IDIOPATHIC PULMONARY FIBROSIS: A GENETIC DISEASE THAT INVOLVES MUCOCILIARY DYSFUNCTION OF THE PERIPHERAL AIRWAYS

Christopher M. Evans, Tasha E. Fingerlin, Marvin I. Schwarz, David Lynch, Jonathan Kurche, Laura Warg, Ivana V. Yang, and David A. Schwartz

Department of Medicine, University of Colorado Denver, School of Medicine, Aurora, Colorado; National Jewish Health, Denver, Colorado; and Department of Immunology, University of Colorado Denver, School of Medicine, Aurora, Colorado

Idiopathic pulmonary fibrosis (IPF) affects 5 million worldwide, disproportionately affects men, is more prevalent with age, and is inexplicably increasing in prevalence (75, 142, 151). Nevertheless, IPF is likely underdiagnosed (72, 151, 165). Death occurs in up to 50% of affected individuals within 3–5 yr following the diagnosis (75, 92, 165). Recently pirfenidone (94) and nintedanib (170) have been shown to slow IPF progression; however, IPF remains a progressive disease and no treatment (short of lung transplantation) is known to impact the poor survival associated with this disease (75). While IPF takes years to develop, most patients with IPF are diagnosed in the advanced stage of disease when little can be done to influence survival. Earlier diagnosis of IPF will identify subjects with more salvageable lung, enable treatment of those with a lower burden of lung disease (121, 163), and may reveal novel molecular targets for intervention that are effective in the preclinical and/or early clinical stages of this progressive disease.

The increasing prevalence of IPF may be explained to some extent by our aging population (75, 151), a greater physician awareness of the disease, screening of first-degree family members affected by the disease, and earlier detection in asymptomatic individuals in which case disease is found incidentally after high-resolution computed tomographic scans (HRCT) for other suspected conditions; however, the observed increase in prevalence is not fully understood. In the United States, IPF has an estimated incidence of 20–30/100,000 individuals, and this number increases with each decade between 50 and 90 yr of age (151). For instance, in 70- to 79-yr-old individuals, the incidence has been reported as high as 75/100,000 (36, 64). However, several studies imply a much higher prevalence of IPF than previously thought even if all asymptomatic fibrotic cases did not progress to established IPF. In a chest HRCT study of al-
most 2,500 smokers undergoing screening for lung cancer, bilateral interstitial lung abnormalities were found in 8%, and these abnormalities were associated with reduced lung volumes (225). Similarly, chest CT abnormalities consistent with IPF were found in 1.8% of those older than 50 yr of age (72). Among asymptomatic members of families with two or more cases of pulmonary fibrosis [familial interstitial pneumonia (FIP)], 14% have interstitial abnormalities on CT scan and 35% have interstitial abnormalities on transbronchial biopsy (111). We have found that the MUC5B promoter variant rs35705950 that is consistently and strongly associated with IPF (13, 47, 66, 146, 156, 178, 191, 226, 236) is also predictive of those with preclinical forms of pulmonary fibrosis (OR = 6.3 per allele, 95% CI 3.1–12.7), suggesting that IPF risk variants may be helpful in identifying subjects with a higher risk of disease development.

While a number of exposures [metal and wood dust (7, 67, 68, 81, 136), viruses (118, 190, 200, 233), and drugs (43, 69, 139)] have been found to be associated with the development of IPF, the most important environmental risk factor for IPF is cigarette smoking. In addition to occupational exposures, such as asbestos (15, 124, 173), ever having smoked cigarettes remains a risk factor for the development of IPF even after smoking cessation (8), suggesting that the response induced by cigarette smoke-related lung injury may become self-sustaining (187). For FIP, we have shown that cigarette smoking is a strong risk factor for the development of disease (189; OR = 3.6; 95% CI = 1.3–9.8). This finding suggests that cigarette smoking contributes significantly even in a genetic model of pulmonary fibrosis.

The morphological hallmark of IPF on imaging and histology is usual interstitial pneumonia (UIP). On the chest radiograph, UIP is characterized by basal predominant fine reticulonodular abnormality, which may be present for many years before clinical presentation with IPF. However, these features are relatively nonspecific (129). HRCT scans with appropriate technique is critical for evaluation of IPF (130). The characteristic HRCT features are reticular abnormality, traction bronchiectasis and honeycombing, with basal and peripheral predominant distribution, and with absence of features (such as mosaicism attenuation or centrilobular nodules) that would suggest an alternative diagnosis (166). A confident radiologic diagnosis of UIP, by an experienced radiologist, based on these clinical features, correctly predicts the histological diagnosis of UIP in over 90% of cases (41, 50, 73, 74, 129, 183). Radiologic honeycombing is the most important single predictor of a UIP diagnosis (73). If honeycombing is absent, but all other HRCT features are highly suggestive of UIP, the likelihood of UIP is more than 80% (29). Thus the radiologic diagnosis of UIP may sometimes be accepted even in the absence of honeycombing, particularly in older individuals in whom the pretest probability of IPF is already high (46). Importantly, the HRCT features of UIP secondary to collagen vascular disease or other causes, such as asbestosis or chronic hypersensitivity pneumonitis, are indistinguishable from those of IPF (198). Identification of an underlying cause for UIP therefore remains an important task for the pulmonary clinician.

The histological expression of IPF is UIP. IPF, either familial or sporadic, represents a lung-limited condition recognized by progressive scarring of the pulmonary interstitium that results from persistent collagen production by proliferating fibroblasts/myofibroblasts and regenerating alveolar epithelia. This represents a constellation of microscopic findings including fibrosis of the alveolar interstitium alternating with normal appearing lung, fibroblastic foci, and subpleural microscopic honeycombing (164, 208). Fibroblastic foci are distinct subepithelial interstitial collections of myofibroblasts that are capable of producing mucopolysaccharide and collagen. Fibroblastic foci are not necessarilyfuse and are more likely to be found in the fibrotic interstitium adjacent to denser areas of fibrosis and in the subpleural regions of the lung (34). It has been put forth that fibroblastic foci represent the site of the initial response to a yet undefined injury to type 1 alveolar epithelial cells, and due to the likely temporally heterogeneous nature of the histological findings in UIP, the alveolar epithelial injury is further postulated to be recurrent over time leading to the development of numerous foci of injury and repair resulting in the progressive accumulation of extracellular matrix (179). Fibroblastic foci are further hypothesized to be the starting point of the uncontrolled fibroproliferative process that eventually encroaches on the normal lung; however, there is inadequate support for this assumption.

An additional and unique histological feature of histological UIP is microscopic honeycombing. This is distinct from radiologic honeycombing and is characterized by clusters of small cysts, usually in a subpleural location and in densely fibrotic areas. Honeycomb cysts are often filled with mucus and occasionally inflammatory cells. These microscopic cystic structures are often lined by both pseudostratified ciliated columnar epithelium similar to that which line the major airways and distal bronchioles (199) and alveolar epithelium consisting of type 2 alveolar epithelial cells. In some cases, honeycomb cysts are entirely lined by alveolar type 2 cells. It has been long thought that honeycomb cysts represent the end stage of IPF since these structures are found in densely fibrotic areas of IPF/UIP; however, the origin and implications of these cystic structures remain debatable. Alternatively, the cysts could be the result of aberrant attempts at regeneration of the terminal airways, possibly via activation of dormant developmental pathways (181) (FIGURE 1).

An additional prominent histological feature, though not specific to UIP, is type 2 alveolar epithelial cell hyperplasia.
These cells function as the progenitors for alveolar type 1 cells, which under normal conditions represent 90% of the alveolar epithelial surface area (181). Additionally, type II alveolar epithelial cells are the source of pulmonary surfactant. Following an injury to the alveolus, reconstitution of alveolar structure by type II cell proliferation and differentiation to type I cells is considered to be essential for maintenance of the epithelial barrier and restoration of gas exchange. In IPF, however, type 2 alveolar epithelial cell hyperplasia persists replacing type 1 cells and likely contributes to collagen accumulation both by the release of fibroblast growth factors, and by transitioning to fibroblasts and eventually myofibroblasts (epithelial-mesenchymal transition) (6, 180, 229). Moreover, while the alveolar epithelial cells undergo apoptosis, the fibroblasts/myofibroblasts appear to be resistant to this process.

