of the minor T allele (table 3). With a median follow-up time of 98.4 months, 104 of 177 patients (58.8%) had a decline in FVC (median time to decline 35.1 months) and 124 of 177 patients (70.1%) had a decline in Tl_{CO} (median time to decline 42 months). Time to decline in FVC, Tl_{CO} or either measure did not correlate with carriage of the T allele, both on univariate and multivariate stepwise regression analysis, after adjustment for age, gender, smoking status and CPI (adjusted HR for time to decline in FVC 1.16, 95% CI 0.7 to 1.9, $p=0.5$; adjusted HR for time to decline in Tl_{CO} 1.19, 95% CI 0.7 to 1.9, $p=0.5$) (see online supplementary figures S2A and S2B). With regard to autoantibody status, no significant difference was observed in allele frequencies between controls and patients with SSc according to the presence of ATA or ACA antibodies (table 2).

Sarcoidosis

The *MUC5B* SNP was not significantly associated with severity of lung function impairment at first presentation or at 2 and 4 years of follow-up (table 4). Furthermore, no significant association was observed with lung disease progression in sarcoidosis, with no significant difference seen between the frequency in controls and patients who did ($n=27$, MAF=0.06, $p=0.67$) or did not ($n=113$, MAF=0.11, $p=0.28$) show a substantial decline in lung function at 4 years of follow-up.

DISCUSSION

We found no association between the *MUC5B* promoter region polymorphism rs35705950 and the presence of lung fibrosis in the context of scleroderma or sarcoidosis while we confirm, in a UK Caucasian population, the association with sporadic IPF. The lack of an association with lung fibrosis in the context of SSc and sarcoidosis suggests that the *MUC5B* variant is not related to shared fibrotic mechanisms across fibrotic lung diseases but is instead associated with an IPF-specific pathway, differing from SSc- and sarcoidosis-associated lung fibrosis which are more likely to be related to immunological/inflammatory triggers.

Table 4 Pulmonary function in sarcoidosis according to carriage of the minor T allele

	Non-carrier	Carrier	p Value
At presentation	n=140	n=38	
Tl _{CO}	75.3 (25.3–125.4)	74.6 (35–109.7)	0.60
FEV ₁	92.8 (14.1–136)	87.5 (22.8–128.5)	0.26
FVC	97.9 (42.4–143.8)	94.6 (43.2–146.5)	0.74
At 2 years	n=121	n=28	
Tl _{CO}	76.6 (24.1–126.5)	72.0 (38.4–110.4)	0.37
FEV ₁	89.1 (13.9–134.5)	93.5 (21.6–128.3)	0.53
FVC	98.5 (47.8–135.9)	100.9 (54.4–143.2)	0.92
At 4 years	n=114	n=26	
Tl _{CO}	76.1 (11.7–126.7)	75.6 (46–119)	0.95
FEV ₁	89.7 (32.8–142.2)	90.2 (34.2–135.9)	0.64
FVC	98.1 (35.1–143)	100.3 (64.4–150.5)	0.47

Data are presented as median (range) percentage predicted.

FEV₁, forced expiratory volume in 1 s; FVC, forced vital capacity; Tl_{CO}, lung carbon monoxide transfer factor.

In IPF, experimental evidence suggests that the respiratory epithelial abnormalities have an important pathogenetic role.^{30–33} Bronchiolisation of the distal lung epithelium with aberrant expression of *MUC5B* has been observed in the lungs of patients with IPF and may be related to abnormal differentiation of the respiratory epithelium.³³ Overexpression of *MUC5B* is preferentially observed in honeycombed lesions;¹ interestingly, honeycombing is present, if at all, only to a minimal degree in the context of idiopathic or SSc-associated fibrotic non-specific interstitial pneumonia (NSIP). One might therefore speculate that aberrant expression of *MUC5B* is involved in the development of architectural distortion not seen to a similar degree in SSc-ILD, mostly characterised by an NSIP pattern.³⁴ If this were the case, one could expect the *MUC5B* polymorphism not to be associated with idiopathic fibrotic NSIP, a hypothesis which will require a multicentre collaborative approach in view of its relative rarity.

