



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

Rachel K. Putman^{1,12}, Gunnar Gudmundsson^{2,12}, Tetsuro Araki^{3,4}, Mizuki Nishino^{3,4}, Sigurdur Sigurdsson⁵, Elías F. Gudmundsson⁵, Gudny Eiríksdóttir⁵, Thor Aspelund^{5,6}, James C. Ross^{7,8}, Raúl San José Estépar^{3,8}, Ezra R. Miller¹, Yoshitake Yamada^{3,4}, Masahiro Yanagawa⁹, Noriyuki Tomiyama⁹, Lenore J. Launer¹⁰, Tamara B. Harris¹⁰, Souheil El-Chemaly¹, Benjamin A. Raby^{1,7}, Michael H. Cho^{1,7}, Ivan O. Rosas¹, George R. Washko^{1,4}, David A. Schwartz¹¹, Edwin K. Silverman^{1,7}, Vilundur Gudnason^{5,6}, Hiroto Hatabu^{3,4} and Gary M. Hunninghake^{1,4}

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, Landspítali 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>

Cite this article as: Putman RK, Gudmundsson G, Araki T, *et al.* The *MUC5B* promoter polymorphism is associated with specific interstitial lung abnormality subtypes. *Eur Respir J* 2017; 50: 1700537 [<https://doi.org/10.1183/13993003.00537-2017>].

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, $p=1\times 10^{-26}$), there was evidence of significant heterogeneity between cohorts ($I^2=81\%$). 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, $p=1\times 10^{-30}$) with minimal heterogeneity ($I^2=0\%$). 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.

The content of this work is not subject to copyright. Design and branding are copyright ©ERS 2017.

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 HL007633. G. Gudmundsson is supported by the Oddur Olafsson Fund, project grant 141513-051 from the Icelandic Research Fund and Landspítali 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 HL007633. Y. Yamada is supported by project grant 141513-051 from the Icelandic Research Fund and Landspítali 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 I01 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: 27120120022C, NIH contracts N01-AG-1-2100 and HHSN27120120022C, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association) and the Althingi (the Icelandic Parliament). V. Gudnason is supported by NIA grant: 27120120022C and project grant 141513-051 from the Icelandic Research Fund. G.M. Hunninghake 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/JRS/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. I^2 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).

