Title

Analysis of protein-altering variants in telomerase genes and their association with MUC5B common variant status in patients with idiopathic pulmonary fibrosis: a candidate gene sequencing study

Journal

The Lancet Respiratory Medicine, 6(8)

ISSN

2213-2600

Authors

Dressen, Amy
Abbas, Alexander R
Cabanski, Christopher
et al.

Publication Date

2018-08-01

DOI

10.1016/s2213-2600(18)30135-8

Peer reviewed

HHS Public Access

Author manuscript

Lancet Respir Med. Author manuscript; available in PMC 2019 August 01.

Published in final edited form as:

Lancet Respir Med. 2018 August ; 6(8): 603–614. doi:10.1016/S2213-2600(18)30135-8.

Analysis of protein-altering variants in telomerase genes and their association with *MUC5B* common variant status in patients with idiopathic pulmonary fibrosis: a candidate gene sequencing study.

Amy Dressen, MS[#1], **Alexander R. Abbas, PhD**[#1], **Christopher Cabanski, PhD**[1], **Janina Reeder, PhD**[1], **Thirumalai R. Ramalingam, PhD**[1], **Margaret Neighbors, PhD**[1], **Tushar R. Bhangale, PhD**[1], **Matthew J. Brauer, PhD**[1], **Julie Hunkapiller, PhD**[1], **Jens Reeder, PhD**[1], **Kiran Mukhyala, MS**[1], **Karen Cuenco, PhD**[1], **Jennifer Tom, PhD**[1], **Amy Cowgill, MA**[1], **Jan Vogel, Msc**[1], **William F. Forrest, PhD**[1], **Harold R. Collard, MD**[2] **[Professor]**, **Paul J. Wolters, MD**[2] **[Professor]**, **Jonathan A. Kropski, MD**[3,4], **Lisa H. Lancaster, MD**[3], **Timothy S. Blackwell, MD**[3,4,5] **[Professor]**, **Joseph R. Arron, MD**[#*,1], and **Brian L. Yaspan, PhD**[#*,1]

[1]Genentech. Inc., South San Francisco, CA, USA

[2]Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of California, San Francisco, CA, USA

[3]Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

[4]Department of Veterans Affairs Medical Center, Nashville, TN, USA

[5]Department of Medicine, Cell and Developmental Biology, and Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN, USA

[#] These authors contributed equally to this work.

Abstract

[*]to whom correspondence should be addressed.

Author Contributions

AD, ARA, JRA and BLY conceived and designed the study. AD and ARA did the statistical analysis. CC, JaR, TRR, and MN contributed to the experimental design and analysis. TRB, MJB, JH, JeR, KM, KC, JT, AC, JV, WFF and AD contributed to whole genome sequence generation, SNP calling and analysis pipeline generation. HRC, PJW, JAK, LHL and TSB were site principal investigators responsible for participant recruitment and data collection. JRA and BLY wrote the first draft of the Article, which was reviewed by all authors.

Conflict of Interest Statements

AD, ARA, JRA, BLY, CC, JaR, TRR, MN, TRB, MJB, JH, JeR, KM, KC, JT, AC, JV, WFF are employees of Genentech who hold stock and stock options in Roche. LHL is or has consulted for Genentech, Boehringer Ingelheim, Global Blood Therapeutics, and Bellerophon and has done disease state education for Genentech and Boehringer Ingelheim.

Ethics Committee approval

All patients included in this study signed informed consents allowing for whole genome sequencing of their DNA.

BACKGROUND: Idiopathic Pulmonary Fibrosis (IPF) risk has a strong genetic component. Studies have implicated variation at several loci, including telomerase reverse transcriptase (*TERT)*, surfactant genes, and a single nucleotide polymorphism (SNP) at chr11p15, rs35705950 in the intergenic region between *TOLLIP* and *MUC5B*. IPF patients with risk alleles at rs35705950 have longer survival from the time of IPF diagnosis than patients homozygous for the non-risk allele, while patients with shorter telomeres have shorter survival times. We hypothesized that rare protein altering variants in genes regulating telomere length are enriched in IPF patients lacking risk alleles at rs35705950.

METHODS: Whole genome sequencing of 1,510 patients with sporadic IPF from phase 3 clinical trials and observational studies was used to assess telomere length and identify rare protein altering variants. We separated patients by rs35705950 genotype and assessed rare functional variation in *TERT* exons and compared genotypes to telomere length and rates of disease progression.

FINDINGS: 2·9% of patients with an rs35705950 risk allele carried a rare protein-altering variant (RV) in *TERT* compared to 7·3% of non-risk allele carriers (odds ratio [OR] 0·40 [95% CI 0·24-0·66], $p=3{\cdot}9 \times 10^{-4}$). Subsequent analyses identified enrichment of rare protein-altering variants in *PARN*, *RTEL1* and rare variation in *TERC* in IPF patients compared to non-IPF controls. In total, IPF patients harbored at least one rare variant in *TERT, PARN, TERC* or *RTEL1* more frequently than non-IPF patients from other clinical trials (8·57% in IPF vs. 2·37% for others $p=2{\cdot}44 \times 10^{-8}$). Patients with a variant in any of the four identified telomerase component genes had 4·78%-16·10% shorter telomeres and an earlier age of onset (65·1 years) than patients without (67·1 years; p=0·004). Patients with shorter telomeres had more rapid rates of lung function decline in the placebo arms of clinical trials (1·7% FVC/kb/year, p=0·002). Despite the aforementioned differences, we found pirfenidone demonstrated treatment benefit regardless of telomere length status.

INTERPRETATION: Rare protein-altering variants in *TERT, PARN, TERC* and *RTEL1* are enriched in IPF patients compared to non-IPF controls, and, in the case of *TERT*, particularly in those not carrying a risk allele at the rs35705950 locus, suggesting that there are multiple genetic factors underlying sporadic IPF that may implicate distinct mechanisms of pathogenesis and rates of progression.

TRIAL REGISTRATION: NCT00287716, NCT01366209, NCT00075998, NCT01872689.

Introduction

Idiopathic pulmonary fibrosis (IPF) is a progressive and lethal fibrotic interstitial lung disease [1] with a typical age of onset of 50-70, and is associated with environmental exposures. Pirfenidone and nintedanib are the two approved therapies for IPF. Both slow the rate of lung function decline relative to placebo but their mechanisms of action are unclear and neither agent reversed disease or substantially affected mortality in 1-year pivotal trials[2-4]. Despite well-defined guidelines for the definition and diagnosis of IPF[5], there is considerable heterogeneity in the molecular and pathological manifestations of disease, which may relate to heterogeneity in the rate of clinical progression[6].

There is a well-recognized genetic component to IPF susceptibility [7,8]. The genetics of both sporadic and familial forms of pulmonary fibrosis implicate two distinct types of lung epithelial cells: bronchial secretory cells and type 2 alveolar epithelial cells (AEC2). Familial studies have implicated genes involved in surfactant production (e.g., *SFPTA1, SFPTA2, SFPTC*)[9-14]. Surfactant proteins are uniquely produced by AEC2, and familial pulmonary fibrosis-associated variants in *SFPTC* results in misfolded SP-C protein, inducing endoplasmic reticulum stress in AEC2s[15]. Multiple genes involved in telomere maintenance have also been associated with IPF in sporadic and familial studies (e.g., *TERT, TERC, PARN* and *RTEL1*)[16-26]. Telomeres are repetitive DNA sequences added during mitosis to protect chromosome ends[27]. Telomere length decreases with increasing cell division and normal aging; IPF patients tend to have abnormally short telomeres in peripheral blood cells and alveolar epithelial cells[28], and among IPF patients, shorter telomere length in peripheral blood is associated with shorter survival[29].

Linkage, candidate gene, and genome-wide association studies (GWAS) have consistently shown strong associations between a common variant at rs35705950 in the chr11p15.5 locus and risk of sporadic and familial pulmonary fibrosis[7,30,31]. The SNP is in the promoter of the *MUC5B* gene and presence of the minor allele corresponds to increased expression of *MUC5B* in terminal bronchioles[32-34] and dramatically increased risk for IPF. The *MUC5B* allele is relatively common in the European ancestry population (MAF = 0·11)[35], yet IPF is a rare disorder with prevalence estimates ranging from 13-63 per 100,000 individuals in the United States[36]. Thus, although *MUC5B* carriers have a dramatically increased risk of developing IPF (OR=4·51)[18], the vast majority of *MUC5B* carriers will not.

While telomere shortening is associated with decreased survival among IPF patients, carriers of the *MUC5B* risk allele with IPF exhibit a slower rate of disease progression than non-carriers[37]. However, familial pulmonary fibrosis associated with genes implicating AEC2s (e.g., surfactant protein or telomerase mutations) typically presents earlier in life, and shorter telomeres are associated with a more aggressive disease course in sporadic IPF cases[28]. Taken together, shorter telomere length is associated with faster disease progression and rs35705950 is associated with slower progression. Thus, we hypothesized rare protein-altering variants that impact telomere length would be enriched in IPF cases lacking rs35705950 risk alleles and would be candidate *MUC5B* risk allele modifiers. To test this hypothesis, we assessed telomerase genetics in both common and rare coding variants in the context of *MUC5B* status for IPF risk and rate of disease progression within a large, well-characterized patient cohort.

Results

IPF Study population

We sequenced 1,510 IPF patients from clinical trials for IFN-γ1b, lebrikizumab and pirfenidone and observational cohorts from Vanderbilt University and UCSF. We stratified patients for analysis based on presence or absence of the risk SNP at the *MUC5B* locus (rs35705950). Detailed characteristics of the pooled study population and by cohort are presented in Table 1. Pirfenidone clinical trial samples were sequenced based on availability of DNA and signed consent for research. As such, not all trial participants were sequenced.

There were no significant differences for several clinical characteristics either overall or by trial (Supplemental Tables 1a-1d) aside from surgical biopsy status, with more DNA consented patients in the overall cohort having “possible or probable UIP” than the unconsented cohort (p=0·002). Non-IPF samples were used as comparators as described in the methods section. Patients were included in the study if they were of genetically determined European ancestry. A subset of patients in the observational cohorts had a family history of disease (N=13 in the Vanderbilt University cohort, N=7 in the UCSF cohort). We tested the Vanderbilt University cohort for differences in *MUC5B* risk allele frequency and candidate gene rare variant status finding no strong differences between the familial and idiopathic cohorts for either (*MUC5B* genotype p=0·73; candidate gene rare variant status p=0·99). As such the patients were included in our subsequent analyses.