Over the past decade, consensus has been reached on the diagnosis of IPF (92, 165, 207, 208). It depends on clinical features (progressive and persistent dyspnea, disease localized to the lungs, and no clear etiology), an HRCT of the lung showing lower zone predominant reticular infiltrates, traction bronchiectasis, and subpleural honeycomb changes, or a surgical lung biopsy demonstrating UIP. It should be noted that microscopic honeycombing does not correlate with HRCT scan honeycombing which may represent dilated airways due to traction from the fibrotic process. Histological honeycombing is almost always present to some degree in UIP; however, this is not the case for HRCT scan honeycombing. If typical HRCT findings are seen, coupled with the expected clinical presentation, and there is no history of connective tissue disease (CTD) and serum autoantibodies are absent, the diagnosis of IPF is fairly secure. However, in up to 33% of IPF subjects when first evaluated, HRCT honeycombing is absent. Instead, the other features of IPF including bilateral peripheral subpleural lower zone predominant reticular infiltrates and traction bronchiectasis are seen. Even in this instance, surgical lung biopsy more often reveals UIP with microscopic honeycombing than NSIP.

Other fibrotic lung diseases have clinical and histological expressions that must be distinguished from IPF. IPF falls into a classification of interstitial lung diseases based on their histological appearance and known as the idiopathic interstitial pneumonias (IIP) (90, 93, 165, 207, 208). IPF/UIP is most common followed by the histological entity of nonspecific interstitial pneumonia (NSIP). NSIP can be idiopathic (iNSIP), but more often represents either a pulmonary manifestation of one of the known CTD or possibly a lung dominant expression of a CTD (LD-CTD). In LD-CTD patients, circulating autoantibodies are present, but clinical features necessary to establish a CTD diagnosis are lacking (48, 176). NSIP can also be one of the histological expressions of chronic hypersensitivity pneumonitis (HP) or a drug-induced lung disease (217). Compared with IPF, NSIP occurring with iNSIP, CTD, or LD-CTD occurs in younger subjects, is more often seen in women and has a much more benign prognosis (210). In pulmonary fibrosis, whether idiopathic, familial, due to chronic hypersensitivity pneumonitis, CVD, LD-CVD, or even asbestosis, either NSIP or UIP may be present.

While IPF is a distinct entity, there are significant variations in the clinical course. The expected course is progressive dyspnea with or without cough with death occurring within 3–5 yr from the consequences of hypoxemia. However, in up to 20% of IPF subjects there is prolonged survival (6–12 yr). In ~5–10% of patients with IPF, acute decompensations occur, referred to as acute exacerbations (33, 186). Pathologically, acute exacerbations are characterized by the histological lesion of organizing diffuse alveolar damage (DAD), which is superimposed on the preexisting UIP. Patients with acute exacerbations experience a marked increase in their dyspnea over 3–7 days, become more hypoxic, and are increasingly difficult to oxygenate (123). Imaging of the chest shows new bilateral diffuse ground glass opacities, and over 50% of episodes of acute exacerbations are fatal in spite of high-dose glucocorticoids and broad-spectrum antibiotics. For those who survive, mortality within the next year is high. There is also an accelerated course in IPF with declines in lung function and death within 3 yr after diagnosis, but without acute exacerbations. In this group, during the initial evaluation, there is often more severe physiological impairment, i.e., forced vital capacity and total lung capacity <60% of predicted, diffusing capacity for carbon monoxide (DLCO) <35% of predicted, hypoxemia requiring supplemental oxygen and
sometimes the presence of pulmonary hypertension. An accelerated course can be anticipated if at a 6-mo interval follow-up, physiological testing demonstrates 5–10% declines in the percentage of the physiological predicted values compared with the initial values (116). Other than the genotype (146, 147, 158), there are few, if any, validated biomarkers which predict outcome, particularly when the patient is first evaluated. For example, it has been shown that IPF subjects with the MUC5B promoter variant rs35705950 have improved survival compared with those without this variant (158) (Figure 2), suggesting that the MUC5B variant either identifies individuals earlier in the course of disease or represents a different pathophysiological subtype of IPF. While a number of biological markers have been shown to be prognostic for existing disease (21, 32, 39, 49, 52, 53, 56, 61, 71, 80, 87, 97, 108, 109, 122, 134, 150, 153, 158, 161, 169, 174, 195, 215, 219, 227, 235, 239), with the exception of the MUC5B promoter variant (72) and MMP7 (174), it is unclear if these biomarkers are useful for predicting preclinical/mild disease.

In the past decade recent novel treatment trials for IPF have focused on inhibition of the fibroproliferative process with the primary end point being improvement or stabilization of the FVC percent predicted (94, 170). Secondary end points included quality of life and/or mortality. Multiple pathways that enhance fibrosis have been interrogated by novel therapies in human IPF and have been negative except for two recent trials testing pirfenidone (94) and nintedanib (170), which have both been shown to slow IPF progression. Pirfenidone, thought to interfere with the profibrotic molecule transforming growth factor (TGF)-β1 (94), and nintedanib, a triple-kinase inhibitor of fibroblast growth factors (170), have been approved for use in IPF. Although in neither trial was there mortality or quality of life benefits, there were impressive decreases in the rate of the FVC decline after 1 yr of treatment. While both pirfenidone and nintedanib slow the rate of disease progression, the long-term effects of these drugs on mortality are unknown; however, review of the trials suggest that pirfenidone and nintedanib may prove to reduce mortality (89).

Bilateral lung transplant remains an option for qualified IPF patients offering a 50% 3-yr survival post-transplant (143). However, a relatively small number of cases qualify or receive this potential life-saving procedure. Other adjunct treatments that may contribute to improve the outcome of IPF include use of oxygen to manage hypoxia, possibly sildenafil for pulmonary hypertension, nocturnal continuous positive airway pressure for obstructive sleep apnea, and the treatment of gastric acid reflux aspiration (31, 76, 132).

II. GENETIC DETERMINANTS OF PULMONARY FIBROSIS

A. Risk of Pulmonary Fibrosis

Current evidence suggests that IPF results from a combination of multiple genetic and nongenetic risk factors, where
the definition of nongenetic risk factors is broadly construed to include both exogenous (e.g., cigarette smoking and air pollution) and endogenous (e.g., systemic diseases and autoimmune conditions) elements that contribute to the overall risk of developing disease. While many nongenetic factors may also be under genetic control (e.g., autoimmune conditions), the concept is that complex diseases, which constitute most human illnesses, develop consequent to the combination of genetic factors and factors that are primarily nongenetic. The relative contribution of genetic versus nongenetic contributors differs across diseases and populations; within a disease like IPF, the relative contributions of each of these risks also differ among individuals. This complexity results in challenges for identification of genetic risk factors, and several different approaches have been used in combination to overcome these challenges for IPF.

Linkage analysis encompasses a group of statistical methods to examine the inheritance pattern of genetic variants within families to determine if there is a relationship between the genetic variants and the phenotype of interest. Linkage analyses rely on family data, which can be made up by a wide range of pedigree structures from extended pedigrees to affected sib pairs; the more distant the relationships between affected individuals, the more powerful the analysis. Linkage analysis is most powerful for rare, highly penetrant alleles, which likely account for a small proportion of pulmonary fibrosis.

Genetic association studies are the most commonly used study design to find disease genes in complex traits. Association studies can be based on a candidate region based on linkage analysis, for candidate genes identified in other ways, or in genome-wide association studies (GWAS). GWAS have historically been conducted using common variants, but are increasingly being explored in the context of whole exome or whole genome resequencing. The challenge for any GWAS is obtaining convincing statistical evidence for association with any given variant or group of variants in the face of many statistical tests. Moreover, association studies often result in common variants that are not particularly penetrant. Each of the above strategies has been applied to pulmonary fibrosis; here we give a brief overview of the evidence for pulmonary fibrosis genetic risk factors.