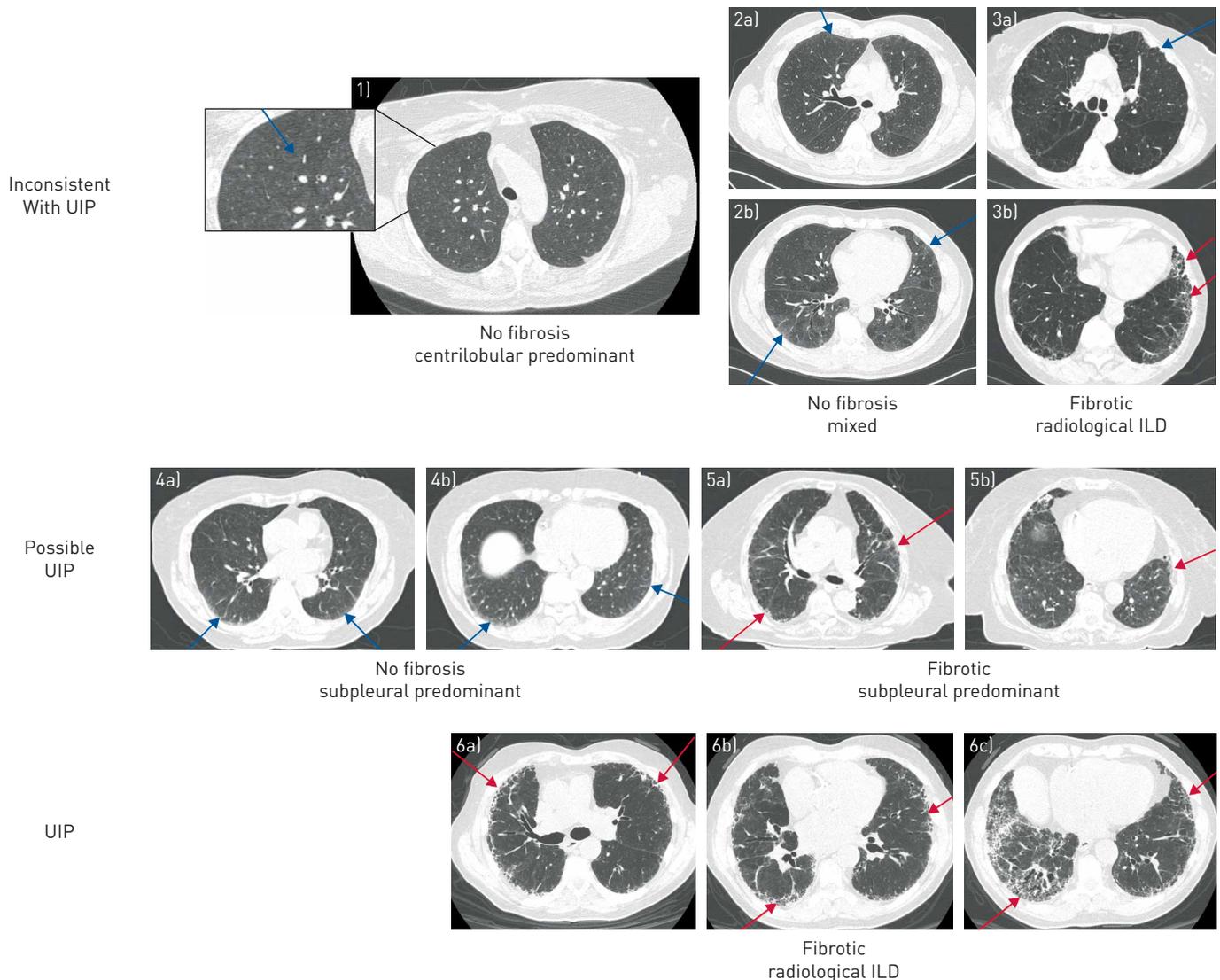


FIGURE 1 Chest computed tomographic (CT) images depicting radiologic subtypes and the overlap between subtypes of interstitial lung abnormalities (ILA). In all panels the blue arrows point to areas of ILA without fibrosis, the red arrows point to areas of ILA with fibrosis, each panel (1–6) represents 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.

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

TABLE 1 Baseline characteristics of participants in AGES-Reykjavik and COPDGene stratified by interstitial lung abnormality (ILA) status and race[#]

	AGES-Reykjavik		p-value	COPDGene (non-Hispanic, white)		p-value	COPDGene (African-American)		p-value
	No ILA	ILA		No ILA	ILA		No ILA	ILA	
Subjects n	3209	377		3667	485		1728	223	
Age years	76±5	78±6	<0.0001	61±9	64±10	<0.0001	54±7	55±8	0.003
Females	1905 (59)	171 (45)	<0.0001	1752 (48)	219 (45)	0.29	705 (41)	126 (57)	<0.0001
Body mass index	27±4	27±5	0.50	29±6	29±6	0.25	29±7	29±7	0.33
Pack-years	2 [#] [0, 25]	20 [#] [0, 50]	<0.0001	40 (29, 56)	45 (34, 63)	<0.0001	34 (22, 46)	35 (24, 47)	0.49
Current smokers	373 (12)	69 (18)	0.0003	1402 (38)	2546 (53)	<0.0001	1393 (81)	183 (82)	0.65
History of COPD				1474 (40)	167 (35)	0.02	381 (22)	54 (25)	0.44
Percentage of the lung less than -950 Hounsfield units				3 (0.9, 9)	1.4 (0.5, 4.8)	<0.0001	1.1 (0.4, 3.3)	0.7 (0.2, 2.4)	<0.0001

Data are presented as mean±SD, n (%) and median (interquartile range), unless otherwise stated. [#]: Includes cigarette and cigar smoking. AGES: Age Gene/Environment Susceptibility; COPD: chronic obstructive pulmonary disease.

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 *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 *MUC5B* genotype was observed to be strongly associated with ILA (OR 2.1, 95% CI 1.8–2.4; $p=1\times 10^{-26}$), despite significant heterogeneity between cohorts ($I^2=81%$) (table 2).