Limitations of this study include the relatively small numbers in each disease group and the lack of a replication cohort for the positive association observed in IPF. However, our IPF study is in effect a replication cohort, as the association has previously been described by two separate groups.^{1, 2} Regarding SSc-ILD and sarcoidosis, it is possible that we were unable to detect a weaker association than the one seen in IPF due to insufficient power. However, the lack of association with SSc-ILD is in accordance with the recent publication by Peljto *et al*³⁵ where the cohort was approximately half the size of that used in this study. Nevertheless, in view of the limitations in terms of power, further studies using larger patient cohorts of non-IPF patients are required to confirm the findings of this study.

Subgroup analyses of putative associations with disease progression within each disease cohort were hampered by relatively small numbers, limiting the power to detect an association. Nevertheless, following adjustment for disease severity, we observed a trend bordering on statistical significance between the *MUC5B* variant and slower decline in FVC in patients with IPF. Interestingly, an association between the *MUC5B* variant and less severe pathological changes was recently reported in familial interstitial pneumonia.³⁶ Although this observation will require further evaluation in separate populations, it provides a possible link between this polymorphism and disease behaviour in IPF, suggesting the need for further exploration. In both the

SSc-ILD and in sarcoidosis groups the analysis was potentially underpowered to pick up a more subtle effect on rate of decline. Furthermore, it is not possible to exclude the fact that the *MUC5B* variant could have an effect on the natural untreated history of the disease as, in both groups, patients observed to decline would have had immunosuppressive treatment initiated or altered. As the choice of treatment, timing and duration varied significantly during the years of follow-up, it is not possible to correct for treatment effects in the analysis. However, we can at least conclude that the *MUC5B* variant is not associated with marked disease progression in a group of unselected SSc-ILD and sarcoidosis patients receiving intended 'best management'.

Seibold *et al* showed that, in the lungs of healthy individuals, high mRNA expression levels of *MUC5B* were significantly associated with carriage of the IPF-associated T allele. Interestingly, although IPF lungs expressed 14.1 times higher levels of *MUC5B* RNA than control lungs, there was no relationship seen between *MUC5B* genotype and expression levels, in contrast to control lungs.¹ It is possible that, in IPF patients homozygous for the G allele, *MUC5B* expression is upregulated as a result of aberrant expression of an upstream/downstream factor. The role played by *MUC5B* overexpression in the pathogenesis of IPF is currently unknown, although a number of possible mechanisms have been proposed.^{1 37} In order to better understand the pathobiology of IPF, further work to investigate the effect of excess *MUC5B* on injury response in the lung is needed. Evaluation of *MUC5B* expression in lung tissue of patients with SSc-ILD, fibrotic sarcoidosis and idiopathic NSIP would also be informative. Overexpression could suggest involvement of genetic variants upstream or downstream of *MUC5B*, or at least the importance of this pathway across fibrotic lung diseases, while the absence of *MUC5B* overexpression would highlight *MUC5B* and related pathways as IPF-specific.

Despite the evidence that *MUC5B* expression correlates with rs35705950,¹ further work is required to establish a direct relationship between this SNP and regulation of expression. It could be acting as a marker for another polymorphism with which it is in linkage disequilibrium. Although Seibold *et al* demonstrated that rs35705950 shows an independent effect on susceptibility to IPF, it is possible that unscreened genetic variants, particularly in unscreened repetitive regions which the authors were unable to assess, may be in strong linkage disequilibrium with rs35705950 and be the true disease-associated variant. This is particularly true in the SSc and sarcoidosis cohorts, in which *MUC5B* polymorphisms have not been previously screened. A broader gene-focused rather than a single SNP-focused study would be necessary to determine if other *MUC5B* variants are involved in the pathogenesis of SSc-ILD or sarcoidosis.

In conclusion, we confirm that the *MUC5B* polymorphism rs35705950 is significantly associated with susceptibility to IPF in this UK Caucasian population. However, in this population, the *MUC5B* variant is not a risk factor for development of lung fibrosis in SSc or sarcoidosis. We found that it is not a marker of more rapidly progressive disease in any of the three groups and, rather, there was a trend for a slower decline in FVC observed in the patients with IPF which will require further investigation.

Acknowledgements We are grateful to all of the patients for allowing us to collect blood samples for genetic association studies over the years.