IPF rs35705950 subset analysis: Rare and common variant burden enrichment in *TERT*

We first focused on the *TERT* gene due to its previously reported associations with IPF risk for both common and rare variants [18,25,38], hypothesizing that rare missense or loss of function variants would be differentially enriched in IPF patients with the risk allele at rs35705950 (those heterozygous or homozygous for the risk allele – termed *MUC5B* risk allele carriers) compared to those without (protective allele homozygotes – termed *MUC5B* non-risk allele carriers). We observed that 2·9% of *MUC5B* risk allele carriers carried a rare functional variant in *TERT* compared to 7·3% of *MUC5B* non-risk allele carriers (OR 0·40 [95% CI 0·24-0·66], $p=3{\cdot}9 \times 10^{-4}$ Table 2). This effect was consistent across cohorts as all cohorts with rare variants in *TERT* showed the same direction of effect (Supplemental Table 1e). Most unique *TERT* variants were missense rather than nonsense (N=34 missense, 5 nonsense), and all patients with a rare variant were heterozygous for that variant. Interestingly, the finding in the risk allele positive subgroup was mostly confined to the rs35705950 heterozygotes, with only two patients homozygous for the risk allele carrying a rare functional variant in *TERT* (1·5% of the rs35705950 homozygous IPF population). A full list of the variants observed is visualized in Figure 1 and annotated in Supplemental Table 2. We see no difference in the frequency of the GWAS common variant in *TERT* (rs2736100) between *MUC5B* risk allele carriers and non-carriers (minor allele frequency (MAF)=0·42 for both, OR 1·01 [95% CI 0·86-1·18], p=0·95) (Table 3).

IPF rs35705950 subset analysis: Rare variant burden enrichment in telomerase complex components and IPF risk loci

We extended the rare variant burden and common variant analyses to include 1) genes identified in previous GWAS, 2) genes encoding the other core telomerase components, and 3) genes associated with the larger telomerase complex. We observed a significant difference in allele frequency for the SNP (rs2609255) at the *FAM13A* locus between rs35705950 carriers (MAF=0·26) and those without (MAF=0·31; OR 0·77 [95% CI 0·65-0·91], $p=2{\cdot}5 \times 10^{-3}$). We did not observe a statistically significant difference in rare functional variation in any other candidate genes between IPF *MUC5B* risk allele carriers and non-carriers. Rare variation in *PARN* had a direction of effect similar to *TERT*, whereas 1·9% of *MUC5B* non-risk allele carriers carried a rare variant in *PARN* compared to 0·9% of *MUC5B* risk allele carriers, however it did not reach statistical significance (p=0·09). We did not observe

differences in rare functional variation in any other telomerase complex components or genes implicated in previous GWAS studies (Tables 4a and 4b).

IPF risk analyses

We next sought to confirm a previous report finding an increase in rare variant burden in *TERT* compared to controls [25]. In our overall IPF population, 4·2% of IPF cases carried a rare functional variant in *TERT*, compared to 1·7% of our non-IPF study population (OR 2·82 [95% CI 1·64-4·97], $p=2{\cdot}38 \times 10^{-4}$; Table 4b). However, the *TERT* RV frequency in our rs35705950 risk allele homozygote IPF population was only 1·5%, but increases to 3·1% in rs35705950 heterozygote IPF patients, and further to 7·3% in the rs35705950 protective allele homozygotes relative to controls (1·7%, 2·1% and 1·5% respectively by rs35705950 genotype).

Intriguingly, as we increase the stringency of the MAF cutoff from 0·01, we see a dramatic increase in the effect size in our overall IPF population compared to controls (MAF <0·005 OR 28·3 [95% CI 9·27-124·7], $p=2{\cdot}02 \times 10^{-7}$; MAF<0.0005 OR 44·5 [95% CI 12·3-288·9], $p=7{\cdot}53 \times 10^{-7}$; Supplemental Table 3). The frequency threshold restriction retains many of the variants in the IPF population (4·2% for MAF<0·01; 2·8% for MAF<0·0005), while removing most of the variants in the control population (1·7% for MAF<0·01; 0·1% for MAF<0·0005).

We replicated the association for the common *TERT* variant rs2736100 (MAF in IPF = 0·42, MAF in non-IPF = 0·49; OR 0·74 [95% CI 0·65-0·84], $p=2{\cdot}94 \times 10^{-9}$). Of note, the INSPIRE clinical trial cohort (N=340), was also included in the publication originally describing the common *TERT* variant, but the effect size holds in the remainder of our IPF population (N=l,170; OR 0·74 [95% CI 0·64-0·85], $p=2{\cdot}12 \times 10^{-5}$), comprising the majority of samples. Similarly, we assessed rare functional variation risk in all of the genes included in this study in addition to *TERT*. Results for all genes are shown in Tables 4a and 4b. We found an excess of rare variation in *PARN, TERC* and *RTEL1* in our IPF cases (*PARN* OR 8·74 [95% CI 2·16-59.65], p=0·007; *TERC* OR 4·27 [95% CI 1·3-16·63], p=0·04; *RTEL1* OR=2·88 [95% CI 1·59-5·29], $p=5{\cdot}44 \times 10^{-4}$). Interestingly, predicted nonsense variants in *RTEL1* were found only in the IPF cases (N=12/l,510; 0·79%) and none of the 1,874 non-IPF controls. The effect size for *PARN, RTEL1* or *TERC* did not change when applying a more stringent MAF cutoff, as with *TERT*. A full list of the variants observed in *PARN, TERC* and *RTEL1* are visualized in Figures 2-4 respectively and annotation for all are included in Supplemental Table 2.

Effect of rare functional variation of *TERT, PARN, TERC*, and *RTEL1* on age of IPF onset and rate of progression

We hypothesized that *TERT, PARN, TERC* and *RTEL1* rare functional variants would impact age of onset and rate of disease progression. Although we could not ascertain precise age of onset for most patients in our studies, the short survival time in IPF implies that age at time of assessment is a reasonable proxy. Patients with a *TERT* rare variant were significantly younger (mean age = 64·6 years) than patients without a rare variant (mean age = 67·4 years; p=0·005). Patients with *RIEL1* (mean age = 65·8) and *TERC* (mean age =

63·1) variants were also younger than patients without a rare variant (mean age = 67·3 for both), however the difference did not meet statistical significance for either. There was no difference in age of onset for a patient with a *PARN* variant (p=0·97). (Supplementary Table 4). The average age in patients with a rare variant in either *TERT, PARN, TERC* or *RTEL1* (mean age = 65·1) was significantly younger than patients without (mean age = 67·1; p=0·004). Conversely, as described above, *MUC5B* risk allele carriers were significantly older (mean age = 68·1 years) than patients without (mean age 65·5 years; $p=1{\cdot}90 \times 10^{-8}$). Next, we tested the effect of a rare functional variant of *TERT, PARN, TERC, or RTEL1* on rate of disease progression (ΔFVC% predicted) over time in the placebo arms of the ASCEND and CAPACITY clinical trial patients in a linear model of change in lung function by telomere length and the individual components of the GAP score. We observed that patients with such a variant had more rapid rates of FVC decline (1·66 % Pred. FVC/month) than patients without one (0·83 % Pred. FVC/monthp=0·02) (Supplemental Figure 1). Thus, rare variation in these genes has a negative impact on age of onset and rate of disease progression in IPF patients. In contrast, IPF *MUC5B* risk allele carriers were older and exhibited a slower rate of disease progression than *MUC5B* non-risk allele patients.

Effect of rare functional variation of *TERT, PARN, TERC*, and *RTEL1* and rs35705950 genotype on telomere length

We hypothesized that variation in *TERT, PARN, TERC* and *RTEL1* may affect telomere length. We first tested the effect of the *TERT* rare and common variants on telomere length, calculated in our patient samples from the WGS data, and expanded our analysis to include WGS from several clinical trials (Supplemental Table 5). Telomere length determined by WGS was significantly correlated with telomere length determined by TRF (telomere restriction fragment) analysis (R^2=0·47, p<0·0001, Supplemental Figure 2). The telomeres of patients with a rare functional variant in *TERT* (mean length = 2·50kb) were significantly shorter than those of patients in our comparison group (patients without a rare variant or a *TERT* common risk allele; mean length = 2·76kb; $p=1{\cdot}54 \times 10^{-5}$, Table 5). We also assessed telomere length as a function of the common *TERT* variant rs2736100 [18]. We observed that patients homozygous for the protective allele at rs2736100 (CC genotype; mean length = 2·76kb) tended to have longer telomeres than the homozygous risk allele patients (AA genotype; mean length = 2·65kb, $p=1.50 \times 10^{-3}$).

We extended our analysis to include the other telomerase complex and maintenance genes identified in our rare variant burden risk analysis: *PARN, TERC* and *RTEL1*. Mean telomere lengths for *PARN, TERC* and *RTEL1* rare variant carriers were 2·66, 2·32 and 2·57kb, respectively (Table 5). IPF patients harbored at least one rare variant in *TERT, PARN, TERC* or *RTEL1* more frequently than non-IPF patients from other clinical trials (8·57% in IPF vs. 2·37% for others, $p=2{\cdot}44 \times 10^{-8}$, Figure 5). Similarly, we observed IPF patients' telomeres were significantly shorter than patients with other diseases, 481bp shorter on average at age 50 (16% decline, p=0·0001, Figure 6). Rare variants in *TERT, PARN, TERC* or *RTEL1* were related to shorter telomeres in most cases among IPF patients and in all cases among all patients. Collectively, patients with rare variants in *TERT, PARN, TERC* or *RTEL1* exhibited telomeres that were on average 348bp shorter per variant. There was no effect on telomere length when stratifying patients by rs35705950 (p=0·94). The effect of a rare

variant on shortening telomeres appears to be cumulative with the common *TERT* SNP, as those patients with at least one variant in *TERT* exhibited further telomere shortening beyond the effect of common *TERT* risk alleles (Figure 7). The interaction between having a rare variant and having at least one common TERT reduced-activity variant was not statistically significant (p=0·84). We did not observe a similar effect of *TERT* common genotype in patients with a rare variant in *PARN, TERC* or *RTEL1*, or for our non-IPF comparison samples.

Telomere length can also be affected by environmental variables such as smoking. We investigated the effect of smoking status on telomere length in our cohorts with available smoking history data. Among our IPF patients with known self-reported smoking status, smoking is not significantly related to telomere length (p>0·05). Among all patients including both IPF and non-IPF in our study with known smoking status, both age and smoking status are associated with telomere length ($p<1 \times 10^{-16}$ for each). Among all patients, in a model of telomere length by age, IPF case status and smoking status, IPF case status $p<1 \times 10^{-16}$ whereas smoking status p=0·001. As such, while there is a modest effect of smoking on telomere length, it is underwhelming compared to IPF case status. This is exemplified by looking at the estimated effect of smoking on telomere length. The estimated effect of smoking status on telomere length is 70 bp, compared to 348 bp per rare variant in *TERT, TERC, PARN* or *RTEL1* (data not shown).