Clustering of pulmonary fibrosis in families first suggested a genetic component to disease risk, and several studies have identified both common and rare genetic variations that are linked and/or associated with disease (203). Cases of pulmonary fibrosis have been reported in closely related family members including monozygotic twins raised in different environments (12, 84, 185), in genetically related members of several families (10, 12, 70, 193), in consecutive generations in the same families (12, 65, 120), and in family members separated at an early age (193). While a single report suggests that familial interstitial pneumonia (FIP) is inherited as an autosomal recessive trait (212), other pedigrees demonstrate an autosomal dominant pattern of inheritance (1, 131, 193) perhaps with reduced penetrance (1, 10, 70, 84, 126, 140, 185, 193). We have reported over 100 families with 2 or more cases of IIP (a category of interstitial lung disease inclusive of IPF) and have found that FIP most closely follows an autosomal dominant inheritance pattern that involves all subtypes of IIP, is strongly associated with cigarette smoking (OR = 3.6; 95% confidence interval 1.3–9.8), and appears to be a complex genetic disorder (189), suggesting that the different histological types of idiopathic interstitial pneumonia may be related etiologically or even pathogenically. These findings highlight the complex nature of IPF and suggest that while a susceptibility gene may predispose to IIP, another event (a modifier gene, a medical condition, or a specific exposure) may result in a unique type of idiopathic interstitial pneumonia. Rare variants in several genes have been reported to segregate with familial interstitial fibrosis, including TERT (4, 211), TERC (4, 211), RTEL (30, 192), and PARN (192) in the telomerase pathway and surfactant protein genes SFTPC (119, 205, 213) and SFTPA2 (224). These studies have taken advantage of large families where available, but have also identified rare variants among those with no family history of disease. Reviewed in detail elsewhere (110), studies of these genes and specific variants implicate telomere maintenance (TERT, TERC, PARN and RTEL) and surfactant dysfunction (SFTPC, SFTPA2) in IPF pathogenesis.

Two large GWAS of patients (familial and sporadic) and controls have been conducted in pulmonary fibrosis (47, 146). Both studies confirmed the role of the MUC5B gene in the risk of IPF (see below) and identified other common variants associated with risk. Noth et al. (146) reported association between pulmonary fibrosis and three polymorphisms in the TOLLIP gene, which is in the same region as the MUC5B gene. Our results demonstrate that common variants in both the TERT and TERC gene regions were associated with IPF (47), suggesting that both common and rare variants in the telomerase genes may be risk factors for disease (FIGURE 3). We also identified seven other loci associated with disease (4q22, FAM13A; 6p24, DSP; 10q24, OBFC1; 13q34, ATP11A; 19p13, DPP9; 7q22, 15q14-15). The pathways implicated by the genes identified in the GWAS confirmed the role of telomere maintenance (TERT, TERC, OBFC1) and host defense (MUC5B, ATP11A, TOLLIP), and identified barrier function (DSP, DPP9) as important for disease etiology. In aggregate, the genome-wide variants account for 30–35% of the risk of familial and sporadic pulmonary fibrosis, suggesting an important role for common genetic variation in both familial and sporadic disease. In fact, the risk of developing IPF for each of the significant genetic loci reported is equivalent for familial and sporadic forms of IPF (FIGURE 4), suggesting a...
similarity between these two types of IPF. This is important in several contexts, including efforts related to screening high-risk populations and to target therapeutics toward these common features that could potentially be applicable to broad groups of patients.

The biological functions of the genes implicated so far suggest that increased injury may be a primary defect for a substantial proportion of those with pulmonary fibrosis, rather than, or in addition to, dysregulated repair mechanisms. The secondary nature of the chronic inflammation and remodeling may provide an additional explanation why many treatments aimed at decreasing inflammation have failed as therapeutics. The genetic findings are key for eventual identification of the causal variant(s) in each gene/region. As we gain a more comprehensive understanding of the mechanism of disease pathogenesis, the hope is that new therapeutics can be developed. As treatments that prolong life are identified, these genetic variants and other biomarkers may prove useful in personalizing approaches to treatment and identifying individuals who are at risk or at earlier stages of disease when the lung is more salvageable.

**B. MUC5B Promoter Variant rs35705950**

We discovered the *MUC5B* promoter variant rs35705950 as a risk variant for IPF using a positional cloning approach of linkage analysis in families followed by case-control association across the region of linkage (178). The linkage analysis was carried out among 88 families of European ancestry and identified a region of suggestive linkage on chromosome 11p15. Resequencing of the region in individuals of European ancestry identified polymorphisms that were then genotyped in familial and sporadic cases in addition to controls of European ancestry. The rs35705950 variant was the strongest and only independently associated variant in the region and has now been replicated in seven subsequent studies among individuals of European ancestry (13, 47, 66, 146, 191, 226, 236). TABLE 1 provides the risk allele carrier frequency, odds ratios, and confidence intervals for having one copy of the risk allele (54.9–67%) compared with having no copies for each of the published studies to date. The frequency of the risk allele (mean allele
frequency) is ~35% among European ancestry cases compared with ~9% among European ancestry controls. The largest study (47) estimated that the odds of developing pulmonary fibrosis for those with one copy of the risk allele were 4.5 times (95% CI: 3.9-5.2) the odds of those with no copies and that the odds for those with two copies are 20.2 times those with no copies (95% CI: 15.2-27.0). Strikingly, the risk associated with rs35705950 is very similar between familial and sporadic disease (178), indicating that the variant contributes substantially to both presentations of IPF.

To date, the MUC5B promoter variant rs35705950 is the strongest known risk factor (genetic and otherwise) for the development of IPF. And while we and others have examined the association between rs35705950 and a variety of lung diseases [asbestosis, sarcoidosis, scleroderma associated interstitial lung disease, chronic obstructive pulmonary disease (COPD), asthma, and lung cancer] (13, 157, 191), the MUC5B promoter variant is only associated with IPF (13, 47, 66, 146, 178, 191, 226, 236) and idiopathic nonspecific interstitial pneumonia (66).

While initial studies of rs35705950 are among those of European ancestry, recent studies have examined the evidence for association with pulmonary fibrosis in other race or ethnic groups (66, 156, 223). The frequency of the T allele at rs35705950 is 11, 8, and 1% among European, South Asian, and East Asian populations, respectively, and was observed only once among 504 individuals (1,008 alleles) from Africa (1,000 genomes) (54). Even though rs35705950 is rare among Asian and African ancestries, the evidence so far indicates that the risk associated with the variant among Asians (66, 223) and Mexicans (156) is similar to the risk observed among European ancestries (TABLE 2). This similarity is important as the MUC5B promoter variant rs35705950 is considered a target for early diagnosis and intervention (72), and rs35705950 is considered for prognostic potential (158).

The elevated risk associated with rs35705950 is substantially higher than for most other common risk variants for complex disease, with the exception of the human leukocyte antigen (HLA) region for some autoimmune diseases such as type I diabetes and systemic lupus erythematosus (SLE), which have odds ratios greater than 10 (42) and 4 (167), respectively, for specific HLA alleles. Other mucin polymorphisms that are associated with gastric and colon cancer have much lower risk estimates than those estimated for rs35705950. For example, a meta-analysis of studies of rs4072037 in the MUC1 gene estimate the OR for the rare allele to be 0.58 (95% CI: 0.44, 0.75) among European ancestries and 0.60 (95% CI: 0.54, 0.66) among Asian ancestries under a dominant model for the rare allele (40), corresponding to an OR of ~1.7 for the common allele.

### Table 1. Summary of studies conducted in NHWs investigating the relationship between the MUC5B promoter variant rs35705950 and IPF

<table>
<thead>
<tr>
<th>Location</th>
<th>IPF (%SNP)</th>
<th>Controls (%SNP)</th>
<th>Odds Ratio GT Versus GG</th>
<th>Odds Ratio TT Versus GG</th>
<th>P Value</th>
<th>Reference Nos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>367 (60.8%)</td>
<td>802 (20.7%)</td>
<td>5.7 (4.3-7.5)</td>
<td>9.6 (4.7-19.4)</td>
<td>8.9 × 10⁻¹¹</td>
<td>236</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>110 (67.0%)</td>
<td>416 (19.6%)</td>
<td>6.6 (4.1-10.7)</td>
<td>11.8 (4.3-33.7)</td>
<td>2.0 × 10⁻¹⁷</td>
<td>191</td>
</tr>
<tr>
<td>France</td>
<td>142 (65.5%)</td>
<td>1383 (20.2%)</td>
<td>6.4 (4.5-9.0)</td>
<td>19.1 (9.0-36.1)</td>
<td>9.0 × 10⁻²⁹</td>
<td>13</td>
</tr>
<tr>
<td>United States</td>
<td>2,492 (63.1%)</td>
<td>1890 (19.6%)</td>
<td>4.5 (3.9-5.2)</td>
<td>20.2 (15.2-27.0)</td>
<td>7.2 × 10⁻⁹⁵</td>
<td>47</td>
</tr>
<tr>
<td>United States</td>
<td>324 (57.2%)</td>
<td>702 (21.3%)</td>
<td>2.4 (2.1-2.8) overall</td>
<td>2.4 × 10⁻¹⁰</td>
<td>146</td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>84 (55.9%)</td>
<td>689 (21.8%)</td>
<td>3.2 (2.2-4.6) overall</td>
<td>1.2 × 10⁻¹⁰</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>71 (54.8%)</td>
<td>35 (8.6%)</td>
<td>11.1 (3.3-37.0) overall</td>
<td>&lt;0.001</td>
<td>66</td>
<td></td>
</tr>
</tbody>
</table>

IPF, idiopathic pulmonary fibrosis; SNP, single nucleotide promoter.