Contributors CJS and EAR contributed to the study design, data acquisition, analysis and interpretation. CF, GEL, DJA contributed to analysis and interpretation. HS, PLM, HA, AMR, CPD and TMM contributed to data

acquisition and critical revision of the manuscript. DMH and AGN contributed to critical revision of the manuscript. AUW contributed to the analysis and interpretation of the data and to critical revision. All authors approved the final version of the manuscript. EAR is guarantor of the manuscript. WASG contributed to statistical analysis.

Ethics approval The ethics committees of the Royal Brompton Hospital and of the Royal Free Hospital gave authorisation for the study.

Patient consent All participants gave written informed consent.

Funding This work was funded by the Raynaud's and Scleroderma Association (BR10), the Asmarley Trust (B0498), Arthritis Research UK (19291 and 19427) and the Scleroderma Society. AGN was partly supported by the NIHR Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London. The project was supported by the NIHR Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London.

Competing interests None.

Provenance and review Not commissioned; externally peer reviewed.

REFERENCES

- Seibold MA, Wise AL, Speer MC, *et al*. A common *MUC5B* promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364:1503–12.
- Zhang Y, Noth I, Garcia JG, *et al*. A variant in the promoter of *MUC5B* and idiopathic pulmonary fibrosis. *N Engl J Med* 2011;364:1576–7.
- Thornton DJ, Devine PL, Hanski C, *et al*. Identification of two major populations of mucins in respiratory secretions. *Am J Respir Crit Care Med* 1994;150:823–32.
- Taylor J, Tyekuceva S, King DC, *et al*. ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. *Genome Res* 2006;16:1596–604.
- Wickstrom C, Davies JR, Eriksen GV, *et al*. *MUC5B* is a major gel-forming, oligomeric mucin from human salivary gland, respiratory tract and endocervix: identification of glycoforms and C-terminal cleavage. *Biochem J* 1998;334 (Pt 3):685–93.
- Kim KC. Role of epithelial mucins during airway infection. *Pulm Pharmacol Ther* 2012;25:415–9.
- Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972–2002. *Ann Rheum Dis* 2007;66:940–4.
- Nardi A, Brillet PY, Letoumelin P, *et al*. Stage IV sarcoidosis: comparison of survival with the general population and causes of death. *Eur Respir J* 2011;38:1368–73.
- Martin JE, Bossini-Castillo L, Martin J. Unraveling the genetic component of systemic sclerosis. *Hum Genet* 2012;131:1023–37.
- Hill MR, Papafili A, Booth H, *et al*. Functional prostaglandin-endoperoxide synthase 2 polymorphism predicts poor outcome in sarcoidosis. *Am J Respir Crit Care Med* 2006;174:915–22.
- Sato H, Williams HR, Spagnolo P, *et al*. CARD15/NOD2 polymorphisms are associated with severe pulmonary sarcoidosis. *Eur Respir J* 2010;35:324–30.
- Heron M, van Moorsel CH, Grutters JC, *et al*. Genetic variation in *GREM1* is a risk factor for fibrosis in pulmonary sarcoidosis. *Tissue Antigens* 2011;77:112–17.
- Wells AU, Cullinan P, Hansell DM, *et al*. Fibrosing alveolitis associated with systemic sclerosis has a better prognosis than lone cryptogenic fibrosing alveolitis. *Am J Respir Crit Care Med* 1994;149:1583–90.
- Kobayashi J, Kitamura S. KL-6: a serum marker for interstitial pneumonia. *Chest* 1995;108:311–15.
- Janssen R, Sato H, Grutters JC, *et al*. Study of Clara cell 16, KL-6, and surfactant protein-D in serum as disease markers in pulmonary sarcoidosis. *Chest* 2003;124:2119–25.
- Yanaba K, Hasegawa M, Hamaguchi Y, *et al*. Longitudinal analysis of serum KL-6 levels in patients with systemic sclerosis: association with the activity of pulmonary fibrosis. *Clin Exp Rheumatol* 2003;21:429–36.
- Yamane K, Ihn H, Kubo M, *et al*. Serum levels of KL-6 as a useful marker for evaluating pulmonary fibrosis in patients with systemic sclerosis. *J Rheumatol* 2000;27:930–4.
- American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am J Respir Crit Care Med* 2000;161:646–64.
- American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med* 2002;165:277–304.
- American Thoracic Society/European Respiratory Society/World Association of Sarcoidosis and Other Granulomatous Disorders. Statement on sarcoidosis. Joint Statement of the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the World Association of Sarcoidosis and Other Granulomatous Disorders