Effect of telomere length and rs35705950 genotype on rate of progression and pirfenidone treatment response

Next, we examined whether telomere length was related to rate of IPF disease progression and how this relationship compares to the prognostic power of GAP (Gender, Age, and lung Physiology) variables for predicting IPF survival[39] In a linear model of change in lung function by telomere length and the individual components of the GAP score in placebo-treated patients we find that shorter telomeres are significantly related to faster progression (decline of 1·7% predFVC/kb/year, p=0·002). This relationship appears stronger than the GAP components themselves in the same model (p>0·1). We observe that patients with baseline telomere length above the median in the placebo arms of the CAPACITY and ASCEND clinical trials had slower disease progression than patients with telomere lengths below the median (Figure 8a). Results were similar for the trials individually (Supplemental Figures 3-4). In an expanded cohort that included both placebo and pirfenidone-treated patients and using a model with terms for treatment and a treatment-telomere interaction, we observed a significant interaction (increased decline of 3.7% predFVC/year with telomeres below median length, p=0·0001) between telomere length and treatment on lung function decline. Shorter telomeres at the study baseline timepoint predict more rapid FVC decline in IPF patients, pirfenidone demonstrated treatment benefit regardless of telomere length status ($p=4{\cdot}24\times10^{-8}$ for telomere length less than the median, $p=441\times10^{-3}$ for telomere length greater than the median). Similarly, we observe that patients with the risk allele at rs35705950 had slower disease progression than patients without (p=0·0067) (Figure 8b). Results were similar for the trials individually (Supplemental Figures 5-6).

Discussion

Genetic studies of sporadic and familial IPF have revealed substantial heritability and strong associations with genes involved both lung epithelial cells and telomere maintenance. The strong risk of developing IPF conferred by the common variant rs35705950 but low prevalence of IPF represents a set of conditions well suited to a genetic modifier screen conditioned on rs35705950 genotype. The previously described slower rates of IPF disease progression observed in rs35705950 carriers[37] and faster rates of disease progression observed in patients with shorter telomeres[28] suggest that genetic factors may underlie mechanistic and clinical heterogeneity in IPF pathogenesis. Here we show in a large cohort of well-characterized IPF patients that: 1) rare functional variants in in telomere maintenance genes are more frequent in IPF compared to controls and particularly in IPF patients lacking the common risk allele at rs3570590; 2) IPF patients have increased rates of rare functional variants in telomere maintenance genes and shorter telomeres than patients with other common diseases; and 3) in placebo arms of clinical studies, IPF patients with shorter telomeres exhibit a faster rate of lung function decline than those with longer telomeres, but the rate of lung function decline is reduced in patients treated with pirfenidone despite differences in telomere length. Taken together, these findings suggest that *MUC5B*-related and telomere-related mechanisms may give rise to pathologically and clinically different subsets of IPF.

A working hypothesis for IPF pathogenesis holds that the disease process originates in AEC2, which exhibit increased levels of endoplasmic reticulum stress and dysregulated proteostasis, mitophagy, and/or autophagy, resulting in increased mitotic rates leading to telomere attrition and cellular senescence and/or apoptosis. This disruption in AEC2 homeostasis precipitates inflammation and mesenchymal cell activation of a wound repair response, leading to excessive myofibroblast activation, extracellular matrix deposition, and interstitial fibrosis, which progressively obliterates normal alveolar architecture and compromises gas exchange[1,40]. Genetic studies of pulmonary fibrosis implicate genes involved in bronchial secretory cells and AEC2s. The common IPF risk variant in the *MUC5B* promoter corresponds to increased *MUC5B* expression in terminal bronchioles[32-34], while rare familial forms of IPF are associated with coding variants in *SFPTC* that result in misfolding of surfactant proteins and endoplasmic reticulum stress in AEC2s[15]. Both sporadic and familial forms of pulmonary fibrosis have been associated with multiple telomerase-related genes [16-26]. While each of these genetic lesions is hypothesized to contribute to increased susceptibility of AEC2s to injury, they are mechanistically distinct.

A complex of RNA template (encoded by *TERC*) and a reverse transcriptase (*TERT*) represent the core elements of telomerase[27] which regulates telomere length; the shelterin complex comprises multiple additional proteins that protect chromosomal ends from double strand break repair machinery during mitosis[41]. When telomeres shorten beyond a critical point, cells undergo cell cycle arrest and undergo apoptosis or senescence. Preclinical models have shown that genetic defects in components of the telomerase complex in alveolar epithelial cells, but not other cell types, predispose animals to the development of spontaneous pulmonary fibrosis with age, and confer increased susceptibility to epithelial cell injury upon bleomycin challenge[42-44]. AEC2s of mice with engineered deficits in

telomerase activity exhibit a senescent phenotype; the senescence-associated secretory proteome (SASP) is hypothesized to stimulate the aberrant wound healing response that manifests as interstitial fibrosis[45].

Telomere shortening is associated with decreased survival among IPF patients[28,29] and, as we show here, more rapid rates of lung function decline in placebo-treated but not pirfenidone-treated patients. However, IPF patients carrying the *MUC5B* risk allele have longer survival from the time of diagnosis than non-carriers[37], and, as we show here, slower rates of disease progression than homozygous non-carriers in placebo-treated patients, and that pirfenidone treatment decreases the rate of disease progression relative to placebo in both *MUC5B* carriers and noncarriers. This differential disease trajectory suggests that *MUC5B*-related and telomerase-related IPF may have different mechanisms of pathogenesis. Considering the parenchymal location of IPF pathology and the bronchial location of mucin expression and secretion, there may be a paracrine effect of MUC5B dysregulation in terminal bronchioles that indirectly injures alveolar epithelium[8], whereas telomerase defects directly impact alveolar epithelial cells leading to a more aggressive disease course.

AEC2s represent a stem cell niche for alveolar epithelial cells[46,47]. During branching morphogenesis of the airways in embryogenesis, alveolar epithelial stem cells undergo numerous divisions to generate the large surface area of lung epithelium to support gas exchange[48]. Further cell divisions postnatally support lung growth, and in adulthood to replace epithelium lost to injury via infection or environmental insults. Hence, due to this considerable mitotic demand, the alveolar epithelial stem cell population may consume its telomeres at a faster rate than other stem cell niches[28], consistent with observations that AEC2s from IPF patients have compromised self-renewal capacity[49]. IPF is a disease of aging, thus we speculate that modest defects in telomerase function present with pathology in lung epithelium before other tissues and account for the strong relationship between telomerase defects and pulmonary fibrosis.

A significant challenge in designing interventional clinical studies in IPF and making treatment decisions for patients with IPF is the variability in rates of disease progression. Here we show in pivotal trials of pirfenidone, *MUC5B* risk allele genotype and telomere shortening correspond to slower and faster rates of disease progression, respectively, as measured by FVC decline over a one-year period. Intriguingly, patients treated with pirfenidone exhibited decreased rates of disease progression regardless of *MUC5B* genotype or baseline telomere length. Future studies should take these and other biomarkers that predict rates of disease progression into account both to: 1) enrich for patients at risk of disease progression to enable potentially shorter, smaller clinical studies that use change in FVC or death as an outcome measure, and 2) to avoid confounding by appropriate stratification of patients according to these prognostic biomarkers at randomization.

Limitations

A major limitation of this study is the lack of screened controls as an IPF comparator group. As such we are accepting the low population prevalence of IPF in our comparator group. While this is a limitation, it is unlikely to affect results due to the rarity of IPF. Furthermore,

our selection of diseased cohorts for our comparator group potentially limits our findings if there are common genetic components to these diseases. Our analyses of disease progression in clinical studies of pirfenidone are post hoc and confined to the subset of patients in those studies from whom matching genetic and outcome data were available; any conclusions related to the efficacy of pirfenidone in genetically defined subsets would need to be confirmed prospectively.

Conclusion

Through whole genome sequencing of a large cohort of IPF patients, we have shown complex genetic relationships between the common SNP at rs35705950 in the *MUC5B-TOLLIP* locus, rare variation in *TERT* and other genes encoding components of the telomerase complex, telomere length, age of onset, and rate of progression. Taken together, our findings suggest that *MUC5B*-related and telomerase-related mechanisms of IPF pathogenesis may yield a common clinical diathesis via distinct pathogenic routes with implications for clinical trial design and patient management.

Methods

Sequencing

The WGS data was created by Illumina X10 sequencers and then processed using the BWA/GATK best practices pipeline. The read depth for the WGS data is 30x. All the sequencing data was subject to quality control as well as checked for concordance with fingerprint data taken prior to sequencing.

Genotyping

WGS short reads were mapped to GRCh38 including alternate (ALT) assemblies using ALT-aware version of BWA to generate bam files. QC of the bam files and variant calling was performed using GATK best practices joint genotyping pipeline to generate a single VCF file. Sample contamination was determined with verifyBamID and samples with a freemix parameter more than 0·03 were excluded. After filtering for GATK genotype quality greater than 90, samples with heterozygote concordance with snp chip data less than 75% were removed. The called variants were then processed using ASDPEx to filter out spurious calls in the ALT regions[50].

Telomere Length

Telomere length was determined by running the Telomere Hunter program on all raw reads passing Illumina QC, adjusting by GC content according to default program parameters, and normalizing to total number of reads analyzed.

Study Population

Our study population consisted of 1,510 IPF patients from clinical trials for IFN-γ1b - INSPIRE (N=340), lebrikizumab – RIFF (N=81) and pirfenidone - CAPACITY (N=312) and ASCEND (N=271), and also from cohorts collected at Vanderbilt University (N=440) and UCSF (N=66).

Clinical trials INSPIRE, CAPACITY and ASCEND have been described previously[2,3,51]. Samples were sequenced based on availability of DNA and signed consent for research. There were no significant differences for several clinical characteristics either overall or by trial (Supplemental Tables 1a-1d) aside from surgical biopsy status, with more DNA consented patients in the overall cohort having "possible or probable UIP" than the unconsented cohort (p=0.002).

Vanderbilt cohort: The Vanderbilt Clinical ILD Registry was started in 2005 and enrollment is offered to all patients seen in the Interstitial Lung Disease clinic at Vanderbilt. All IPF patients in the Vanderbilt Clinical ILD registry from whom DNA was available for WGS were included in the Vanderbilt cohort. Patients with a family history of ILD (Familial Interstitial Pneumonia) were excluded from the WGS cohort. In this registry, a diagnosis of IPF is adjudicated by ILD expert clinicians (LHL, JAK) according to ATS/ERS consensus criteria. Only patients with probable or definite IPF were included in the Vanderbilt WGS cohort.

