The MUC5B promoter polymorphism is associated with specific interstitial lung abnormality subtypes

Rachel K. Putman1,12, Gunnar Gudmundsson2,12, Tetsuro Araki3,4, Mizuki Nishino3,4, Sigurdur Sigurdsson5, Elías F. Gudmundsson5, Gudny Eiríksdottir5, Thor Aspelund6,7, James C. Ross7,8, Raúl San José Estépar3,8, Ezra R. Miller1, Yoshitake Yamada3,4, Masahiro Yanagawa9, Noriyuki Tomiyama9, Lenore J. Launer10, Tamara B. Harris10, Souheil El-Chemaly1, Benjamin A. Raby1,7, Michael H. Cho1,7, Ivan O. Rosas1, George R. Washko1,4, David A. Schwartz1,4, David A. Schwartz11, Edwin K. Silverman1,7, Vilmundur Gudnason5,6, Hiroto Hatabu3,4 and Gary M. Hunninghake1,4

Affiliations: 1Pulmonary and Critical Care Division, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA. 2Dept of Respiratory Medicine and Sleep, Faculty of Medicine, Landspital University Hospital, University of Iceland, Reykjavik, Iceland. 3Dept of Radiology, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA. 4Center for Pulmonary Functional Imaging, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA. 5Icelandic Heart Association, Kopavogur, Iceland. 6University of Iceland, Reykjavik, Iceland. 7The Channing Division of Network Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA. 8Surgical Planning Laboratory, Dept of Radiology, Brigham and Women’s Hospital, Boston, MA, USA. 9Dept of Radiology, Osaka University Graduate School of Medicine, Suita-city, Osaka, Japan. 10Intramural Research Program, National Institute of Aging, NIH, Bethesda, MD, USA. 11Dept of Medicine, University of Colorado, Denver, CO, USA. 12These authors contributed equally to this work.

Correspondence: Gary M. Hunninghake, Pulmonary and Critical Care Division, Dept of Medicine, Brigham and Women’s Hospital, 75 Francis Street, Boston, MA 02115, USA. E-mail: ghunninghake@partners.org

@ERSpublications
The MUC5B genotype is associated with specific subtypes of ILA, with varying heterogeneity in underlying populations http://ow.ly pcmH30dRsZ7


ABSTRACT The MUC5B promoter polymorphism (rs35705950) has been associated with interstitial lung abnormalities (ILA) in white participants from the general population; whether these findings are replicated and influenced by the ILA subtype is not known. We evaluated the associations between the MUC5B genotype and ILA in cohorts with extensive imaging characterisation.

We performed ILA phenotyping and MUC5B promoter genotyping in 5308 and 9292 participants from the AGES-Reykjavik and COPDGene cohorts, respectively.

We found that ILA was present in 7% of participants from the AGES-Reykjavik, 8% of non-Hispanic white participants from COPDGene and 7% of African-American participants from COPDGene. Although the MUC5B genotype was strongly associated (after correction for multiple testing) with ILA (OR 2.1, 95% CI 1.8–2.4, p=1×10−26), there was evidence of significant heterogeneity between cohorts (I²=81%). When narrowed to specific radiologic subtypes, (e.g. subpleural ILA), the MUC5B genotype remained strongly associated (OR 2.6, 95% CI 2.2–3.1, p=1×10−10) with minimal heterogeneity (I²=0%). Although there was no evidence that the MUC5B genotype influenced survival, there was evidence that MUC5B genotype improved risk prediction for possible usual interstitial pneumonia (UIP) or a UIP pattern in non-Hispanic white populations.

The MUC5B promoter polymorphism is strongly associated with ILA and specific radiologic subtypes of ILA, with varying degrees of heterogeneity in the underlying populations.
Introduction
Specific patterns of radiologic abnormalities on chest computed tomography (CT) scans (termed interstitial lung abnormalities; ILAs) [1, 2], could represent an early or mild stage of pulmonary fibrosis or other interstitial lung diseases (ILDs). Evidence in support of that hypothesis includes physiologic and clinical outcome data demonstrating that ILAs are associated with measures of reduced pulmonary function [1–5] and exercise tolerance [6], an increased rate of respiratory symptoms [2] and death [7]. Further evidence linking ILA to pulmonary fibrosis includes the fact that the genetic polymorphism most consistently associated with idiopathic pulmonary fibrosis (IPF) (the minor allele of the single nucleotide polymorphism (SNP) rs35705950 in the promoter region of the mucin 5B (MUC5B) gene) [8] is associated with ILA in the Framingham Heart Study (FHS) [2]. Despite the latter finding, it is not known whether the association between the MUC5B promoter polymorphism and ILA would replicate in additional populations, and whether specific radiologic patterns affect the associations.