### Table 2. Summary of studies conducted in Asians and Hispanics investigating the relationship between the MUC5B promoter variant rs35705950 and IPF

<table>
<thead>
<tr>
<th>Location</th>
<th>IPF (%SNP)</th>
<th>Controls (%SNP)</th>
<th>Odds Ratio GT Versus GG</th>
<th>Odds Ratio TT Versus GG</th>
<th>P Value</th>
<th>Reference Nos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mexico</td>
<td>82 (50.0%)</td>
<td>108 (12.5%)</td>
<td>7.90 (3.63-18.31) overall</td>
<td>3.8 × 10⁻⁹³</td>
<td>156</td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>228 (2.2%)</td>
<td>102 (0.0%)</td>
<td>Unable to calculate</td>
<td></td>
<td></td>
<td>156</td>
</tr>
<tr>
<td>China</td>
<td>165 (6.7%)</td>
<td>1013 (1.6%)</td>
<td>4.33 (1.99-9.42) overall</td>
<td>0.001</td>
<td>223</td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>44 (6.8%)</td>
<td>310 (1.8%)</td>
<td>4.34 (1.02-18.49) overall</td>
<td>0.03</td>
<td>66</td>
<td></td>
</tr>
</tbody>
</table>

IPF, idiopathic pulmonary fibrosis; SNP, single nucleotide promoter.
C. Rare Versus Common IPF Risk Alleles

Like other complex diseases, both rare and common variants contribute to risk of pulmonary fibrosis, with rare variants typically having larger effects than the more common variants. However, in IPF the strongest known risk factor for pulmonary fibrosis, the MUC5B promoter variant rs35705950, has a larger effect on disease risk than anticipated by most models of the relationship between disease allele frequency and impact on risk, given the relatively common frequency of the risk allele (FIGURE 5). The rare variants identified in genes like TERT are associated with a large increased risk of disease compared with common variants, but by virtue of their rare frequency, individually account for a small proportion of disease. The more common variants generally have lesser effects, but because they are so common, have the potential to explain more of the population burden of disease. It is important to note that given the low incidence of disease, even the substantial relative increases in risk associated with any of the common variants still result in a low absolute risk for IPF. The impacts of rs35705950 and other common risk factors have similar impacts among both familial and sporadic cases, indicating that studies of either group may have important implications for the other.

D. Mucin Gene Expression and Disease

The risk allele at rs35705950 is a gain-of-function variant associated with increased expression of MUC5B in unaffected subjects (178) and among subjects with IPF (Helling et al., unpublished data). The MUC5B promoter variant rs35705950 is located in a highly conserved, regulatory region 3 kb upstream from the transcription start site of the MUC5B gene. Our results indicate that the variant allele is associated with increased MUC5B mRNA expression in lung tissue from unaffected individuals and patients with IPF (FIGURE 6A; Refs. 177, 178). Moreover, we have found that enhanced expression of MUC5B is localized to the distal airway, respiratory bronchiole, and the honeycomb cyst (FIGURE 6, B AND C; Ref. 177), and that the MUC5B promoter variant rs35705950 enhances expression of MUC5B in the bronchiolo-alveolar epithelia (141). These findings suggest that dysregulation of MUC5B, especially in the peripheral airspace, is a consequence of the MUC5B promoter variant and that overexpression of MUC5B in this region of the lung may be involved in the pathogenesis of IPF.

However, the MUC5B promoter variant rs35705950 does not fully explain the observed differential expression of MUC5B between cases and controls (178, 232). Polymorphisms in the proximal promoter region within 1 kb of the transcription start site of MUC5B are also associated with expression changes in MUC5B (88). Among Japanese, three of these variants (rs17228946, rs17235346, and rs17235353) associated with increased expression were associated with the lung disease panbronchiolitis (88). Moreover, the 4-kb promoter region of the MUC5B gene has 3 CpG islands and other regulatory sites, suggesting that expression of MUC5B may be a function of genetic and non-genetic factors. Ongoing studies are examining epigenetic and environmental influences such as cigarette smoking on MUC5B expression and whether these effects are modified by rs35705950.

Dysregulation of mucin expression is commonly observed in gastric and colorectal cancer (86, 154, 168, 194, 197, 220, 221). The pathways implicated by these studies and animal models are strongly suggestive of alterations in host defense and barrier function as drivers of disease risk. For example, the mouse knockout of MUC2 upregulates the TLR pathway (17), and knockout of MUC4 is associated with increased colonization of Helicobacter pylori (149). Erosion of barrier function due to alterations in MUC1 and MUC4 are important for metastasis in gastric (MUC1)
(197), ovarian (MUC4) (160), and lung cancer (MUC4) (160). There are few examples of specific sequence variants in mucin genes that are convincingly associated with either disease or expression differences, however. One exception is the SNP noted above in MUC1 that is associated with gastric cancer (rs4072037); the exonic SNP is associated with alternative splicing of the MUC1 gene (144). This same SNP may be associated with serous borderline tumors, mucinous borderline, and invasive tumors of the ovary and endometroid tumors (228).

E. Evolutionary Selection

The magnitude of increased risk associated with the risk allele of rs35705950, which occurs at a baseline frequency of ~9% of the European-ancestry population, is striking. Because pulmonary fibrosis typically has an onset well after reproductive age, the T allele at rs35705950, which increases MUC5B expression, is not expected to be subject to negative selection due to increasing risk for pulmonary fibrosis. Recent work demonstrating the necessity of MUC5B for mucociliary clearance, infection control, and immune regulation in mice indicates that increased MUC5B may provide a selective advantage with improved host defense, hence resulting in positive selective pressure on the T allele at rs35705950. In fact, among patients with IPF, the MUC5B promoter variant is associated with an improved prognosis (158) and a lower bacterial burden (137), suggesting a relationship between the variant allele and improved host defense. In light of these observations and the fact that rs35705950 is rare among East Asians and Africans, investigations of the evidence for selection at rs35705950 are underway; the genome-wide data available allow studies of the strength of evidence of selection at the MUC5B locus as well as other loci. Comparison of these data with those in other populations can provide estimates of the age and genealogy of the risk variant that may shed light on important selective events relevant to the genesis of idiopathic pulmonary fibrosis.

III. HOMEOSTATIC AND PATHOBIOLOGICAL FUNCTIONS OF LUNG MUCINS

As a crucial mechanism of homoeostatic innate defense, mucociliary clearance requires the coordinated functions of its two components: mucus and cilia. Mucus is comprised of water, salts, cells, cellular debris, and other macromolecules that are held together in a mesh formed by polymeric mucin glycoproteins (45, 206). Mucins are released by nonciliated secretory cells, club cells in the distal lung (11, 44, 204). In the extracellular space, the mucus gel they form is positioned above an epithelial layer that is also densely populated with ciliated cells. The interactions between mucus and cilia are tightly regulated by factors that include the...
quantities and qualities of mucins produced and the maintenance of a cilia structures and motility. Indeed, dysregulated mucin production, aberrant cilia assembly, and poor control of the salt and water contents in the mucus layer are factors that are known to impact mucociliary clearance function and lead to disease development and exacerbation. The regulation and control of mucin and cilia functions in health and disease are presented below.