RIFF clinical trial cohort A: Patients enrolled in RIFF cohort A were 40 years or older, have a diagnosis of IPF based on the 2011 ATS/ERS/JRS/ALAT consensus statement on IPF[52] within the previous 5 years from time of screening (confirmed at baseline), and have a central review assessment of an HRCT and SLB if available performed during the screening period or within 12 months prior to the start of screening with additional multidisciplinary discussion to finalize diagnosis in the event of disparate results for HRCT and SLB. Additional inclusion criteria include FVC > 40% and < 100% of predicted at screening, DLco > 25% and <90% of predicted at screening, the ability to walk > 100 meters unassisted in 6 minutes, and no background IPF therapy for >4 weeks prior to randomization. Relevant exclusion criteria include evidence of other known causes of ILD, a lung transplant expected within 12 months of screening, evidence of clinically significant lung disease other than IPF (e.g. asthma or COPD), post-bronchodilator (FEV1)/FVC ratio <0·7 at screening, post-bronchodilator response, and hospitalization due to an exacerbation of IPF within 4 weeks prior to or during screening.

UCSF cohort: The UCSF patients were drawn from a longitudinal, prospective, registry of pulmonary fibrosis patients seen in the UCSF interstitial lung disease clinic who were MUC5B risk allele carriers. The diagnosis of IPF was made by multidisciplinary team discussions, after in-person assessments, according to published guidelines[52].

Non-IPF control cohort: Non-IPF controls were obtained from clinical trials cohorts from age-related macular degeneration, asthma and rheumatoid arthritis. All patients included in the study were ≥40 years of age and of genetically determined European ancestry based on comparison to samples from the HapMap project.

Sample QC

There were 1,764 IPF samples and 2,420 non-IPF controls before applying quality control measures. Samples were excluded if the call rate was < 90% (12 IPF; 0 non-IPF). IBD analysis was used to detect and filter out relatedness in the dataset (50 IPF; 7 non-IPF).

Samples were removed if they showed excess heterozygosity with more than three standard deviations of the mean (27 IPF; 18 non-IPF). For IBD analysis (patient relatedness), the PI_HAT cutoff for pairwise relatedness is >04. Principal components were generated including the HapMap populations, and samples were excluded if they did not cluster near the CEU population (165 IPF; 521 non-IPF). There are 1,510 IPF patients and 1874 non-IPF controls that were included in the final analysis.

SNP QC and Batch Effects

The common variant analysis was restricted to SNPs with MAF >= 1% and the rare variant analysis was restricted to exonic SNPs with MAF < 1%. Variants were included in the rare variant analysis if the PolyPhen score was damaging (>=0·5) or if they were high impact variants (stop gain, stop loss, frameshift, etc).

Sample genotypes were set to missing if the GQ score was < 20 and SNPs were removed if the missingness was > 5%. SNPs were filtered if the HWE P was < $1X10^{-8}$. Missingness, HWE and MAF filters were applied to each batch separately and variants passing the criteria in all batches were included in the final analysis. In addition, differential missingness ($p<lX10^{-6}$) by sequence site was used to filter out SNPs.

DNA samples were sequenced at two sites over three years. As such, there is the potential for batch effects. To test for if our QC could adequately account for this, we divided our IPF population by batch and tested for differences in rare variant frequency for the genes in our analysis. For all genes tested in this study, we did not see any significant differences in carrier frequency resulting from batch (p>0.05 for all).

Analysis

Logistic regression was used to test association in the common variant analysis, while adjusted for age and sex. PLINK was used for the common variant analysis.

The rare variant gene burden results looked at the cumulative effect of rare variants (MAF < 1%) in the various subsets of analyses. rvtest was used to perform the CMC gene burden test and this was adjusted for age and sex.

Pairwise comparisons were performed with Student’s T-test after confirming normal distribution and the validity of the assumption of equal variance of data. Linear modeling was performed in R. Change in FVC was modeled with a mixed effects model of time, patient, arm, age and the variable to be tested (*MUC5B*, telomere length).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Role of Funding Source

Genentech funded the whole genome sequencing of DNA from participants used in this study. HRC and PJW were supported in part by the Nina Ireland Program in Lung Health. NIH K08HL130595, Francis Family Foundation,

and the Pulmonary Fibrosis Foundation provided support for storage and collection of specimens and clinical data and relevant salary support for JAK. Employees of Genentech were responsible for the analysis, interpretation of data and writing of the report. AD and ARA had complete access to the data. JRA and BLY were responsible for the decision to submit the manuscript for publication.

FUNDING: Genentech, Inc., NIH K08HL130595, Francis Family Foundation, Pulmonary Fibrosis Foundation, Nina Ireland Program in Lung Health, P01HL92870, Department of Veterans Affairs

References

1. Richeldi L, Collard HR, Jones MG. Idiopathic pulmonary fibrosis. Lancet 2017; 389(10082): 1941–52. [PubMed: 28365056]
2. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. Lancet 2011; 377(9779): 1760–9. [PubMed: 21571362]
3. King TE Jr., Bradford WZ, Castro-Bernardini S, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. N Engl J Med 2014; 370(22): 2083–92. [PubMed: 24836312]
4. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med 2014; 370(22): 2071–82. [PubMed: 24836310]
5. Raghu G, Rochwerg B, Zhang Y, et al. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline: Treatment of Idiopathic Pulmonary Fibrosis. An Update of the 2011 Clinical Practice Guideline. Am J Respir Crit Care Med 2015; 192(2): e3–19. [PubMed: 26177183]
6. Ley B, Collard HR, King TE Jr., Clinical course and prediction of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 2011; 183(4): 431–40. [PubMed: 20935110]
7. 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; 96(4): 1567–91. [PubMed: 27630174]
8. Kropski JA, Blackwell TS, Loyd JE. The genetic basis of idiopathic pulmonary fibrosis. Eur Respir J 2015; 45(6): 1717–27. [PubMed: 25837031]
9. Nathan N, Giraud V, Picard C, et al. Germline SFTPA1 mutation in familial idiopathic interstitial pneumonia and lung cancer. Hum Mol Genet 2016; 25(8): 1457–67. [PubMed: 26792177]
10. van Moorsel CH, Ten Klooster L, van Oosterhout MF, et al. SFTPA2 Mutations in Familial and Sporadic Idiopathic Interstitial Pneumonia. Am J Respir Crit Care Med 2015; 192(10): 1249–52. [PubMed: 26568241]
11. Wang Y, Kuan PJ, Xing C, et al. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. Am J Hum Genet 2009; 84(1): 52–9. [PubMed: 19100526]
12. Nogee LM, Dunbar AE 3rd, Wert SE, Askin F, Hamvas A, Whitsett JA. A mutation in the surfactant protein C gene associated with familial interstitial lung disease. N Engl J Med 2001; 344(8): 573–9. [PubMed: 11207353]
13. Lawson WE, Grant SW, Ambrosini V, et al. Genetic mutations in surfactant protein C are a rare cause of sporadic cases of IPF. Thorax 2004; 59(11): 977–80. [PubMed: 15516475]
14. van Moorsel CH, van Oosterhout MF, Barlo NP, et al. Surfactant protein C mutations are the basis of a significant portion of adult familial pulmonary fibrosis in a dutch cohort. Am J Respir Crit Care Med 2010; 182(11): 1419–25. [PubMed: 20656946]
15. Lawson WE, Polosukhin VV, Stathopoulos GT, et al. Increased and prolonged pulmonary fibrosis in surfactant protein C-deficient mice following intratracheal bleomycin. Am J Pathol 2005; 167(5): 1267–77. [PubMed: 16251411]
16. Armanios MY, Chen JJ, Cogan JD, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. N Engl J Med 2007; 356(13): 1317–26. [PubMed: 17392301]
17. Tsakiri KD, Cronkhite JT, Kuan PJ, et al. Adult-onset pulmonary fibrosis caused by mutations in telomerase. Proc Natl Acad Sci U S A 2007; 104(18): 7552–7. [PubMed: 17460043]
18. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet 2013; 45(6): 613–20. [PubMed: 23583980]
19. Stuart BD, Choi J, Zaidi S, et al. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. Nat Genet 2015; 47(5): 512–7. [PubMed: 25848748]

20. Kannengiesser C, Borie R, Menard C, et al. Heterozygous RTEL1 mutations are associated with familial pulmonary fibrosis. Eur Respir J 2015; 46(2): 474–85. [PubMed: 26022962]
21. Cogan JD, Kropski JA, Zhao M, et al. Rare variants in RTEL1 are associated with familial interstitial pneumonia. Am J Respir Crit Care Med 2015; 191(6): 646–55. [PubMed: 25607374]
22. Kropski JA, Reiss S, Markin C, et al. Rare Genetic Variants in PARN are Associated with Pulmonary Fibrosis in Families. Am J Respir Crit Care Med 2017.
23. Borie R, Tabeze L, Thabut G, et al. Prevalence and characteristics of TERT and TERC mutations in suspected genetic pulmonary fibrosis. Eur Respir J 2016; 48(6): 1721–31. [PubMed: 27836952]
24. Newton CA, Batra K, Torrealba J, et al. Telomere-related lung fibrosis is diagnostically heterogeneous but uniformly progressive. Eur Respir J 2016; 48(6): 1710–20. [PubMed: 27540018]
25. Petrovski S, Todd JL, Durheim MT, et al. An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. Am J Respir Crit Care Med 2017.
26. Alder JK, Cogan JD, Brown AF, et al. Ancestral mutation in telomerase causes defects in repeat addition processivity and manifests as familial pulmonary fibrosis. PLoS Genet 2011; 7(3): e1001352. [PubMed: 21483807]
27. Armanios M Telomeres and age-related disease: how telomere biology informs clinical paradigms. J Clin Invest 2013; 123(3): 996–1002. [PubMed: 23454763]
28. Alder JK, Chen JJ, Lancaster L, et al. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. Proc Natl Acad Sci U S A 2008; 105(35): 13051–6. [PubMed: 18753630]
29. Stuart BD, Lee JS, Kozlitina J, et al. Effect of telomere length on survival in patients with idiopathic pulmonary fibrosis: an observational cohort study with independent validation. Lancet Respir Med 2014; 2(7): 557–65. [PubMed: 24948432]
30. Seibold MA, Wise AL, Speer MC, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. N Engl J Med 2011; 364(16): 1503–12. [PubMed: 21506741]
31. Noth I, Zhang Y, Ma SF, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. Lancet Respir Med 2013; 1(4): 309–17. [PubMed: 24429156]
32. Seibold MA, Smith RW, Urbanek C, et al. The idiopathic pulmonary fibrosis honeycomb cyst contains a mucocilary pseudostratified epithelium. PLoS One 2013; 8(3): e58658. [PubMed: 23527003]
33. Nakano Y, Yang IV, Walts AD, et al. MUC5B Promoter Variant rs35705950 Affects MUC5B Expression in the Distal Airways in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med 2016; 193(4): 464–6. [PubMed: 26871673]
34. Helling BA, Gerber AN, Kadiyala V, et al. Regulation of MUC5B Expression in Idiopathic Pulmonary Fibrosis. Am J Respir Cell Mol Biol 2017.
35. Yates A, Akanni W, Amode MR, et al. Ensembl 2016. Nucleic Acids Res 2016; 44(D1): D710–6. [PubMed: 26687719]
36. Ley B, Collard HR. Epidemiology of idiopathic pulmonary fibrosis. Clin Epidemiol 2013; 5: 483–92. [PubMed: 24348069]
37. Peljto AL, Zhang Y, Fingerlin TE, et al. Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. JAMA 2013; 309(21): 2232–9. [PubMed: 23695349]
38. Mathai SK, Yang IV, Schwarz MI, Schwartz DA. Incorporating genetics into the identification and treatment of Idiopathic Pulmonary Fibrosis. BMC Med 2015; 13: 191. [PubMed: 26400796]
39. Ley B, Ryerson CJ, Vittinghoff E, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. Ann Intern Med 2012; 156(10): 684–91. [PubMed: 22586007]
40. Pardo A, Selman M. Lung Fibroblasts, Aging, and Idiopathic Pulmonary Fibrosis. Ann Am Thorac Soc 2016; 13(Supplement_5): S417–S21. [PubMed: 28005427]
41. de Lange T Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev 2005; 19(18): 2100–10. [PubMed: 16166375]
42. Naikawadi RP, Disayabutr S, Mallavia B, et al. Telomere dysfunction in alveolar epithelial cells causes lung remodeling and fibrosis. JCI Insight 2016; 1(14): e86704. [PubMed: 27699234]