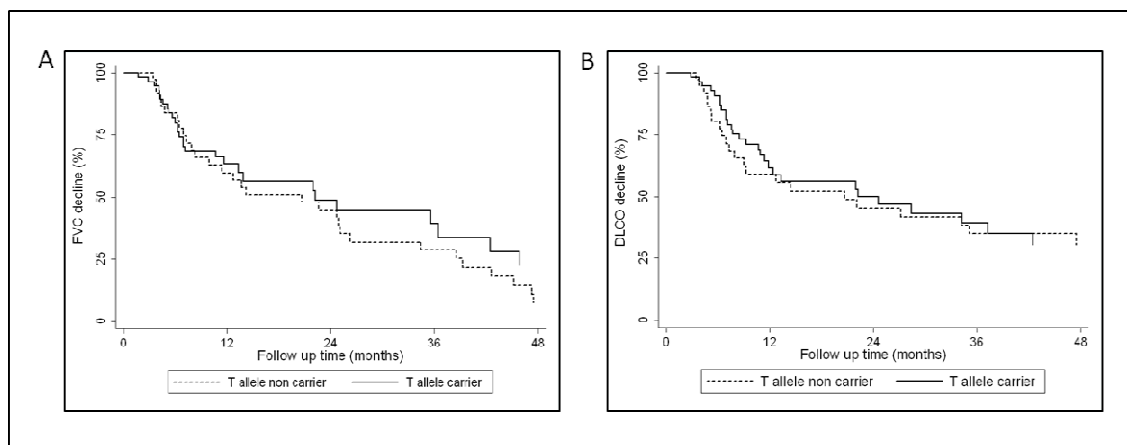
- (WASOG) adopted by the ATS Board of Directors and by the ERS Executive Committee, February 1999. *Am J Respir Crit Care Med* 1999;160:736–55.
- 21 Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum* 1980;23:581–90.
 - 22 Bradley B, Branley HM, Egan JJ, *et al*. Interstitial lung disease guideline: the British Thoracic Society in collaboration with the Thoracic Society of Australia and New Zealand and the Irish Thoracic Society. *Thorax* 2008;63(Suppl 5): v1–58.
 - 23 Goh NS, Desai SR, Veeraraghavan S, *et al*. Interstitial lung disease in systemic sclerosis: a simple staging system. *Am J Respir Crit Care Med* 2008;177:1248–54.
 - 24 Bunn CC, Denton CP, Shi-Wen X, *et al*. Anti-RNA polymerases and other autoantibody specificities in systemic sclerosis. *Br J Rheumatol* 1998;37:15–20.
 - 25 Reveille JD, Solomon DH. Evidence-based guidelines for the use of immunologic tests: anticentromere, Scl-70, and nucleolar antibodies. *Arthritis Rheum* 2003;49:399–412.
 - 26 Renzoni EA. Interstitial lung disease in systemic sclerosis. *Monaldi Arch Chest Dis* 2007;67:217–28.
 - 27 Wells AU, Desai SR, Rubens MB, *et al*. Idiopathic pulmonary fibrosis: a composite physiologic index derived from disease extent observed by computed tomography. *Am J Respir Crit Care Med* 2003;167:962–9.
 - 28 Goh NS, Veeraraghavan S, Desai SR, *et al*. Bronchoalveolar lavage cellular profiles in patients with systemic sclerosis-associated interstitial lung disease are not predictive of disease progression. *Arthritis Rheum* 2007;56:2005–12.
 - 29 Dudbridge F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered* 2008;66:87–98.
 - 30 Barbas-Filho JV, Ferreira MA, Sesso A, *et al*. Evidence of type II pneumocyte apoptosis in the pathogenesis of idiopathic pulmonary fibrosis (IFP)/usual interstitial pneumonia (UIP). *J Clin Pathol* 2001;54:132–8.
 - 31 Chilosi M, Poletti V, Murer B, *et al*. Abnormal re-epithelialization and lung remodeling in idiopathic pulmonary fibrosis: the role of deltaN-p63. *Lab Invest* 2002;82:1335–45.
 - 32 Willis BC, Liebler JM, Luby-Phelps K, *et al*. Induction of epithelial-mesenchymal transition in alveolar epithelial cells by transforming growth factor-beta1: potential role in idiopathic pulmonary fibrosis. *Am J Pathol* 2005;166:1321–32.
 - 33 Plantier L, Crestani B, Wert SE, *et al*. Ectopic respiratory epithelial cell differentiation in bronchiolised distal airspaces in idiopathic pulmonary fibrosis. *Thorax* 2011;66:651–7.
 - 34 Bours D, Wells AU, Nicholson AG, *et al*. Histopathologic subsets of fibrosing alveolitis in patients with systemic sclerosis and their relationship to outcome. *Am J Respir Crit Care Med* 2002;165:1581–6.
 - 35 Peljto AL, Steele MP, Fingerlin TE, *et al*. The pulmonary fibrosis-associated MUC5B promoter polymorphism does not influence the development of interstitial pneumonia in systemic sclerosis. *Chest* 2012;142:1584–8.
 - 36 Cosgrove GP, Groshong SD, Peljto AL, *et al*. The MUC5B promoter polymorphism is associated with a less severe pathological form of familial interstitial pneumonia (FIP) (abstract). *Am J Respir Crit Care Med* 2012;185:A6865.
 - 37 Boucher RC. Idiopathic pulmonary fibrosis—a sticky business. *N Engl J Med* 2011;364:1560–1.