A. Mucin Genes

MUC5AC and MUC5B, and their orthologs (“Muc” in mice) are expressed throughout the upper and lower respiratory tract. In human upper airways, MUC5AC is found in the nasopharynx, while MUC5B is found in nasal and oral gland secretions. In tracheobronchial conducting airways, both mucins are expressed by surface and submucosal gland mucous cells, but their expression ordinarily favors higher relative expression of MUC5AC by surface epithelia and MUC5B by the glands. In bronchioles (small airways that lack cartilage or glands), very little MUC5AC is present, and MUC5B predominates. In mice, which lack cartilaginous or gland-rich airways distal to the trachea, in the baseline state surface epithelial cell expression of Muc5b predominate (45, 175, 234).

In diseased conditions in humans, or in animal models of lung disease, mucin expression and localization change dramatically. In human airways, MUC5AC is induced in both submucosal glands and surface epithelia (99, 152, 230). While this finding with MUC5AC is quite consistent, the effects of disease states on MUC5B vary substantially. The variable expression of MUC5B in human disease is best studied in bronchial biopsies, which favor detection of transcript and protein levels in surface epithelia. In human asthma, whereas MUC5AC expression in bronchial biopsies increases significantly and consistently, MUC5B expression decreases (152). Importantly, this decrease is driven primarily by allergic status: patients with the highest levels of allergic inflammation display MUC5B expression levels that are lower than healthy controls and less severely allergic patients (230). Perhaps not surprisingly, the mechanisms that regulate the control and expression of MUC5AC/Muc5ac are characterized more substantially than MUC5B/Muc5b.

Data demonstrating MUC5AC/Muc5ac expression mechanisms have been extensively studied in the context of asthma. The type 2 inflammatory cytokine IL-13 induces MUC5AC/Muc5ac expression in human asthma and in mouse models (114). This process has been linked to the signal transduction and activator of transcription (STAT)-6 pathway (113). However, the lack of a cis STAT-6 DNA binding motif in the MUC5AC/Muc5ac promoter regions suggests that its role is indirect (234). Rather, signaling pathways that involve control through IL-13 and the epi-

dermal growth factor receptor and Notch, ultimately leading to the inactivation of forkhead box A2 (FOXA2) and activation of SAM pointed domain-containing Ets transcription factor (SPDEF) in IL-13, have been characterized extensively (22, 23, 107, 155, 196, 222). Other pathways downstream of inflammation, such as mitogen-activated protein kinases (MAPks), nuclear factor κB (NF-κB), and hypoxia inducible factor-1 (HIF-1) have also been described (24, 25, 234). The roles of these pathways in MUC5AC/Muc5ac regulation have been recently reviewed elsewhere (24, 25, 234).

Given its more complex patterns of expression, much less is known about MUC5B/Muc5b regulation. MUC5B/Muc5b is abundantly expressed in naive airways from early life time points onward (175). However, whereas IL-13 potently stimulates MUC5AC expression, IL-13 negatively affects MUC5B levels in human asthma (152, 230). This is supported further by data in differentiated human airway epithelial cell cultures, where IL-13 directly downregulates MUC5B expression (152, 237). Levels of MUC5B expression by surface epithelia also decrease in early stages of COPD (19, 79). However, MUC5B increases later in COPD (19, 98), and also in cystic fibrosis (CF) (37, 99) and IPF (177, 178) consistent with a lack of an IL-13 driven component in these diseases. In contrast to what is observed in humans, mice show little negative control of Muc5b expression by IL-13. Muc5b increases two- to threefold in inflammation driven by allergen sensitization/challenge or transgenic overexpression of IL-13 (114, 234). However, in a chronic cigarette smoke exposure model of COPD, a transient downregulation followed by recovery of Muc5b expression occurs (Evans, unpublished observation). Taken together, these data suggest that the complex control of MUC5B/Muc5b expression involves multiple levels of regulatory machinery. Indeed, the MUC5B/Muc5b promoter contains both core elements that are highly conserved including the 3-kb single nucleotide promoter variant rs35705950 that is present in primates and controls MUC5B levels by >30-fold in the lungs (FIGURE 6A).

Whether this site is actively bound by transcription factors or comprises an added layer of epigenetic control is yet to be determined. Evidence for epigenetic control of MUC5B has been shown in studies using carcinoma cell lines from various tissues (216), but the role of the SNP in controlling MUC5B expression in these cell lines and in nonpulmonary organs in vivo is not yet known.

B. Mucin Structures

Whereas the molecular control of mucin expression of MUC5AC/Muc5ac and MUC5B/Muc5b have been studied most extensively, the biochemical features of these mucins at the glycoprotein level have been described but not extensively tested at functional levels. In part, this is because of their extraordinarily large sizes (in the mega-Dalton range)
and the characteristics as thick viscoelastic polymers. Both MUC5AC/Muc5ac and MUC5B/Muc5b contain disulfide-rich NH₂ and COOH termini that form disulfide linked polymers (FIGURE 7). These give rise to polymers that can vary in assemblies of 2 to >20 oligomers. Disruption of these with thiol reducing agents such as dithiothreitol or N-acetyl cysteine reduces their sizes and increase mucociliary transport.

Glycosylation is another property that lends to mucin size and charge heterogeneity (35, 45, 206). Glycosylation is controlled by numerous enzymes that include core machinery that introduce O-linked N-acetylgalactosamines to serine and threonine amino acids, followed by Core-linkage and terminal glycosylation machinery that result in very large and diverse carbohydrate structures. Terminal glycans include N-acetylglucosamine, galactose, fucose, and sialic acid, which vary in size, polarity, and charge density, which can also be affected by sulfation. Importantly, terminal glycosylation of mucins utilizes in many cases the same machinery that determines blood antigen types (e.g., Lewis and H antigens). In human asthma, the presence of H antigen secretor status (fucosyltransferase 2-mediated α1,2 fucosylation) has been linked to heightened risk for severe asthma exacerbation and hospitalization (78). This could relate either to changes in host defense or physical obstruction, but these have not been determined.

C. Secreted Polymeric Mucin Functions

The pathological obstructive effects of secreted mucus reflect aberrant control of its homeostatic regulation. Mucus can either be tightly adherent to or loosely associated with underlying epithelia. In the colon, tight association is responsible for keeping the intestinal flora and underlying tissues separated. In the lungs, loosely associated mucus is responsible for trapping inhaled particles, including bacte-

![Image](http://physrev.physiology.org/)

**FIGURE 7.** Polymeric mucin structures. MUC5AC and MUC5B apoprotein backbones are comprised by NH₂-terminal (white) and COOH-terminal (gray) cysteine-rich domains that are conserved in polymeric mucins and von Willebrand factor. The central portions of MUC5AC and MUC5B are comprised of a mucin glycosylation domain (green) that is rich in serines and threonines that are sites of O-linked glycan attachments. The glycosylation domain is interrupted by additional cysteine-rich CysD domains (red). A: the COOH-terminal von Willebrand-like domain (VWD) has a cysteine-knot structure that is the site of disulfide dimerization in the endoplasmic reticulum; each dimer is ~1 μm in length. In the Golgi, dimers further assemble at their NH₂ termini via covalent disulfide bonds and other noncovalent interactions to form extensive polymeric structures that are 10s to 100s of microns long. B: in cysteine-rich CysD domains, intramolecular disulfide bonds also assemble. CysDs contain hydrophobic amino acids and form loops that may promote mucin mesh network alignment. C: in the mucin glycosylation domain, glycan linkages occur on the hydroxylated ends of serine and threonine residues through the initial attachment of N-acetylgalactosamine (GalNac). This initial O-glycosylation step is followed by galactose and/or N-acetylgalactosamine (GlcNac) attachments in specific core linkages. These can be elaborated with additional galactose and GlcNac moieties and be modified further by the attachment of fucose, sialic acid, and sulfate end groups. These terminal groups carry charges and structural characteristics that mediate mucin biophysical properties and interactions with host-derived and foreign materials.
ria, while still allowing for transport out of the airways by ciliary and cough driven forces, followed by elimination by expectoration or swallowing (45, 206).

These clearance mechanisms also remove endogenous debris, including dying epithelial cells and leukocytes. While the means of transport of these is somewhat intuitive, recent data also show that in addition to physical elimination, mucin glycans also interact with carbohydrate interacting lectins that are present on leukocyte surfaces in a highly selective fashion. Sialic acid residues that are present in α2,3 linkages on the surface of MUC5B/Muc5b bind to sialic acid binding immunoglobulin-like lectin receptor (Siglec)-8, which is present on human eosinophils, and Siglec-F, which is present on mouse eosinophils and resident alveolar macrophages (9, 135). These interactions have recently been shown to mediate apoptosis in human and mouse eosinophils (100), but the effects of Muc5b sialoside-Siglec-F signaling in alveolar macrophages are not yet determined. Interactions between siglecs, as well as other glycan-selective immunoreceptors such as dectins, have only recently been identified. Their interactions have opened a new field for studying how first and second lines of innate host defense are linked.