We hypothesised that the MUC5B genotype would be associated with ILA, and that these associations would depend on specific radiologic patterns of ILA. To test these hypotheses, we evaluated the association between ILA (and radiologic subtypes of ILA) and the MUC5B genotype in participants from the Age Gene/Environment Susceptibility (AGES)-Reykjavik Study and participants from the Genetic Epidemiology of COPD Study (COPDGene). Based on the results, additional analyses were performed to determine if the MUC5B genotype would influence survival, and if it could help to improve risk prediction for ILA.

Methods
Study design
Protocols for participant enrolment in the AGES-Reykjavik study and COPDGene have been previously reported [1, 9, 10]. The AGES-Reykjavik study is a longitudinal birth cohort derived from the Reykjavik Study, which was established in 1967, and includes men and women born in Reykjavik, Iceland from 1907 to 1935 who are now followed by the Icelandic Heart Association [9]. Of the 5764 participants recruited from January 2002 to February 2006, 5308 (92%) had both chest CT and genotypic information available, and were included in the analysis. COPDGene is a multicentre longitudinal study of smokers, designed to identify the epidemiologic and genetic risk factors for chronic obstructive pulmonary disease (COPD). Participants were excluded from COPDGene if they had a history of known lung disease other than asthma, emphysema or COPD [10]. Of the 10364 participants recruited between November 2007 and April 2010, 9292 (90%) had both chest CT scans and genotypic information passing quality control, and were included in the analysis (this number includes 64 participants excluded from primary COPDGene analyses due to the presence of bronchiectasis or ILD identified on chest CT scans after recruitment). Of the 9292 participants included from COPDGene, 6134 (66%) were non-Hispanic white participants and 3158 (34%) were African-Americans. Written informed consent was obtained from all participants, including consent for genetic studies. The institutional review boards of the Brigham and Women’s Hospital and participating centres approved this study.

This article has supplementary material available from erj.ersjournals.com.

Received: March 14 2017 | Accepted after revision: May 31 2017

Support statement: COPDGene is supported by NIH Award Number R01 HL089897 and R01 HL089856. R.K. Putman is supported by NIH Grant Number T32 HL07633. G. Gudmundsdottir is supported by the Oddur Olafsson Fund, project grant 141513-051 from the Icelandic Research Fund and Landspitali Scientific Fund A-2015-030 and A-2016-023. M. Nishino is supported by NIH Grant Number K23 CA157631. J.C. Ross is supported by NIH Grant Number K25 HL130637. R. San José Estépar is supported by NIH Grant Numbers: K25 HL104085 and R01 HL116473. E.R. Miller is supported by NIH Grant Number T32 HL07633. Y. Yamada is supported by project grant 141513-051 from the Icelandic Research Fund and Landspitali Scientific Fund A-2015-030. S. El-Chemaly is supported by NIH Grant Number R01 HL130275. B.A. Raby is supported by NIH Grant Numbers: U01 AI095219, R01 HL0866601, R01 HL118455, R01 HL123546, R01 HL130974 and P01 HL132825. M.H. Cho is supported by NIH Grant Number R01 HL113264. I.O. Rosas is supported by NIH Grant Numbers: R01 HL130974, R01 HL129920 and P01 HL114501. G.R. Washko is supported by NIH Grant Number: R01 HL122464 and R01 HL116473. D.A. Schwartz is supported by NIH Grant Numbers: R25 ES025476, P01 HL092870, R01 HL097163, R33 HL120770, UH2 HL123442 and the Veterans Administration Grant Number 101 BX001534. E.K. Silverman is supported by NIH Grants: R01 HL089856, R01 HL113264, R33 HL120794 and P01 HL114501. The Age, Gene/Environment Susceptibility-Reykjavik Study was supported by NIA grant: 2712012002C, NIH contracts N01-AG-1-2100 and HHSN2712012002C, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association) and the Althingi (the Icelandic Parliament). V. Gudnason is supported by NIA grant: 2712012002C and project grant 141513-051 from the Icelandic Research Fund. G.M. Hunningshake and this work were supported by NIH Grant Numbers: R01 HL111024, R01 HL130974 and project grant 141513-051 from the Icelandic Research Fund. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com
Genotyping
All genotyping of the MUC5B promoter polymorphism (rs35705950) was done using TaqMan Genotyping Assays (Applied Biosystems, ThermoFisher Scientific, Foster City, CA, USA) [2, 8].