43. Povedano JM, Martinez P, Flores JM, Mulero F, Blasco MA. Mice with Pulmonary Fibrosis Driven by Telomere Dysfunction. Cell Rep 2015; 12(2): 286–99. [PubMed: 26146081]
44. Alder JK, Barkauskas CE, Limjunyawong N, et al. Telomere dysfunction causes alveolar stem cell failure. Proc Natl Acad Sci U S A 2015; 112(16): 5099–104. [PubMed: 25840590]
45. van Deursen JM. The role of senescent cells in ageing. Nature 2014; 509(7501): 439–46. [PubMed: 24848057]
46. Treutlein B, Brownfield DG, Wu AR, et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 2014; 509(7500): 371–5. [PubMed: 24739965]
47. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. Nature 2014; 507(7491): 190–4. [PubMed: 24499815]
48. Metzger RJ, Klein OD, Martin GR, Krasnow MA. The branching programme of mouse lung development. Nature 2008; 453(7196): 745–50. [PubMed: 18463632]
49. Liang J, Zhang Y, Xie T, et al. Hyaluronan and TLR4 promote surfactant-protein-C-positive alveolar progenitor cell renewal and prevent severe pulmonary fibrosis in mice. Nat Med 2016; 22(11): 1285–93. [PubMed: 27694932]
50. Jager M, Schubach M, Zemojtel T, Reinert K, Church DM, Robinson PN. Alternate-locus aware variant calling in whole genome sequencing. Genome Med 2016; 8(1): 130. [PubMed: 27964746]
51. King TE Jr., Albera C, Bradford WZ, et al. Effect of interferon gamma-1b on survival in patients with idiopathic pulmonary fibrosis (INSPIRE): a multicentre, randomised, placebo-controlled trial. Lancet 2009; 374(9685): 222–8. [PubMed: 19570573]
52. 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(6): 788–824. [PubMed: 21471066]

Research in Context

Evidence before this study

We searched PubMed between Feb 28 2017 and March 9 2018 with the search terms “pulmonary fibrosis”, “genome wide”, “familial”, “rs35705950” and “telomere”. Previous linkage studies of familial pulmonary fibrosis and exome sequencing studies of idiopathic pulmonary fibrosis (IPF) patients have identified variants in genes in telomere maintenance genes, notably *TERT, TERC, PARN*, and *RTEL1*. Furthermore, previous genome-wide association studies have identified and confirmed several independent loci that confer susceptibility to idiopathic pulmonary fibrosis risk. From these studies, the locus with the strongest effect size is on chr11p15.5 which contains the genes *MUC5B* and *TOLLIP*. The risk variant (SNP rs35705950) at this locus is carried by approximately 60% of IPF patients. Additionally, it has been shown that carriers of this risk allele tend to survive longer from the time of IPF diagnosis than non-risk allele carriers. Aside from the differences in survival time, there is a lack of information on the differences underlying IPF patients carrying the *MUC5B* risk allele and those that do not.

Added value of this study

To our knowledge, this study contains the greatest number of IPF patients with whole genome sequence published to date, comprising 1,510 IPF patients and an additional 1,874 non-IPF controls. Furthermore, it is the first to compare differences in rates of disease progression and clinical benefit from pirfenidone in IPF clinical trial populations based on presence or absence of the *MUC5B* risk allele. We report statistically significant differences between IPF patients with a *MUC5B* risk allele compared to those without for the frequency of rare missense or loss of function variants in *TERT*. Furthermore, we replicate findings that rare variation in *TERC*, *PARN* and *RTEL1* is also enriched in IPF patients compared to controls and show that presence of these variants is negatively correlated with telomere length in IPF patients. Finally, we show the effect of telomere length and presence of these variants on disease progression and pirfenidone treatment response.

Implications of all the available evidence

MUC5B genotype and telomere length significantly influence the rate of lung function decline in IPF. Patients with the *MUC5B* risk allele have slower disease progression, while patients lacking the *MUC5B* allele have faster disease progression and are more likely to have rare genetic defects in telomere maintenance. While *MUC5B* risk allele status and telomere length may differ between IPF patients, there is evidence for placebo-adjusted benefit from pirfenidone regardless of these baseline characteristics. The effects of *MUC5B* genotype, telomere maintenance genetics, and telomere length on IPF progression will have important ramifications for patient management, target discovery, and the design and interpretation of clinical trials in IPF.

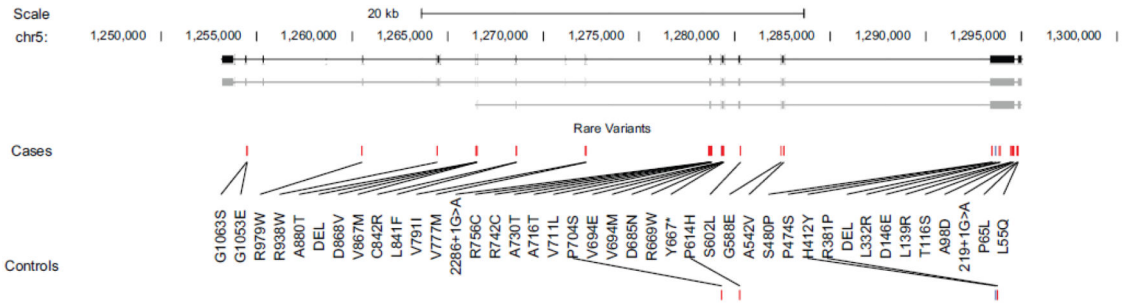

Figure 1.

Gene diagram of *TERT* and rare variant location. SNPs are identified by amino acid change and position and colored red if the MAF is ≤ 0.001, blue if the MAF > 0.001. DEL indicates a frameshift affecting >1 amino acids.

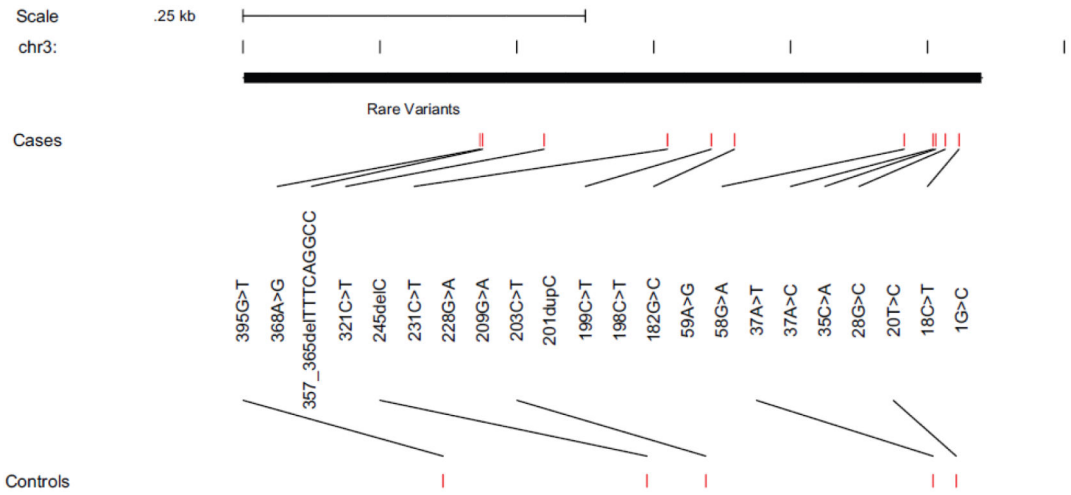

Figure 2.

Gene diagram of *TERC* and rare variant location. SNPs are identified by nucleic acid position and base change and colored red if the MAF is < 0.001, blue if the MAF > 0.001.

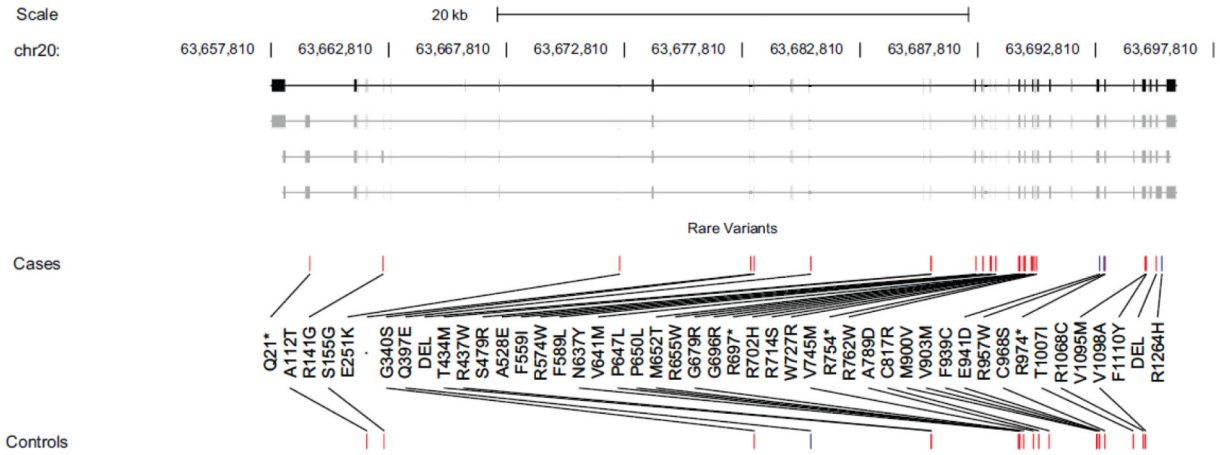

Figure 3.