Supplementary Table 1. Genotype frequency in control and ILD patient cohorts

	GG	GT	TT	p value
Control (n=416)	0.81	0.17	0.02	
IPF (n=110)	0.38	0.53	0.09	9.11x10 ⁻¹⁷
SSc (n=440)	0.78	0.22	0.01	0.09
no ILD (n=211)	0.80	0.19	0.01	0.57
SSc-ILD (n=229)	0.76	0.24	0.01	0.06
Sarcoidosis (n=180)	0.79	0.20	0.01	0.58
Stage 0-III (n=98)	0.77	0.21	0.02	0.54
Stage IV (n=73)	0.82	0.18	0.00	0.53

Data are presented as genotype frequency. p value is versus the control cohort. SSc: systemic sclerosis, ILD: interstitial lung disease

Supplementary Figure 1. Time to decline in FVC (panel A) and in DLco (panel B) in IPF patients, according to carriage of the MUC5B T allele. T allele non-carrier n=38, T allele carrier n=57. No significant difference in time to decline in FVC (HR: 0.78, 95% CI: 0.47-1.30, p=0.3), DLco (HR: 1.03, 95% CI: 0.61-1.73, p=0.9) was seen according to carriage of the disease-associated T allele on univariate analysis. However, after multivariate stepwise regression, leaving only significant covariates in the equation (CPI), the minor T allele was associated with a longer time to decline in FVC, just reaching the limits of statistical significance (HR: 0.59, 95% CI: 0.35-1.005, p=0.052), while no association was seen with time to decline in DLco (HR: 0.96, 95% CI: 0.56-1.62, p=0.9) even after adjustment for CPI.



Supplementary Figure 2. Time to decline in FVC (panel A) and in DLco (panel B) in SSc-ILD patients, according to carriage of the MUC5B T allele. T allele non-carrier n=140, T allele carrier n=37. Time to decline in FVC, DLco, or either measure, did not correlate with carriage of the T allele, both on univariate (adjusted HR for time to decline in FVC: 1.16, 95% CI: 0.73-1.84, p=0.5, adjusted HR for time to decline in DLco: 1.46, 95% CI: 0.95-2.23, p=0.08), and multivariate stepwise regression analysis, after adjustment for age, gender, smoking status, and CPI (adjusted HR for time to decline in FVC: 1.16, 95% CI: 0.7-1.9, p=0.5, adjusted HR for time to decline in DLco: 1.19, 95% CI: 0.7-1.9, p=0.5).