Chest CT characterisation
Methods for characterising ILA in the initial 2508 participants from COPDGene and those from AGES-Reykjavik have been previously described [1, 7]. Similar methods were used to characterise ILA in the remaining 6784 participants from the COPDGene. Chest CT scans were evaluated by up to three readers (chest radiologists and pulmonologists) using a sequential reading method [11]. ILA were defined as specific non-dependent patterns of increased lung density including ground-glass, reticular abnormalities, diffuse centrilobular nodules, nonemphysematous cysts and honeycombing or traction bronchiectasis, affecting greater than 5% of any lung zone (figure 1). Chest CT scans with focal or unilateral ground-glass or reticular abnormalities, or patchy ground-glass abnormalities were considered indeterminate, (additional details in online supplementary material).

To determine if the associations between ILA and the MUC5B genotype were dependent on specific radiologic patterns, further imaging-based classification was then performed on all scans with ILA present. First, ILA was classified by the presence or absence of definite fibrosis (defined as evidence of pulmonary parenchymal architectural distortion, such as traction bronchiectasis or honeycombing) [2, 5, 7] into two groups: ILA with definite fibrosis and ILA without fibrosis. The scans with ILA were then classified by consistency with a usual interstitial pneumonia (UIP) pattern (inconsistent, possible and UIP) according to ATS/ERS/JSRS/ALAT criteria [12]. Finally, chest CT scans with ILA were classified by the type and location of radiologic densities observed [1, 5] (online supplementary material and figure 1). All ILA subtyping was performed by a consensus of at least three readers, who were blind to any participant specific information. Quantitative measures of emphysema (percentage of lung below 950 Hounsfield units) were measured using the Airway inspector (www.airwayinspector.org) [13].

Statistical analyses
All genetic analyses were performed using additive genetic models [8]. Logistic regression was used to assess the MUC5B SNP associations with ILA and ILA subtypes, and Cox proportional hazards models were used to analyse the time-to-mortality. In Cox models, all variables were assessed and none was found to violate the proportional hazards assumption. Multivariable models were adjusted for age, sex and smoking behaviour (pack-years). Mega analysis was performed by pooling the participant level data, and p-values reported for the combined cohorts were corrected for multiple testing using a Bonferroni correction. I² values to assess heterogeneity between cohorts were calculated using the DerSimonian and Laird method [14]. To evaluate the ability of the MUC5B genotype to predict ILA (and ILA subtypes), we first evaluated clinical variables and risk factors for ILA based on previous reports [1, 2, 5, 7] and significant findings from our association analyses. Recursive partitioning using Hosmer-Lemeshow tests were used to assess goodness of fit for clinical variables (online supplementary material). Receiver operating characteristic curves were then generated to obtain areas under the curve (AUC) and create C-statistics, and Wald tests assessed whether the addition of the MUC5B minor allele improved the ability to predict ILA. All analyses were performed using the SAS version 9.4 software (SAS Institute Inc, Cary, NC, USA). All p-values were two-sided and a level of 0.05 was considered statistically significant.

Results
ILA prevalence and baseline characteristics
The prevalence of participants with ILA, indeterminate ILA status and without ILA in the AGES-Reykjavik cohort [7] has been previously reported and the percentages were similar when subset to participants with genotypic information. In AGES-Reykjavik, ILAs were present in 377 (7%), 3209 (60%) showed no ILA and 1722 (32%) had an indeterminate ILA status (table 1). In non-Hispanic white participants from COPDGene, 485 (8%) had ILA, 3667 (60%) showed no ILA and 1982 (32%) had an indeterminate ILA status. In COPDGene, 223 African-American participants (7%) had ILA, 1728 (55%) showed no ILA and 1207 (38%) had an indeterminate ILA status (table 1). In AGES-Reykjavik, 4.4% (n=236) had possible UIP, 0.32% (n=17) had UIP and 2.4% (n=128) had definite fibrosis. In non-Hispanic white participants from COPDGene, 3.4% (n=210) had possible UIP, 0.2% (n=12) had UIP and 1.6% (n=101) had definite fibrosis. In African-American participants from COPDGene, 2.1% (n=66) had possible UIP, 0.09% (n=3) had UIP and 0.8% (n=25) had definite fibrosis (online supplementary table S2).
The baseline characteristics in AGES-Reykjavik and COPDGene, stratified by race, are indicated by the presence or absence of ILA in table 1. Baseline characteristics of participants from COPDGene with an indeterminate ILA status are presented in online supplementary table S1, and have been previously published in the AGES-Reykjavik cohort [7]. In all cohorts, participants with ILA were significantly older than those without ILA. In both AGES-Reykjavik and non-Hispanic white participants from COPDGene, participants with ILA had a greater number of pack-years and were more likely to be actively smoking, as compared to those without ILA; whereas in African-Americans from COPDGene, no significant differences were associated with ILA in terms of pack-years or current smoking status.