D. Pathological Effects of Mucins

When mucus is overproduced, excessively secreted, and/or overly viscous and adherent, materials that need to be eliminated accumulate and drive disease exacerbations. This is particularly prominent in cystic fibrosis, but adherent mucus plugs are also seen in severe asthma (77, 115, 182), COPD (91), and IPF (178). Mucus accumulation results in plugging that is linked to airflow obstruction and can lead to fatal asphyxiation, but may also compromise underlying cell function. In a mouse model of asthma, genetic deletion of Muc5ac resulted in protection from mucus plugging and airway hyperreactivity acutely (231). Under chronic conditions, mucus accumulation can also lead to pathogen and leukocyte accumulation, which result in impaired immune defense and chronic tissue injury.

E. Cilia

Cilia are organelles that protrude from plasma membranes, with two types of cilia: motile and nonmotile. Motile cilia are present in specialized cells that are involved in fluid and particle transport. Nonmotile cilia are present in almost all cells, which typically have a single primary cilium whose functions include sensory signal transduction. Both are constructed from tubulin-formed axonemes that also contain structural anchors such as dyneins and nexin. In motile cilia, these form a ring of nine microtubule doublets with a central pair of singlet microtubules (9 + 2) axonemes, but in nonmotile cilia, the central pairs are absent, resulting in outer ring only (9 + 0) axonemes (16).

Motile cilia are present in specialized cells involved in particle transport. In the lungs, ciliated cells are abundant throughout the conducting airways and are typically present in roughly equal numbers compared with nonciliated/secretory cells. Ciliated airway epithelial cells are easily identified by their apical surfaces with numerous motile cilia in clusters of 100–300 projecting into the airway lumenspace and surface mucus. Although the ciliated and mucus layers in the lumen were once considered to be separate entities, so-called “sol” and “gel” phases, recent studies demonstrate that this separation is not so distinct. Due to the presence of highly glycosylated membrane mucins (MUC1, MUC4, and MUC16) along and between cilia, the perilin layer (PCL) forms a grafted brush that allows for the free movement of cilia. The overlying, viscous layer of mucus that serves to transport adsorbed particles and pathogens is positioned atop the grafted brush, thus creating what is now referred to as a gel-on-brush structure (18). Both the membrane mucins lining the cilia surfaces and the mucus gel itself are highly glycosylated and hydrated. Charged sugars on the mucins participate with ions such as sodium, calcium, chloride, and bicarbonate to maintain an osmotically balanced hydrogel. Dysfunction of these components could lead to a loss of hydration, which causes cilia to collapse as water is removed from the PCL. For example, excess osmoles in mucus pulls water from the PCL, reducing its thickness via collapse of cilia (63). This results in defective mucus transport, particle/debris accumulation, and impaired repair processes following injury (FIGURE 8).

Disease states that can result in cilia dysfunction include conditions of impaired ion transport such as cystic fibrosis, structural abnormalities of cilia components in primary ciliary dyskinesia (PCD) (101), and in diseases in which the polymeric mucins MUC5B and MUC5AC are overproduced (e.g., asthma) (133, 148). The recent discovery that the MUC5B promoter variant rs35705950 causes excessive MUC5B production in small airways suggests that this could result in impaired cilia-mediated clearance in the small airways. As a result of defective ciliary function, clearance is impaired, leading to chronic airway infection and inflammation, bronchiectasis, and distal lung remodeling occur.

In addition to motile cilia, primary cilia (highly conserved, nonmotile, solitary structures) are present in almost all nondividing cells of the airway and transduce essential signaling information from the extracellular milieu. A large number of ciliopathies (20, 51) as well as polycystic kidney disease (PKD) (38) are associated with defects in the primary cilium. The primary cilium is critical for multiple developmental pathways (55), and their dysregulation is prominent in IPF lung (181). Furthermore, it has also been shown that primary ciliogenesis precedes motile cilia during development and that primary cilia are transiently present in adult mouse airway during repair after respiratory virus injury.
Thus the long-term results of chronic injury-repair processes. Under conditions where mucins are overproduced or hypersecreted (such as the MUC5B promoter variant rs35705950), or when electrolyte homeostasis is disrupted (e.g., through excessive Na⁺ absorption), cilia collapse and aggregates of mucus adhere to airway surfaces, potentially worsening injury. Under injurious conditions, resulting from excess mucus or other forms of lung injury, aberrant repair could lead to abnormal activation of progenitor basal-like cells, resulting in partially differentiated ciliated cells that retain primary cilia, have poorly developed apical motile cilia, and developmental programs that are aberrantly activated.

Tissue damage can itself perpetuate the activation of inflammatory cascades that if persistent can also chronically evoke injury-repair processes. While tissue repair may ordinarily result in resolution of injury and a return to homeostasis, chronic injury and/or aberrant repair can result in remodeling of tissues to pathological states. In diseased lungs, remodeling is seen as chronic mucous cell metaplasia/hyperplasia, small airway and alveolar fibrosis, and emphysematous alveolar destruction. These processes have been shown to be associated with endoplasmic reticulum (ER) stress or hypoxia in intracellular compartments, culminating in the release of proinflammatory and tissue growth factors such as TGF-α and TGF-β and destructive proteases such as matrix metalloproteinases and cathepsins into the extracellular environments. Thus the long-term results of chronic mucus hypersecretion can result in enhanced lung injury.

IV. CLINICAL RELEVANCE OF MUC5B TO PULMONARY FIBROSIS

MUC5B is a validated target in IPF. As noted above, we have found that 1) a common gain-of-function (Figure 6A) promoter variant in MUC5B rs35705950 is the strongest risk factor (genetic and otherwise), accounting for 30–35% of the risk of developing IPF (178) (confirmed in 9 independent studies) (13, 66, 146, 156, 178, 191, 226, 236), including our GWAS [OR for T (minor) allele = 4.51; 95% CI = 3.91–5.21; P = 7.21 × 10⁻⁹⁵ (47)]; 2) rs35705950 can potentially be used to identify individuals with preclinical pulmonary fibrosis (72) and is predictive of radiologic progression of preclinical pulmonary fibrosis (3); 3) IPF subjects with the MUC5B promoter variant rs35705950 have improved survival compared with those without this variant (158) (Figure 2); and 4) MUC5B may be involved in the pathogenesis of pulmonary fibrosis. Specifically, MUC5B message and protein are expressed in bronchiolo-alveolar epithelia of IPF (141, 177) (Figure 6, B and C) and the characteristic IPF honeycomb cysts (177, 178). Reflecting on these advances, it is reasonable to consider whether the MUC5B promoter variant rs35705950 could be used to assist in the diagnosis or clinical management of patients with IPF.