Gene diagram of *RTEL1* and rare variant location. SNPs are identified by amino acid change and position and colored red if the MAF is < 0.001, blue if the MAF > 0.001. DEL indicates a frameshift affecting >1 amino acids. "." indicates variation affecting a predicted splice site.

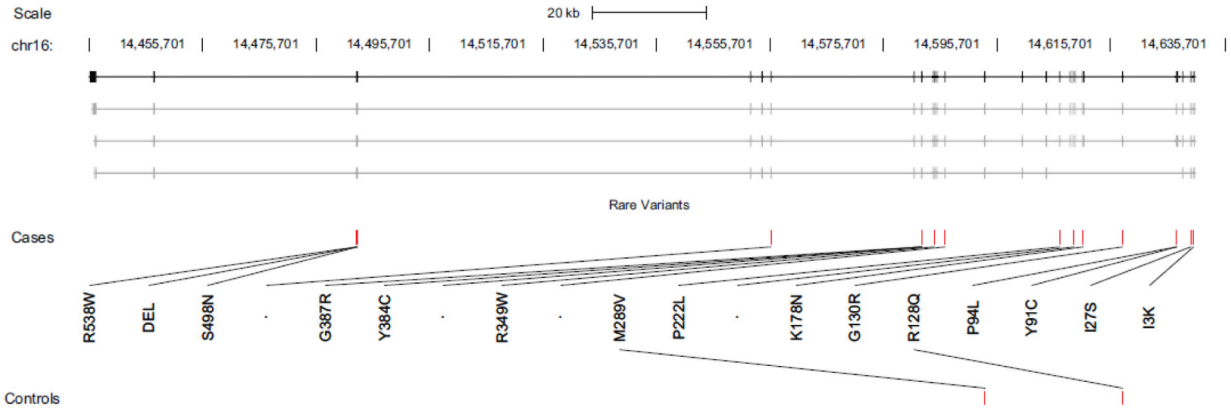

Figure 4.

Gene diagram of *PARN* and rare variant location. SNPs are identified by amino acid change and position and colored red if the MAF is < 0.001, blue if the MAF > 0.001. DEL indicates a frameshift affecting >1 amino acids. "." indicates variation affecting a predicted splice site. Exons have 200 bp added both up and downstream to enhance visibility.

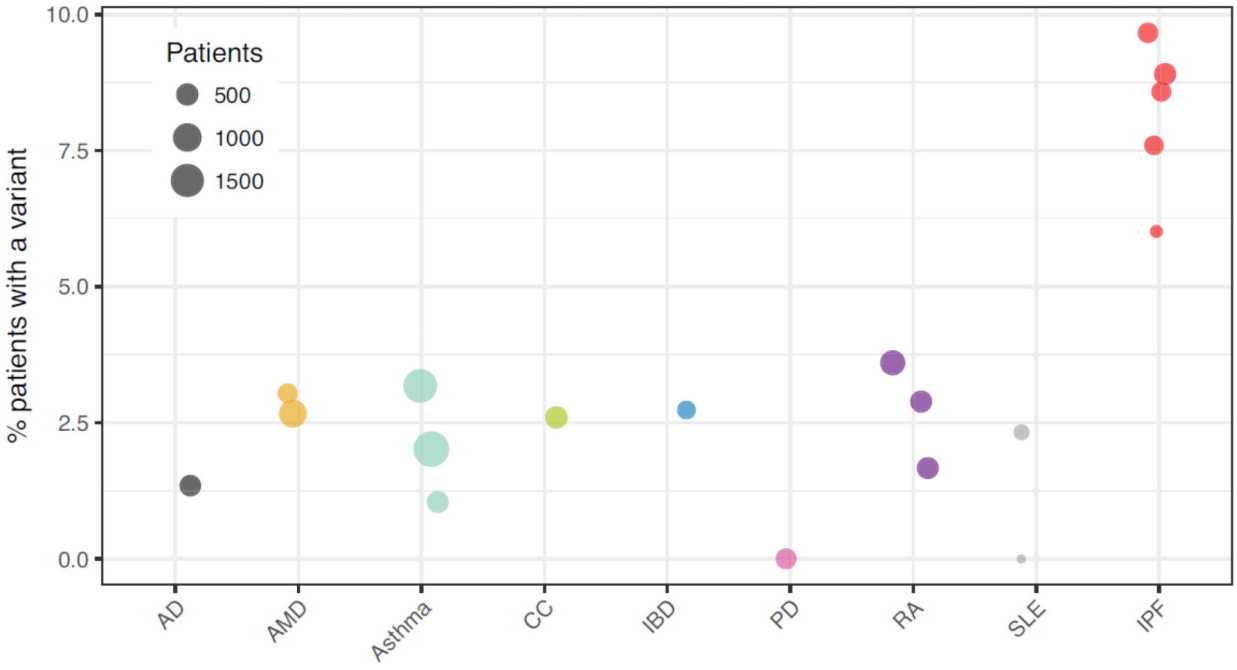

Figure 5.

Rates of telomere complex gene rare protein-altering variants in clinical trial patients. Each patient was assessed for the presence of a missense/nonsense variant in TERT, or RTEL1 or any variant in TERC. The fraction of patients in each trial that have at least one such variant are plotted. The size of each point is proportional to the number of patients in that trial.

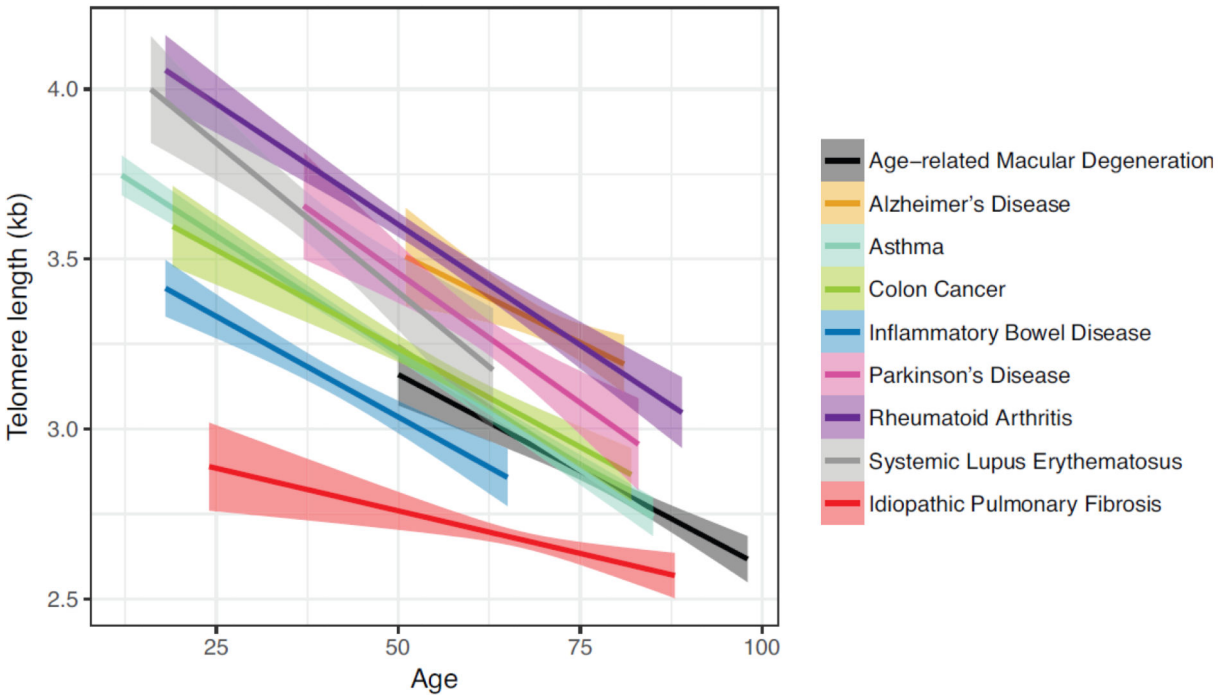

Figure 6.

Telomere lengths of clinical trial patients. Mean telomere lengths as measured by TelomereHunter analysis of whole genome sequencing of blood samples from patients in various clinical trials. Age vs. telomere length by disease. Each line is a least squares fit to all patients from a given primary disease indication. Shaded areas are 95% confidence interval estimates for the linear fit to the true population.

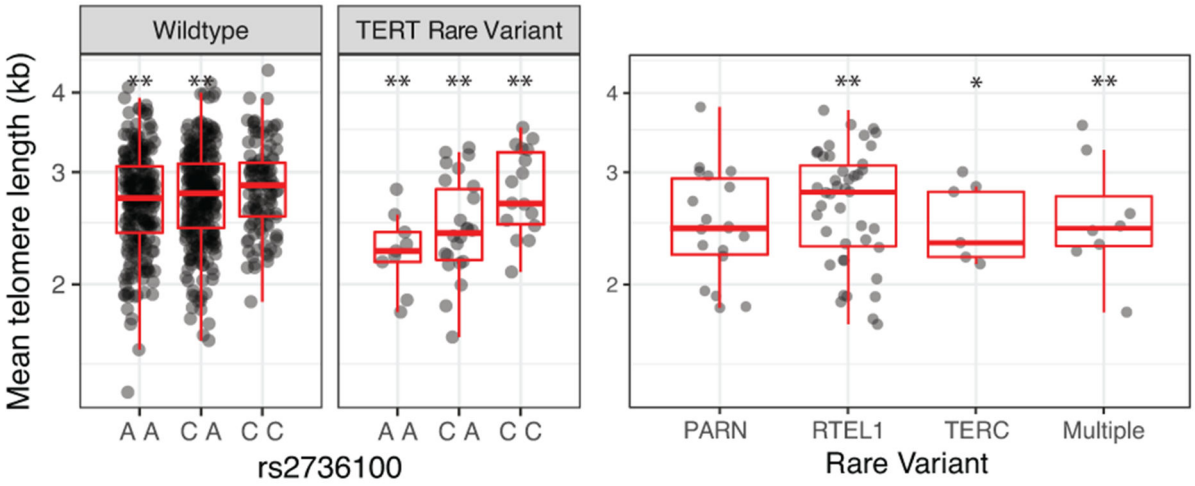

Figure 7.