**Interstitial lung abnormalities and the MUC5B promoter polymorphism**

The minor allele frequency of the MUC5B promoter SNP (rs3570950) was 12.7% in AGES-Reykjavik, 10.3% in non-Hispanic white participants from COPDGene and 2% in African-American participants.

FIGURE 1 Chest computed tomographic (CT) images depicting radiologic subtypes and the overlap between subtypes of interstitial lung abnormalities (ILA). In all panels the blue arrows point to areas of ILA without fibrosis, the red arrows point to areas of ILA with fibrosis, each panel (1–6) represents one participant. Panels 1–3 demonstrate patterns of ILA that are inconsistent with usual interstitial pneumonia (UIP); panels 4–5 demonstrate patterns of ILA that are possible UIP; and panel 6 is a pattern of ILA that is consistent with UIP. Panel 1 represents non-fibrotic, centrilobular predominant ILA, with an area zoomed in to highlight the centrilobular ground-glass nodules. Panel 2 represents a non-fibrotic, mixed pattern of ILA; in 2a, the blue arrow points to subpleural reticulation; in 2b the arrows demonstrate both subpleural and centrilobular ground-glass. Panel 3 represents fibrotic (see the red arrows in 3b), radiologic interstitial lung disease (ILD) that is inconsistent with UIP, due to the pleural plaque (blue arrow) in 3a. Panel 4 represents non-fibrotic, subpleural predominant ILA; blue arrows point to subpleural reticulation. Panel 5 represents fibrotic, subpleural predominant ILA, with red arrows in both panels pointing to traction bronchiectasis. Panel 6 represents fibrotic, radiologic ILD; red arrows highlight traction bronchiectasis and honeycombing.
from COPDGene (consistent with the reported population diversity allelic frequency in the SNP database). The SNP was found to be in Hardy–Weinberg equilibrium in all cohorts. At least one copy of the \( \text{MUC5B} \) promoter polymorphism was noted in 44% (166 out of 377), 27% (131 out of 485) and 5% (12 out of 223) of those with ILA in AGES-Reykjavik, non-Hispanic white participants, and African-American participants from COPDGene, respectively. After adjustment for multiple testing, the \( \text{MUC5B} \) genotype was observed to be strongly associated with ILA (OR 2.1, 95% CI 1.8–2.4; \( p=1\times10^{-26} \)), despite significant heterogeneity between cohorts (I\(^2\)=81%) (table 2).

The \( \text{MUC5B} \) promoter polymorphism and radiologic patterns of ILA

Whereas there was some variability in associations between the \( \text{MUC5B} \) promoter polymorphism and radiologic subtypes of ILA across cohorts, consistent patterns emerged. For example, after adjustment for covariates and despite moderate heterogeneity between cohorts (I\(^2\)=59%), participants with the \( \text{MUC5B} \) genotype were consistently associated with definite fibrosis (OR 3.0, 95% CI 2.4–3.7; \( p=8\times10^{-22} \)) compared to those without ILA, table 2, figure 1. There was also evidence that in addition to consistent association (or lack thereof) with the \( \text{MUC5B} \) genotype, when narrowed to specific radiologic phenotypes,