The MUC5B promoter variant rs35705950 may prove helpful in identifying asymptomatic cases of early IPF. For both the familial and sporadic forms IPF, it is generally accepted that pulmonary fibrosis can present a number of years before symptoms become apparent, suggesting that the diagnosis of IPF can be made before the lung has been extensively and irreversibly damaged. While most patients with IPF are diagnosed when the disease is advanced, an earlier diagnosis of IPF will detect patients with a lower burden of lung disease (111, 121, 163), more salvageable lung, and may reveal novel molecular targets for intervention that are only effective in early stages of IPF. Preclinical pulmonary fibrosis (PrePF) is an emerging clinical phenotype that is characterized by specific identifiable chest CT abnormalities (subpleural reticular changes, honeycomb- ing, traction bronchiectasis, centrilobular nodules, and ground glass) (163). Preclinical pulmonary fibrosis has been
reported more frequently among smokers and in families with two or more cases of pulmonary fibrosis (163, 189), and we have found that the MUC5B promoter variant rs35705950 is predictive of those with PrePF (OR = 6.3 per allele; 95% CI 3.1–12.7; Ref. 72). Others have found that among asymptomatic first-degree family members of FIP, 14% have interstitial changes on CT scan and 35% have interstitial abnormalities on transbronchial biopsy (111). Moreover, we have recently found that rs35705950 is predictive of radiologic progression of PrePF (OR = 2.8 per allele; 95% CI 1.8–4.4; Ref. 3) which was associated with a greater decline in FVC (P = 0.0001) and an increased risk of death (HR = 3.7; 95% CI 1.3, 10.7; P = 0.02). Interstitial lung abnormalities on chest CT scan, a radiologic feature of PrePF, has recently been shown to be associated with a greater risk of all-cause mortality in four independent cohorts (162), suggesting that in addition to having radiologic features of IPF, PrePF is likely a harbinger of IPF. While the MUC5B promoter variant is predictive of PrePF, rs35705950 is present in 19% of the population (178), and IPF occurs infrequently (<0.1%; Refs. 151, 165). Thus additional biomarkers are needed to identify individuals with PrePF within at-risk populations. In aggregate, these findings suggest that 1) IPF is underdiagnosed (72, 111); 2) PrePF is prevalent in an identifiable at-risk population, such as asymptomatic family members of FIP families or even sporadic IPF (72, 111); 3) ~75% of the cases of PrePF are progressive (3); 4) radiologic progression of preclinical is associated with increased morbidity and mortality (3, 162); and 5) the MUC5B variant rs35705950 in combination with peripheral blood biomarkers and radiologic screening should prove useful in identifying those with PrePF. Thus we have put forward a model that uses the MUC5B promoter variant rs35705950 to identify a population that is at risk of both PrePF and IPF (FIGURE 9). We postulate that the development of IPF takes many years and that the MUC5B promoter variant can be used to identify individuals early in their path of disease development. While strategies have not been developed for primary and secondary prevention of IPF, identification of both an at-risk population and those with PrePF facilitates intervention prior to the development of advanced lung fibrosis with the goal of preventing the persistent and extensive fibroproliferative response that destroys and displaces normal lung tissue.

**FIGURE 9.** Indicates that the “at-risk” population and the population with preclinical pulmonary fibrosis is large (19% with the MUC5B promoter variant and 1.8% of individuals ≥ 50 years of age, respectively), IPF is diagnosed in a small population with established end-stage disease (111, 163), and preclinical pulmonary fibrosis can be identified using the MUC5B variant rs35705950. In addition, recent findings (3) indicate that preclinical pulmonary fibrosis (detected via chest CT scan) is associated with a poor prognosis, suggesting that preclinical pulmonary fibrosis may be a precursor of IPF.
The MUC5B promoter variant rs35705950 and other genotypes may prove helpful in managing patients with IPF. While IPF is a distinct entity, there are substantial variations in the clinical course of this disease. IPF is a progressive lung disease with death occurring on average within 3–5 yr of diagnosis. Adverse prognosticators include age, cigarette smoking history, and features of more extensive lung disease (finger clubbing, lung volumes, radiologic features, and oxygen desaturation) (96). However, some patients (5–10%) have acute exacerbations (33, 186), and up to 20% have a prolonged survival. While we have shown that IPF subjects with the MUC5B promoter variant rs35705950 have improved survival compared with those without this variant (158) (FIGURE 2), others have found that both a TOLLIP allele (146) and a TLR3 allele (147) also provide prognostic information among individuals with IPF. These findings suggest that IPF subtypes defined by genetic variants may prove helpful in prognosticating among individuals with IPF, and further consideration should be given to inclusion of these prognostic factors in trials designed to test the utility of novel therapeutic agents.

V. PROPOSED MODELS OF MUC5B-INDUCED PULMONARY FIBROSIS

While it is indisputable that the MUC5B promoter variant rs35705950 is a validated gain-of-function etiologic target in PrePF and IPF, the mechanisms that link enhanced production of MUC5B to the development of pulmonary fibrosis are not obvious or even partially understood. However, accumulating evidence suggests that the distal airway epithelium is the likely site of MUC5B overexpression and may at least localize the investigative efforts to understand the fibroproliferative effects of MUC5B in the IPF lung. Given the above considerations and the emerging understanding of the pathogenesis of IPF, there are at least three (and undoubtedly many more) competing hypotheses that link enhanced production of MUC5B in the bronchiole-alveolar region to the development of pulmonary fibrosis.

One line of reasoning focuses on the relationship between overexpression of MUC5B, the involvement of the terminal airway, microscopic honeycomb cyst formation, and repair/regeneration of the distal airspace in IPF (159, 164, 181, 208). The MUC5B promoter variant rs35705950 leads to enhanced expression of MUC5B in the bronchiole-alveolar epithelia (FIGURE 6, B AND C; Ref. 141), and several histopathological analyses (159, 177, 178) have confirmed involvement of MUC5B in the pathogenesis of IPF in that region of the lung. Microscopic honeycomb cysts contain abundant MUC5B, and our recent observations highlight the relationship between MUC5B overexpression and microscopic honeycomb cyst formation (FIGURE 10) (232). Accumulating evidence suggests that the distal airway epithelium is the principal origin of alveolar stem cells. Bronchiole-alveolar epithelial cells give rise to alveolar type I (ATI) and alveolar type II (ATII) cells in bleomycin-induced pulmonary fibrosis in mice (171), and it has recently been demonstrated in mice that distal airway stem cells expressing Krt5 proliferate into interbronchial regions of alveolar...
ablation following H1N1 infection and assemble into alvelo-like structures (112). Furthermore, activated rare lineage-negative, Krt5+ epithelial stem/progenitor cells proliferate and migrate widely to occupy heavily injured areas in the distal lung following H1N1 infection and bleomycin challenge (214). Persistent Notch signaling after H1N1 injury leads to parenchymal cysts, indicative of abortive regeneration, analogous to current concepts regarding honeycomb cyst formation in the IPF lung. In H1N1 lung injury, it has been proposed that honeycomb cysts represent daughter structures derived from the aberrant differentiation of pulmonary progenitors expressing KRT5/Krt5 (112, 159, 171, 172, 214).

In IPF, we have found that KRT5+ cells can be found within the fibrotic interstitium, and KRT5+ cells are in close association with MUC5B-rich honeycomb cysts (177). Cells lining honeycomb cysts in IPF appear to manifest intermediate features of ciliated and distal epithelial phenotypes, suggesting aberrant differentiation (177, 232). We have recently found that lung tissue samples from ~40% of patients with IPF are highly enriched for transcripts of cilium genes, MUC5B, and MMP7, and this molecular phenotype is associated with microscopic honeycombing (FIGURE 10 and Ref. 232). In patient samples, bronchiolized epithelium expresses SOX2 and MUC5B but not SPDEF, consistent with abnormal programming of distal airway epithelia in IPF (159). Moreover, airway epithelium overlying fibroblastic foci contains p63-positive basal cells that are capable of undergoing epithelial to mesenchymal transition (EMT) (85). Importantly, a striking number of cells in honeycomb cysts and bronchiolized epithelium express β-catenin (28), a gene in the Wnt developmental pathway with an established role in IPF (28, 60, 102, 103, 159, 171, 172, 214). The strong association of MUC5B expressing cells with KRT5+ cells lining honeycomb cysts suggests a role for MUC5B expression in the establishment or maintenance of honeycomb cysts by KRT5+ cells, and points to a role for the bronchiolo-alveolar epithelium as a potential site of initiation of the fibroproliferative response with involvement of MUC5B and stem/progenitor cells.

Based on these considerations, we hypothesize that excessive production of MUC5B by stem cells that attempt to regenerate injured bronchiolar and alveolar epithelium disrupt normal developmental pathways and highjack the normal reparative mechanisms in the distal lung, resulting in chronic fibroproliferation and honeycomb cyst formation (FIGURE 11). Given the large size and extensive glycosylation of MUC5B/Muc5b protein, we postulate that its production presents a significant metabolic stress, which can interfere with differentiation of airway stem cells. This effect is enhanced in carriers of the gain-of-function MUC5B variant rs35705950 allele. Alternatively, excess MUC5B/Muc5b could affect KRT5/Krt5 cells, perhaps by modifying growth factor signaling pathways. While these disrupted development mechanisms may directly initiate fibroproliferative responses, it is notable that the murine bleomycin model can generate fibrosis without honeycombing (232). Although speculative, it is more likely that honeycomb cysts represent unsuccessful reparative responses, which then interfere with basic signaling events leading to the emergence of reprogrammed cells. These structures subsequently prevent establishment of normal alveolar homeostasis and may promote persistent foci of fibroproliferative activity.