Telomere lengths of clinical trials patients by *TERT* common variant (rs2736100) genotype and rare variant status. "A" is the risk allele for rs2736100, while "C" is the protective allele. "Wildtype" here refers to patients with a TERT coding sequence containing no missense mutations, while "TERT RV" indicates patients with at least one missense mutation. Boxes indicate the median and quartile values, while the whiskers extend to the most distant value no further than 1.5 times the interquartile range from the quartiles.

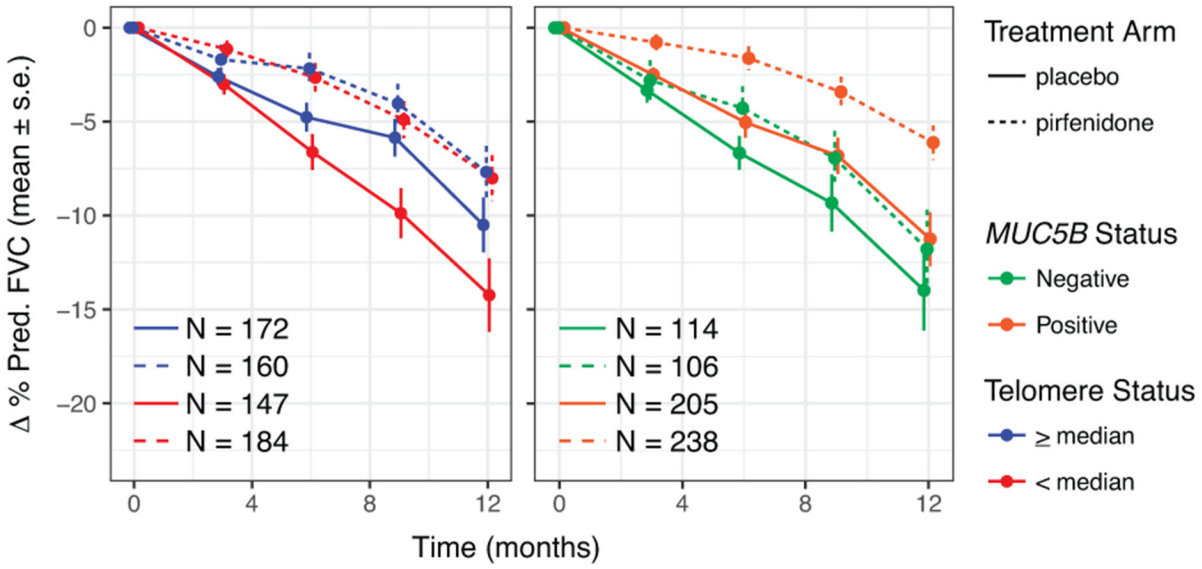

Figure 8.

Profile of mean percent decline over time of forced vital capacity (FVC) in the ASCEND and CAPACITY phase III clinical trials of pirfenidone for IPF. Patients in the ASCEND and CAPACITY trials were stratified by whether they received pirfenidone treatment (dashed lines) or a placebo control treatment (solid lines), and by whether they had peripheral blood telomeres that were longer (blue lines) or shorter (red lines) than the median length of the cohort. Error bars are Standard Error of the Mean. Figure A represents telomere length stratified by the median, Figure B is by *MUC5B* promoter variant status (orange denotes patients positive for the *MUC5B* risk-allele; green denotes patients negative for the *MUC5B* risk allele).

Table 1:

Population demographics - all

Characteristics	All IPF	MUC5B+ IPF	MUC5B− IPF	Controls	P1	P2
N	1510	1046	464	1874		
Age, mean (sd)	67·29 (7·98)	68·07 (7·66)	65·54 (8·39)	56·38 (9·32)	$< 2 \times 10^{-16}$	$1{\cdot}89 \times 10^{-8}$
Female, N(%)	391 (25·89%)	260 (24·86%)	131 (28·23%)	1367 (72·95%)	$< 2 \times 10^{-16}$	0·18
Ever smoked, N(%)	1000 (69·49%)	675 (68·81%)	325 (70·96%)	713 (38·05%)	$< 2 \times 10^{-16}$	0·44
rs35705950						
Counts, T·T/T·G/G·G	136/910/464	-	-	26/376/1472		
MAF	0·39	-	-	0·11		
P1 is the p-value for IPF vs non-IPF controls. P2 is the p-value for MUC5B+ IPF vs MUC5B− IPF.						

ASCEND						
Characteristics	**All IPF**	**MUC5B+ IPF**	**MUC5B− IPF**	**Controls**	**P1**	**P2**
N	271	200	71	1874		
Age, mean (sd)	68·34 (7·06)	69·06 (6·73)	66·3 (7·6)	56·38 (9·32)	$< 2 \times 10^{-16}$	0·005
Female, N(%)	67 (24·72%)	45 (22·5%)	22 (30·99%)	1367 (72·95%)	$< 2 \times 10^{-16}$	0·21
Ever smoked, N(%)	181 (66·79%)	131 (65·5%)	50 (70·42%)	713 (38·05%)	$< 2 \times 10^{-16}$	0·54
rs35705950						
Counts, T·T/T·G/G·G	23/177/71	-	-	26/376/1472		
MAF	0·41	-	-	0·11		
P1 is the p-value for IPF vs non-IPF controls. P2 is the p-value for MUC5B+ IPF vs MUC5B− IPF.						

CAPACITY						
Characteristics	**All IPF**	**MUC5B+ IPF**	**MUC5B− IPF**	**Controls**	**P1**	**P2**
N	312	216	96	1874		
Age, mean (sd)	66·56 (7·83)	67·47 (7·65)	64·5 (7·88)	56·38 (9·32)	$< 2 \times 10^{-16}$	0·002
Female, N(%)	94 (30·13%)	64 (29·63%)	30 (31·25%)	1367 (72·95%)	$< 2 \times 10^{-16}$	0·87
Ever smoked, N(%)	208 (66·67%)	143 (66·2%)	65 (67·71%)	713 (38·05%)	$< 2 \times 10^{-16}$	0·89
rs35705950						
Counts, T·T/T·G/G·G	33/183/96	-	-	26/376/1472		
MAF	0·40	-	-	0·11		
P1 is the p-value for IPF vs non-IPF controls. P2 is the p-value for MUC5B+ IPF vs MUC5B− IPF.						

INSPIRE						
Characteristics	**All IPF**	**MUC5B+ IPF**	**MUC5B− IPF**	**Controls**	**P1**	**P2**
N	340	239	101	1874		
Age, mean (sd)	66·49 (7·51)	66·97 (7·01)	65·35 (8·49)	56·38 (9·32)	$< 2 \times 10^{-16}$	0·06
Female, N(%)	97 (28·53%)	66 (27·62%)	31 (30·69%)	1367 (72·95%)	$< 2 \times 10^{-16}$	0·65
Ever smoked, N(%)	241 (70·88%)	169 (70·71%)	72 (71·29%)	713 (38·05%)	$< 2 \times 10^{-16}$	1·00

INSPIRE						
Characteristics	**All IPF**	**MUC5B+ IPF**	**MUC5B− IPF**	**Controls**	**P1**	**P2**
rs35705950						
Counts, T·T/T·G/G·G	21/218/101	-	-	26/376/1472		
MAF	0·38	-	-	0·11		
P1 is the p-value for IPF vs non-IPF controls. P2 is the p-value for MUC5B+ IPF vs MUC5B− IPF.						

UCSF						
Characteristics	**All IPF**	**MUC5B+ IPF**	**MUC5B− IPF**	**Controls**	**P1**	**P2**
N	66	66	0	1874		
Age, mean (sd)	73·73 (9·63)	74·3 (9·48)	-	56·38 (9·32)	$< 2 \times 10^{-16}$	-
Female, N(%)	13 (19·7%)	13 (19·7%)	-	1367 (72·95%)	$< 2 \times 10^{-16}$	-
Ever smoked, N(%)	46 (69·7%)	46 (69·7%)	-	713 (38·05%)	$1{\cdot}4 \times 10^{-16}$	-
rs35705950						
Counts, T·T/T·G/G·G	9/57/0	-	-	26/376/1472		
MAF	0·57	-	-	0·11		
P1 is the p-value for IPF vs non-IPF controls. P2 is the p-value for MUC5B+ IPF vs MUC5B− IPF.						

RIFF Cohort A						
Characteristics	**All IPF**	**MUC5B+ IPF**	**MUC5B− IPF**	**Controls**	**P1**	**P2**
N	81	54	27	1874		
Age, mean (sd)	69·93 (7·02)	69·41 (7·22)	70·96 (6·62)	56·38 (9·32)	$< 2 \times 10^{-16}$	0·34
Female, N(%)	15 (18·52%)	11 (20·37%)	4 (14·81%)	1367 (72·95%)	$< 2 \times 10^{-16}$	0·76
Ever smoked, N(%)	58 (71·6%)	37 (68·52%)	21 (77·78%)	713 (38·05%)	$2{\cdot}9 \times 10^{-9}$	0·54
rs35705950						
Counts, T·T/T·G/G·G	11/43/27	-	-	26/376/1472		
MAF	0·40	-	-	0·11		
P1 is the p-value for IPF vs non-IPF controls. P2 is the p-value for MUC5B+ IPF vs MUC5B− IPF.						

Vanderbilt						
Characteristics	**All IPF**	**MUC5B+ IPF**	**MUC5B− IPF**	**Controls**	**P1**	**P2**
N	440	274	166	1874		
Age, mean (sd)	66·34 (8·27)	67·08 (7·74)	65·13 (8·97)	56·38 (9·32)	$< 2 \times 10^{-16}$	0·01
Female, N(%)	105 (23·86%)	61 (22·26%)	44 (26·51%)	1367 (72·95%)	$< 2 \times 10^{-16}$	0·35
Ever smoked, N(%)	312 (71·72%)	195 (71·69%)	117 (71·78%)	713 (38·05%)	$< 2 \times 10^{-16}$	1·00
rs35705950						
Counts. T·T/T·G/G·G	39/235/166	-	-	26/376/1472		
MAF	0·36	-	-	0·11		
P1 is the p-value for IPF vs non-IPF controls. P2 is the p-value for MUC5B+ IPF vs MUC5B− IPF.						