### Table 2: Association between interstitial lung abnormalities (ILAs) and the \( \text{MUC5B} \) promoter polymorphism

<table>
<thead>
<tr>
<th>ILA subtype</th>
<th>AGES-Reykjavik p-value (non-Hispanic, white)</th>
<th>COPDGene p-value (non-Hispanic, white)</th>
<th>COPDGene p-value (African-American)</th>
<th>COPDGene p-value (African-American)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILA</td>
<td>( 2.7 ) (2.2, 3.2) ( 1\times10^{-22} )( ^\text{c} )</td>
<td>( 2.0 \times 10^{-6} ) ( 0.19 )</td>
<td>( 1.9 \times 10^{-8} ) ( 0.09 )</td>
<td>( 1.9 \times 10^{-8} ) ( 0.09 )</td>
</tr>
<tr>
<td>ILA without fibrosis( ^f )</td>
<td>( 2.3 ) (1.8, 2.9) ( 5\times10^{-13} )</td>
<td>( 0.0044 )</td>
<td>( 0.99 ) (0.5, 2.2)</td>
<td>( 0.97 )</td>
</tr>
<tr>
<td>Definite fibrosis</td>
<td>( 3.3 ) (2.4, 4.4) ( 5\times10^{-15} )</td>
<td>( 2.1 ) (1.4, 3.1)</td>
<td>( 6.2 ) (2.3, 16.8) ( 0.003 )</td>
<td>( 59.0 ) (3.0, 2.4, 3.7)</td>
</tr>
<tr>
<td>Centrilobular</td>
<td>( 1.3 ) (0.7, 2.5) ( 4\times10^{-4} ) ( 0.09 )</td>
<td>( 1.2 ) (0.4, 3.8)</td>
<td>( 0.79 ) ( 1.5 )</td>
<td>( 0.4 ) (0.2, 1.1)</td>
</tr>
<tr>
<td>Subpleural</td>
<td>( 3.0 ) (2.3, 3.7) ( 2\times10^{-19} )</td>
<td>( 2.3 ) (0.98, 5.4) ( 0.06 )</td>
<td>( 0 ) (2.6, 2.5, 3.1)</td>
<td>( 1\times10^{-20} )</td>
</tr>
<tr>
<td>Mixed</td>
<td>( 2.3 ) (1.6, 3.3) ( 2\times10^{-5} )</td>
<td>( 1.06 ) (0.3, 3.4)</td>
<td>( 0.92 )</td>
<td>( 0 ) (1.5, 2.1)</td>
</tr>
<tr>
<td>Radiologic ILD( ^# )</td>
<td>( 4.4 ) (2.2, 9.0)</td>
<td>( 4.8 ) (2.5, 8.9) ( 1\times10^{-6} )</td>
<td>( 0 ) (4.4, 2.8, 6.8) ( 5\times10^{-10} )</td>
<td>( 0.0003 )</td>
</tr>
<tr>
<td>Inconsistent with UIP( ^% )</td>
<td>( 2.2 ) (1.7, 3.0)</td>
<td>( 0.75 ) (0.7, 1.3)</td>
<td>( 0.75 ) ( 1\times10^{-11} )</td>
<td>( 0 ) (1.1, 1.7)</td>
</tr>
<tr>
<td>Possible UIP</td>
<td>( 2.8 ) (2.2, 3.6) ( 9\times10^{-17} )</td>
<td>( 2.5 ) (1.9, 3.3)</td>
<td>( 2.7 ) (1.1, 6.3) ( 30 )</td>
<td>( 0 ) (2.7, 3.2, 3.2)</td>
</tr>
<tr>
<td>UIP</td>
<td>( 4.4 ) (2.2, 9.0) ( 4\times10^{-5} )</td>
<td>( 4.6 ) (2.0, 10.4) ( 0.003 )</td>
<td>( 0 ) (4.1, 2.1, 8.1)</td>
<td>( 0 ) (4.1, 2.1, 8.1)</td>
</tr>
</tbody>
</table>

\( ^\text{c} \): analyses of the \( \text{MUC5B} \) genotype were performed using additive genetic models; odds ratios are per copy of the \( \text{MUC5B} \) minor allele; \( ^f \): models are adjusted for age, sex and tobacco exposure; \( ^\# \): adjusted using the DerSimonian and Laird method; \( ^\% \): p-value corrected for multiple testing using Bonferroni correction; \( ^\% \): fibrosis is evidence of pulmonary parenchymal architectural distortion; \( ^\# \): analysis was not done in African-Americans; no participants with radiologic interstitial lung disease (ILD) had the \( \text{MUC5B} \) genotype; \( ^\% \): analysis was not done in African-Americans; no participants with usual interstitial pneumonia (UIP) had the \( \text{MUC5B} \) genotype; AGES: Age Gene/Environment Susceptibility.
there was minimal heterogeneity between cohorts. After adjustment for covariates, the $MUC5B$ promoter polymorphism was consistently associated with a possible UIP pattern (OR 2.7, 95% CI 2.3–3.2; $p=1\times10^{-30}$), with essentially no between-cohort heterogeneity ($\Gamma^2=1\%$), (table 2, figure 1). However, there was no evidence of any association with the $MUC5B$ promoter polymorphism when ILA was limited to those with a centrilobular pattern (OR 0.91, 95% CI 0.63–1.3; $p=1.0$, $\Gamma^2=15\%$), (table 2, figure 1). Additional results, subset to participants by age are presented in online supplementary tables S3 and S4.