A second line of reasoning focuses on the possibility that IPF is a mucociliary disease caused by recurrent injury/inflammation/repair at the bronchoalveolar junction, which is initiated and exacerbated by overexpression of MUC5B and reduced mucociliary function, retention of particles, and enhanced lung injury (FIGURE 12). Based on the relationship between the MUC5B promoter variant rs35705950 and excess production of MUC5B specifically at the bronchoalveolar junction, we hypothesize that too much MUC5B impairs mucociliary function (14, 18), causes excessive retention of inhaled substances (air pollutants, particles and chemicals from cigarette smoke, microorganisms, etc.), and, over time, the foci of lung injury lead to scar tissue and persistent fibroproliferation that expands and displaces normal lung tissue. Excess concentrations of MUC5B may also physically impair ciliary function (18), further enhancing the retention of inhaled particles, dissolved chemicals, and microorganisms. Given the importance of environmental exposures, such as asbestos and silica, in the development of other forms of interstitial lung disease, it is logical to speculate that common inhaled particles, such as those associated with cigarette smoke or air pollution, might cause exaggerated interstitial injury in individuals who have defects in mucociliary clearance and mucosal host defense.

A third line of reasoning focuses on the relationship between MUC5B and motile cilia in the distal airway. We identified a cilium gene expression signature in the IPF lung contains both primary and motile cilia genes (FIGURE 10). A role of primary cilia in the pathogenesis of IPF would be consistent with the aberrant repair process following the injury to bronchiolo-alveolar regions of the lung and the concept that honeycomb cysts could be the result of aborted attempts to regenerate the terminal airways via activation of developmental pathways that signal in the primary cilium (55) and have been implicated extensively in IPF (181) (FIGURE 8). Alternatively, a role of cilium gene expression and function in multiciliated airway epithelial cells in the distal airways and bronchioles would be consistent with the idea that accumulation of mucus, caused by the MUC5B promoter variant, in the bronchoalveolar region of the lung impairs mucociliary function (FIGURE 8). Available mouse models are helping to clarify this somewhat, but more work is needed to define the pathobiology of MUC5B among carriers of the variant rs35705950 allele in humans.
Lastly, given the large size and heavily glycosylated characteristic of airway mucins (45), we have been particularly interested in the relationship between the MUC5B promoter variant rs35705950, enhanced transcription of MUC5B and posttranscriptional processing of MUC5B, endoplasmic reticulum (ER) stress and the unfolded protein response (UPR), as a pathway to pulmonary fibrosis. In fact, ER stress is observed in a number of different diseases (62), including IPF (5, 26, 104–106, 117, 118, 125, 138, 201), and our preliminary findings indicate that excess production of Muc5b induces ER stress and activates UPR both at baseline and after profibrotic stimulation. The adaptive or homeostatic UPR attempts to restore normal protein folding by jointly coordinating three different responses: IRE1/XBP1, the PERK/ATF4, and BiP/ATF6 (83). These three molecular pathways of UPR involve a specific membrane ligand and result in the activation of specific gene targets (57, 83). When proteostasis is abnormal, stress induced by the accumulation of unfolded proteins in the ER is a feature of secretory cells that undergo apoptosis. While ER stress has been observed in IPF (5, 26, 104–106, 117, 118, 125, 138, 201), in experimental models of lung fibrosis, activation of the UPR in alveolar epithelial cells can lead to EMT (202, 238) and fibroproliferative lung disease (117, 201). ER stress is also at play in the airway epithelium and is required for MUC5B production (127). Not surprisingly, ER stress and apoptosis interfere with the normal epithelial response to injury, wound healing, and tissue repair which are hallmarks of IPF (95), can be modeled in vitro (184), and may be particularly relevant to IPF that is associated with the gain-of-function MUC5B promoter variant.

VI. PERSPECTIVE

The discovery that MUC5B plays a key role in the etiology of IPF, accounting for 30–35% of individuals who develop this disease, is important for the following reasons. First, the assay for MUC5B promoter variant rs35705950 is an inexpensive genetic test that could be offered to high-risk asymptomatic individuals to identify an at-risk population or identify individuals in the early stages of disease. While high-risk populations, such as first-degree relatives of patients with either familial or sporadic IPF, are obvious tar-
gets for screening, it is possible that screening for IPF will be more widely deployed as therapeutic interventions become more effective in the early stages of disease and have fewer side effects. Preventing the destruction of lung tissue in an asymptomatic individual is a much more effective strategy than treating patients with IPF once extensive lung damage has occurred. Second, lung mucins may prove to be an effective therapeutic target for the treatment of IPF. We postulate that terminal bronchiolar and possibly alveolar epithelial inflammation/injury/repair as well as subsequent fibroproliferation induced by MUC5B can be overcome, in part, by decreasing the concentration of MUC5B in the distal airspace. Moreover, we think that our ability to restore MUC5B homeostasis, improve mucociliary clearance, and reduce the effects of excessive mucus accumulation may prove especially effective in limiting the progression of preclinical/mild stages of disease. However, given the spatial and temporal distribution of the microscopic lesions in IPF, reducing the concentration of MUC5B in the distal airspace may also prove effective in stabilizing established IPF. Third, since IPF subjects with the MUC5B promoter variant rs35705950 have improved survival compared with those without this variant (158), it is logical to control for the MUC5B genotype (and potentially other genotypes) when considering the impact of novel therapeutic agents. Lastly, the discovery that MUC5B plays a role in the etiology of IPF provides a novel direction in understanding the basic mechanisms involved in the pathogenesis of IPF. Although we have learned a great deal about many of the basic biological mechanisms involved in IPF, the disease is heterogeneous and the final common pathways of fibrogenesis are not well understood (145, 209). The MUC5B discovery not only identified a novel target to consider, this discovery also highlights the potential involvement of the distal airway, the bronchiolo-alveolar region, and mucociliary transport in the development of this complex and incurable disease.

ACKNOWLEDGMENTS

C. M. Evans, T. E. Fingerlin, and M. I. Schwarz contributed equally to this manuscript.
Address for reprint requests and other correspondence: D. A. Schwartz, Univ. of Colorado, 12631 East 17th Ave., B178, Aurora, CO 80045 (e-mail: david.schwartz@ucdenver.edu).

GRANTS

This work was supported by National Institutes of Health (NIH) Grants R01-HL097163, R21-HL120770, UH2-HL123442, P01-HL092870, and R25-ES025476 as well as Department of Veterans Affairs Merit Award 1101BX001534. C. M. Evans was supported by NIH Grants R01-HL083936, R21-ES023384, and R01-HL130938 as well as American Heart Association Grant 14GRNT19990040.

DISCLOSURES

D. A. Schwartz has a patent on the use of the MUC5B promoter variant rs35705950 for the risk prediction, diagnosis, prognosis, and treatment of pulmonary disorders (US Patent No. 8,673,565); has a provisional patent on the early recognition and treatment of IPF; and has developed a company (ElevenP15) based on the early recognition and treatment of IPF.

REFERENCES


222. Wan H, Kaestner KH, Ang SL, Ikegami M, Finkelman FD, Stahlman MT, Fulkerson PC,
221. Walsh MD, Clendenning M, Williamson E, Pearson SA, Walters RJ, Nagler B, Packenas
220. Wakatsuki K, Yamada Y, Narikiyo M, Ueno M, Takayama T, Tamaki H, Miki K,
217. Vourlekis JS, Schwarz MI, Cool CD, Tuder RM, King TE, Brown KK. Nonspecific
214. Vaughan AE, Brumwell AN, Xi Y, Gotts JE, Brownfield DG, Treutlein B, Tan K, Tan V,
213. Van Moorsel CH, van Oosterhout MF, Barlo NP, de Jong PA, van der Vis JJ, Ruven HJ,
211. Tokumaru M, Kasuga I, Kusumaa S, Tsuruoka T, Kiyokawa H, Kuwabara S, Ichinose
210. Travis WD, Hunninghake G, King TE Jr, Lynch DA, Colby TV, Galvin JR, Brown KK,
208. Turco LA, Sessions GD, de Vera GN, Hruban RH, Colby TV, Hansell DM, Shiendley ML,
207. Takashima K, Kajii T. Interstitial pulmonary fibrosis in two sisters. Possible autosomal
205. Takazawa T, Hayashi Y, Sugimoto S, Aoki Y, Kasahara H, Morita M. Expression of