Table 2:

TERT rare variant frequency stratified by the *MUC5B* variant

Gene	MUC5B+ IPF Freq	MUC5B− IPF Freq	N SNPs	OR (95% CI)	P
TERT	2·8%	7·3%	39	0·40 [0·24, 0·66]	$3{\cdot}9 \times 10^{-4}$

Table 3:

Common GWAS SNPs from Fingerlin et al. stratified by the MUC5B promoter variant

SNP	Chr	Position	Nearby genes	Function	Risk Allele	MUC5B+ IPF MAF	MUC5B− IPF MAF	OR [95% CI]	P
rs6793295	3	169800607	*LRRC34*	Missense	C	30%	32%	0·94 [0·79, 1·11]	0·45
rs2609255	4	88890044	*FAM13A*	Intronic	G	26%	31%	0·77 [0·65, 0·91]	$2{\cdot}5 \times 10^{-3}$
rs2736100	5	1286401	*TERT*	Intronic	C	42%	42%	1·01 [0·86, 1·18]	0·95
rs2076295	6	7562999	*DSP*	Intronic	G	56%	55%	1·00 [0·86, 1·17]	0·96
rs4727443	7	99995723	*AZGP1P1-ZKSCAN1*	Intergenic	A	46%	45%	1·06 [0·91, 1·24]	0·45
rs11191865	10	103913084	*OBFC1*	Intronic	A	56%	56%	0·98 [0·84, 1·16]	0·85
rs1278769	13	112882313	*ATP11A*	3 UTR	A	19%	20%	0·95 [0·78, 1·15]	0·59
rs2034650	15	40425103	*IVD-BAHD1*	Intergenic	G	43%	42%	1·02 [0·87, 1·19]	0·80
rs1981997	17	45979401	*MAPT*	Intronic	N/A	N/A	N/A	N/A	N/A
rs12610495	19	4717660	*DPP9*	Intronic	G	35%	35%	0·98 [0·84, 1·16]	0·85
rs62025270	15	85756967	*AKAP13*	Intronic	A	25%	24%	1·01 [0·83, 1·21]	0·94

Table 4a:

Rare variant burden test for telomerase complex components

Gene	MUC5B+ IPF Freq	MUC5B− IPF Freq	N SNPs	OR [95% CI]	P	All IPF Freq	Control Freq	N SNPs	OR [95% CI]	P
ACD (TPP1)	0·38%	0·65%	4	0·64 [0·14, 3·37]	0·57	0·46%	0·53%	9	1·73 [0·46, 6·26]	0·41
CTC1	3·73%	2·37%	17	1·53 [0·79, 3·19]	0·23	3·31%	3·42%	30	0·85 [0·5, 1·42]	0·53
DKC1	N/A	N/A	0	N/A	N/A	N/A	N/A	0	N/A	N/A
GAR1	0·19%	0·00%	2	N/A	0·89	0·13%	0·21%	3	0·57 [0·06, 3·87]	0·57
NHP2	1·24%	0·65%	1	1·87 [0·59, 8·27]	0·34	1·06%	0·75%	5	1·75 [0·62, 5·09]	0·30
NOP10	N/A	N/A	0	N/A	N/A	0·00%	0·05%	2	N/A	0·95
PARN	0·90%	1·90%	17	0·45 [0·17, 1·16]	0·09	1·24%	0·10%	19	8·74 [2·16, 59·65]	0·007
POT1	0·86%	0·86%	2	1·11 [0·35, 4·15]	0·87	0·86%	0·80%	4	0·67 [0·27, 1·65]	0·38
RTEL1	2·77%	3·66%	31	0·8 [0·43, 1·53]	0·49	3·05%	1·76%	47	2·88 [1·59, 5·29]	$5{\cdot}44 \times 10^{-4}$
STN1	N/A	N/A	0	N/A	N/A	N/A	N/A	0	N/A	N/A
TCAB1	N/A	N/A	0	N/A	N/A	N/A	N/A	0	N/A	N/A
TEN1	0·00%	0·43%	2	N/A	0·87	0·13%	0·11%	4	1·57 [0·14, 16·95]	0·70
TERC (TR)	0·76%	0·86%	22	0·81 [0·26, 2·73]	0·71	0·79%	0·21%	22	4·27 [1·30, 16·64]	0·04
TERF1 (TRF1)	0·10%	0·00%	1	N/A	0·88	0·07%	0·43%	4	0·2 [0·01, 1·58]	0·19
TERF2 (TRF2)	0·67%	0·86%	3	0·73 [0·22, 2·85]	0·62	0·73%	0·59%	6	1·28 [0·44, 3·72]	0·65
TEKF2IP (RAP1)	0·10%	0·00%	1	N/A	0·90	0·07%	0·11%	3	0·37 [0, 24·0]	0·70
TERT	2·87%	7·33%	39	0·4 [0·24, 0·66]	$3{\cdot}9 \times 10^{-4}$	4·24%	1·71%	41	2·82 [1·64, 4·97]	$2{\cdot}38 \times 10^{-4}$
TINF2 (TIN2)	0·10%	0·22%	2	0·4 [0·02, 10·21]	0·52	0·13%	0·21%	6	1·00 [0·1, 7·53]	1·00
TZAP (ZBTB48)	N/A	N/A	0	N/A	N/A	N/A	N/A	0	N/A	N/A
WRAP53	0·96%	0·22%	7	4·98 [0·92, 92·48]	0·13	0·73%	0·75%	13	1·03 [0·35, 2·94]	0·96

Table 4b:

Rare variant burden tests for previously implicated IPF susceptibility loci

Gene	MUC5B+ IPF Freq	MUC5B− IPF Freq	N SNPs	OR [95% CI]	P	All IPF Freq	Control Freq	N SNPs	OR [95% CI]	P
AKAP13	0·002	0·000	2	N/A	0·89	0·001	0·003	7	0·37 [0·04, 2·55]	0·32
ATP11A	0·015	0·011	14	1·27 [0·49, 3·94]	0·65	0·014	0·014	26	0·95 [0·46, 1·96]	0·90
AZGP1P1	N/A	N/A	0	N/A	N/A	N/A	N/A	0	N/A	N/A
BAHD1	0·007	0·002	9	2·46 [0·43, 46·21]	0·40	0·005	0·004	14	2·48 [0·57, 11·29]	0·23
DPP9	0·016	0·011	4	1–54 [0·59, 4·78]	0·41	0·015	0·007	8	1–53 [0·65, 3–72]	0·34
DSP	0·023	0·026	20	0·83 [0·41, 1·76]	0·61	0·024	0·035	47	0·79 [0·46, 1·33]	0·37
FAM13A	0·003	0·009	6	0–26 [0·05, 1·19]	0·08	0·005	0·004	14	1·13 (0·27, 4·79]	0·87
IVD	0·011	0·013	7	0–72 [0·26, 2·52]	0·54	0·011	0·006	7	1–22 [0·36, 3·13]	0·70
LRRC34	0·005	0·009	8	0·53 [0·14, 2·18]	0·35	0·006	0·004	10	2·00 [0·58, 6·95]	0·27
MAPT	N/A	N/A	0	N/A	N/A	N/A	N/A	0	N/A	N/A
MUC5B	0·038	0·052	28	0·69 [0·41, 1·19]	0·18	0·065	0·066	63	0·83 [0·57, 1·2]	0·30
OBFC1	0·001	0·000	1	N/A	0·89	0·001	0·001	2	0–67 [0·02, 28·59]	0·83
PARN	0·009	0·019	17	0·45 [0·17, 1·16]	0·09	0·012	0·001	19	8·74 [2·16, 59·65]	0·007
RTEL1	0·028	0·037	31	0·8 [0·43, 1·53]	0·49	0·030	0·018	47	2·88 [1·59, 5·29]	$5{\cdot}44 \times {}^{-4}$
SFTPA1	0·003	0·000	1	N/A	0·91	0·002	0·001	1	2·26 [0·29, 21·9]	0·44
SFTPA2	0·000	0·004	1	N/A	0·88	0·001	0·001	1	0·42 [0·04, 4·75]	0·45
SFTPC	0·004	0·002	3	3·11 [0·43, 62·95]	0·32	0·003	0·004	4	1·71 [0·44, 6·31]	0·43
TERC	0·008	0·009	22	0·81 [0·26, 2·73]	0·71	0·008	0·002	22	4·27 [1·3, 16·64]	0·042
TERT	0·029	0·073	39	0·4 [0·24, 0·66]	$3{\cdot}9 \times 10^{-4}$	0·042	0·017	41	2·82 [1·64, 4·97]	$2{\cdot}38 \times 10^{-4}$
TOLLIP	0·004	0·000	5	N/A	0·90	0·003	0·004	11	1·14 [0·21, 5·61]	0·87
ZKSCAN1	0·013	0·011	2	1·22 [0·46, 3·83]	0·71	0·013	0·022	4	0·67 [0·32, 1·36]	0·28

Table 5:

Patient telomere length by rare variant status in IPF and non-IPF

IPF Cases						
Gene	**SNP type**	**Avg telomere length**	**N**	**Δ from reference**	**change from reference (%)**	**p-value**
TERT	rare	2·50	87	0·26	9·55	$1{\cdot}54 \times 10^{-5}$
TERC	rare	2·32	15	0·45	16·10	0·03
PARN	rare	2·66	18	0·10	3·69	0·22
RTEL1	rare	2·57	88	0·20	7·16	$3{\cdot}16 \times 10^{-4}$
2+ RV	rare	2·55	6	0·21	7·74	0·01
TERT	common - AA	2·65	440	0·11	4·02	$1{\cdot}38 \times 10^{-3}$
TERT	common - AC	2·66	626	0·10	3·69	$1{\cdot}5 \times 10^{-3}$
TERT	common - CC	2·76	248	reference	reference	reference
All Patients						
TERT	rare	2·93	125	0·29	9·01	$4{\cdot}55 \times 10^{-12}$
TERC	rare	2·30	19	0·92	28·57	$3{\cdot}56 \times 10^{-4}$
PARN	rare	2·57	19	0·65	20·19	$1{\cdot}49 \times 10^{-4}$
RTEL1	rare	2·57	121	0·65	20·19	$2{\cdot}9 \times 10^{-27}$
2+ RV	rare	2·55	6	0·67	20·81	$3{\cdot}66 \times 10^{-6}$
TERT	common - AA	3·12	2806	0·10	3·11	$2{\cdot}6 \times 10^{-7}$
TERT	common - AC	3·14	4888	0·08	2·48	$1{\cdot}01 \times 10^{-6}$
TERT	common - CC	3·22	2292	reference	reference	reference