The $MUC5B$ promoter polymorphism and ILA prediction

Based on the consistent associations between the $MUC5B$ promoter polymorphism and ILA subtypes, we sought to determine whether knowledge of the $MUC5B$ genotype alone could predict definite fibrosis, and a possible UIP or UIP pattern, on chest CT. In all cohorts, the $MUC5B$ genotype improved risk prediction for definite fibrosis (C-statistic 0.64, 95% CI 0.60–0.69, $p=0.0001$; C-statistic 0.57, 95% CI 0.52–0.62, $p=0.0007$; C-statistic 0.58, 95% CI 0.50–0.65, $p=0.0005$ in the AGES-Reykjavik, non-Hispanic white participants, and African-American participants from COPDGene, respectively). In the non-Hispanic white populations, the $MUC5B$ genotype improved risk prediction for the presence of a possible UIP or UIP pattern (C-statistic 0.66, 95% CI 0.61–0.71, $p<0.0001$; C-statistic 0.60, 95% CI 0.57–0.63, $p<0.0001$ in AGES-Reykjavik and in non-Hispanic white participants from COPDGene, respectively). Risk prediction was not improved in African-Americans (C-statistic 0.52, 95% CI 0.49–0.56, $p=0.06$ in African-American participants from COPDGene), see table 3, figure 2.

We next sought to determine whether carrying the $MUC5B$ genotype could add to the clinical characteristics and increase risk prediction for ILA subtypes. When added to models of best-fitting clinical characteristics (age, sex and pack-years), the $MUC5B$ genotype improved risk prediction for definite fibrosis in AGES-Reykjavik (C-statistic 0.70–0.75, $p=0.004$ for comparison), but not in populations from COPDGene (C-statistic 0.76–0.76, $p=0.22$ for comparison; and C-statistic 0.70–0.73, $p=0.34$ for comparison in non-Hispanic white participants and African-American participants from COPDGene, respectively) (table 3). When added to models of best-fitting clinical characteristics, the $MUC5B$ genotype improved risk prediction for the presence of a possible UIP or UIP pattern in white populations (C-statistic 0.70–0.76, $p=0.001$ for comparison; and C-statistic 0.71–0.75, $p=0.0008$ for comparison in the AGES-Reykjavik and non-Hispanic white participants from COPDGene, respectively). However, similar findings were not observed in African-American participants from the COPDGene (C-statistic 0.70–0.70, $p=0.50$ for comparison) (table 3 and figure 2). Additional models for risk prediction are presented in table 3 and online supplementary table S5.

### TABLE 3 MUC5B genotype and prediction of interstitial lung abnormality

<table>
<thead>
<tr>
<th></th>
<th>MUC5B minor allele</th>
<th>Clinical data</th>
<th>Clinical data+MUC5B minor allele</th>
<th>p-value for the comparison of clinical with Clinical+MUCB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C-statistic (95% CI)</td>
<td>p-value</td>
<td>C-statistic (95% CI)</td>
<td>p-value</td>
</tr>
<tr>
<td><strong>AGES-Reykjavik</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| ILA               | 0.61 (0.59, 0.64)  | <0.0001       | 0.64 (0.61, 0.66) | <0.0001       | 0.72 (0.69, 0.74) | <0.0001       |<0.0001
| Subpleural and radiologic ILD | 0.63 (0.60, 0.67) | <0.0001       | 0.66 (0.63, 0.70) | <0.0001       | 0.72 (0.69, 0.76) | <0.0001       |<0.0001
| Definite fibrosis* | 0.64 (0.60, 0.69) | <0.0001       | 0.70 (0.65, 0.75) | <0.0001       | 0.75 (0.70, 0.79) | <0.0001       |0.004
| Possible UIP and UIP | 0.64 (0.61, 0.67) | <0.0001       | 0.64 (0.61, 0.68) | <0.0001       | 0.72 (0.69, 0.75) | <0.0001       |0.001
| **Non-Hispanic white subjects in COPDGene** |             |               |                                |               |                        |               |
| ILA               | 0.54 (0.52, 0.62)  | <0.0001       | 0.57 (0.55, 0.60) | <0.0001       | 0.58 (0.56, 0.61) | <0.0001       |0.20
| Subpleural and radiologic ILD | 0.60 (0.57, 0.63) | <0.0001       | 0.72 (0.69, 0.75) | <0.0001       | 0.75 (0.72, 0.78) | <0.0001       |0.0006
| Definite fibrosis* | 0.57 (0.52, 0.62) | <0.0001       | 0.76 (0.71, 0.80) | <0.0001       | 0.76 (0.72, 0.81) | <0.0001       |0.22
| Possible UIP and UIP | 0.60 (0.57, 0.63) | <0.0001       | 0.71 (0.68, 0.75) | <0.0001       | 0.75 (0.72, 0.78) | <0.0001       |0.0008
| **African-Americans in COPDGene** |             |               |                                |               |                        |               |
| ILA               | 0.51 (0.49, 0.52)  | 0.3           | 0.58 (0.54, 0.62) | 0.0006       | 0.59 (0.55, 0.62) | 0.0001       |0.52
| Subpleural and radiologic ILD | 0.52 (0.49, 0.55) | 0.1           | 0.67 (0.62, 0.74) | <0.0001       | 0.59 (0.55, 0.62) | <0.0001       |0.47
| Definite fibrosis* | 0.58 (0.50, 0.65) | 0.0005       | 0.67 (0.62, 0.74) | <0.0001       | 0.73 (0.62, 0.85) | 0.005        |0.34
| Possible UIP and UIP | 0.52 (0.49, 0.56) | 0.06          | 0.68 (0.63, 0.77) | <0.0001       | 0.70 (0.63, 0.76) | <0.0001       |0.50

* analyses using the MUC5B genotype were performed using additive genetic models; ‡ clinical data includes age, sex and pack-years of tobacco exposure; * definite fibrosis is evidence of pulmonary parenchymal architectural distortion. AGES: Age Gene/Environment Susceptibility; ILA: interstitial lung abnormality; ILD: interstitial lung disease; UIP: usual interstitial pneumonia.
Survival and the MUC5B promoter polymorphism

Finally, we sought to determine whether the MUC5B promoter polymorphism influenced survival amongst participants with ILA. Over a median follow-up time of 8.3 years (interquartile range (IQR) 4.8, 9.6) in AGES-Reykjavik, of the 378 participants with ILA, 210 (56%) died. Of the participants with ILA from COPDGene, with both mortality and genetic information available, over a median follow-up time of 5.4 years (IQR 4.6, 6.1), 59 (15%) of the 399 non-Hispanic white participants with ILA died, and 13 (8%) of the 165 African-Americans with ILA died (none of the African-American deaths occurred in participants with the MUC5B promoter polymorphism). No association was observed between the MUC5B minor allele and survival (hazard ratio (HR) 1.0, 95% CI 0.8–1.3; p=0.95; and HR 1.2, 95% CI 0.75–2.0; p=0.41, in the AGES-Reykjavik and in non-Hispanic white participants from COPDGene, respectively). Similar results were observed when ILA was subset to include only those with various ILA subtypes.

Discussion

This study adds important contributions to our understanding of the extent of the associations between ILA and the MUC5B promoter polymorphism (rs35705950), and the origins of pulmonary fibrosis in general. First, this study replicates the association between the MUC5B genotype and ILA [2]. Moreover, we found that the MUC5B promoter variant is associated with fibrotic ILA in African-Americans, a population with low allelic frequency of the rs35705950 polymorphism. Second, this study provides important information on the consistency of ILA subtypes that are associated with the MUC5B genotype; more specifically, it provides the evidence for consistent and minimal heterogeneity of associations between specific overlapping ILA subtypes (e.g. subpleural ILA and possible UIP). Finally, although our
study provides evidence that the MUC5B genotype, in addition to clinical characteristics, could help to improve risk prediction for important ILA subtypes (e.g., possible UIP or UIP patterns) in non-Hispanic white populations, it also demonstrates that the MUC5B genotype is unlikely to be useful in differentiating between individuals with ILA, who have a better or worse chance of survival.

The consistency of replication between the MUC5B genotype and ILA subtypes provides further support to the growing evidence [2] that some chest CT imaging patterns can reliably identify a common phenotype that shares a genetic background in patients with IPF [8, 15–17]. These findings, coupled with evidence that research participants with undiagnosed ILA are more likely to have physiologic decrements [1–4, 18], elevated fibrosis biomarkers [19, 20], and when followed over time can experience imaging progression [5], accelerated lung function decline [5] and an increased rate of mortality, [5, 7] all bolster the case that some subtypes of ILA likely represent an early and/or mild case of undiagnosed pulmonary fibrosis.

Although the pathogenic processes leading to pulmonary fibrosis that result from the MUC5B promoter polymorphism are not entirely understood, some steps in this process have been elucidated. Increasing copies of the minor allele of the MUC5B promoter polymorphism are associated with increased promoter activity [21], leading to an overall increased expression of MUC5B in the lung [8, 22], and specifically in the bronchiolar epithelium [21]. In IPF patients, increased expression of cilium-associated genes (including MUC5B) is associated with increased amounts of honeycombing [23]. Although it remains unclear how increased expression of MUC5B results in pulmonary fibrosis, our findings add to those noted in IPF patients, which demonstrate that increased MUC5B expression in the lung tends to result in a radiologic appearance dominated by subpleural reticular infiltrates and fibrosis both in patients with IPF [24] and in those with undiagnosed interstitial abnormalities.

To interpret these findings properly, it is important to consider the characteristics of the study populations. We previously demonstrated an association between the MUC5B genotype and ILA in a white population from the FHS [2]. The AGES-Reykjavik cohort, although similar to the FHS, in that it is also a general population sample of non-Hispanic white participants, is unique, in that it is comprised entirely of older adults from a geographically and genetically isolated population from Iceland [25, 26]. In contrast, COPDGene includes a population of smokers both with and without COPD, and excludes those known to have significant ILD. Consistent replication in these populations, and the minimal between-cohort heterogeneity seen with specific radiologic patterns, provides further evidence that the MUC5B promoter polymorphism confers a strong risk to develop a subpleural fibrotic process that can be detected in adults regardless of mitigating factors, such as differences in geography and smoking prevalence. The MUC5B promoter polymorphism, although relatively common in European and American populations (with at least one copy occurring in ~20% of Europeans and ~11% of Americans), is rare in African populations (~0.6%) [27]. The prevalence of at least one copy of the minor allele of the MUC5B promoter polymorphism in African-Americans (which has not previously been reported) from COPDGene was 4%. Additional studies will be needed to understand the unique factors that contribute to an ILA prevalence of 7% in this population.

In addition, our findings demonstrate that the MUC5B genotype improves risk prediction, particularly for detecting the presence of a possible UIP or UIP pattern among non-Hispanic white populations. This finding is remarkable, given the more modest improvements in risk that have been noted in multimarker genetic prediction models for established clinical disease entities, such as breast cancer [28] and cardiovascular events, [29] and the lack of evidence that multimarker genetic profiles can improve risk prediction for subclinical atherosclerosis [30]. Our findings suggest that the MUC5B genotype, in addition to important clinical variables, could be helpful in determining those individuals most likely to develop an early stage of pulmonary fibrosis. In contrast, our findings do not suggest that the MUC5B genotype will help to identify those with ILA who have an improved survival, as has been noted in patients with IPF [31]. Instead, our findings demonstrate that the MUC5B genotype is important, but not the only factor that can increase the risk for ILA (and ILA progression) [5], which when present, can lead to an increased rate of mortality [7].

This study has several limitations. First, although we were able to demonstrate similar associations between the MUC5B promoter polymorphism and specific radiologic subtypes of ILA in African-Americans from COPDGene, the smaller sample size and lower prevalence of the minor allele, might have limited the statistical power to demonstrate an improvement in risk prediction. Second, although the MUC5B genotype is associated with a possible UIP pattern across all populations, the magnitude of this association was less than that observed in patients with clinically identified IPF [8, 16]. Finally, we cannot rule out the possibility that the small sample size, within some ILA subtypes specifically, could have limited our statistical power to detect an association between the MUC5B genotype and survival in subgroup analyses.
In conclusion, our study demonstrates that the MUC5B promoter polymorphism is associated with undiagnosed chest CT findings that are consistent with an early stage of pulmonary fibrosis. Our study also provides some specificity for these associations by demonstrating that the MUC5B genotype is associated with subpleural ILA and a possible UIP pattern, but not with predominantly centrilobular abnormalities. In addition, the MUC5B genotype might help to predict the presence of specific subtypes of ILA on chest CT scans. Although it is not known whether the treatment of early stages of pulmonary fibrosis will help to prevent an accelerated decline in pulmonary function [5] and early mortality [7] with which they are associated, the fact that the MUC5B genotype could improve risk detection for a possible UIP pattern suggests a path forward.

Acknowledgements


References