The *MUC5B* promoter polymorphism and radiologic patterns of ILA

Whereas there was some variability in associations between the *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 *MUC5B* genotype were consistently associated with definite fibrosis (OR 3.0, 95% CI 2.4–3.7; $p=8\times 10^{-22}$) compared to those without ILA, table 2, figure 1. There was also evidence that in addition to consistent association (or lack thereof) with the *MUC5B* genotype, when narrowed to specific radiologic phenotypes,

TABLE 2 Association between interstitial lung abnormalities (ILAs) and the *MUC5B* promoter polymorphism[#]

ILA subtype	AGES-Reykjavik		COPDGene (non-Hispanic, white)		COPDGene (African-American)		I ²⁺	Cohorts combined	
	Adjusted [¶] OR (95% CI)	p-value	Adjusted [¶] OR (95% CI)	p-value	Adjusted [¶] OR (95% CI)	p-value		Adjusted [¶] OR (95% CI)	p-value [§] (corrected)
ILA	2.7 (2.2, 3.2)	1×10^{-22}	1.6 (1.3, 2.0)	2×10^{-6}	1.5 (0.8, 2.8)	0.19	81%	2.1 (1.8, 2.4)	1×10^{-26}
ILA without fibrosis ^f	2.3 (1.8, 2.9)	5×10^{-13}	1.5 (1.2, 1.9)	0.0004	0.99 (0.5, 2.2)	0.97	76%	1.8 (1.5, 2.1)	2×10^{-12}
Definite fibrosis	3.3 (2.4, 4.4)	5×10^{-15}	2.1 (1.4, 3.1)	0.0001	6.2 (2.3, 16.8)	0.0003	59%	3.0 (2.4, 3.7)	8×10^{-22}
Centrilobular	1.3 (0.7, 2.5)	0.44	0.64 (0.4, 1.1)	0.09	1.2 (0.4, 3.8)	0.79	15%	0.91 (0.63, 1.3)	1.0
Subpleural	3.0 (2.3, 3.7)	2×10^{-19}	2.4 (1.9, 3.2)	5×10^{-11}	2.3 (0.98, 5.4)	0.06	0%	2.6 (2.2, 3.1)	1×10^{-30}
Mixed	2.3 (1.6, 3.3)	2×10^{-5}	1.05 (0.6, 1.7)	0.86	1.06 (0.3, 3.4)	0.93	66%	1.5 (1.1, 2.0)	0.05
Radiologic ILD ^{##}	4.4 (2.2, 9.0)	4×10^{-5}	4.8 (2.5, 8.9)	1×10^{-6}			0%	4.4 (2.8, 6.8)	5×10^{-10}
Inconsistent with UIP ^{¶¶}	2.2 (1.7, 3.0)	1×10^{-7}	0.95 (0.7, 1.3)	0.75	1.0 (0.4, 2.4)	0.94	55%	1.4 (1.1, 1.7)	0.01
Possible UIP	2.8 (2.2, 3.6)	9×10^{-17}	2.5 (1.9, 3.3)	1×10^{-11}	2.7 (1.1, 6.3)	0.03	1%	2.7 (2.3, 3.2)	1×10^{-30}
UIP	4.4 (2.2, 9.0)	4×10^{-5}	4.6 (2.0, 10.4)	0.0003			0%	4.1 (2.1, 8.1)	0.0003

[#]: analyses of the *MUC5B* genotype were performed using additive genetic models; odds ratios are per copy of the *MUC5B* minor allele; [¶]: models are adjusted for age, sex and tobacco exposure; ⁺: I^2 calculated using the DerSimonian and Laird method; [§]: p-value corrected for multiple testing using Bonferroni correction; ^f: 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 *MUC5B* genotype; ^{¶¶}: analysis was not done in African-Americans; no participants with usual interstitial pneumonia (UIP) had the *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; $p=1\times 10^{-30}$), with essentially no between-cohort heterogeneity ($I^2=1\%$), (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; $p=1.0$, $I^2=15\%$), (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[#]

	<i>MUC5B</i> minor allele		Clinical data [¶]		Clinical data+ <i>MUC5B</i> minor allele		p-value for the comparison of clinical with Clinical+ <i>MUC5B</i>
	C-statistic (95% CI)	p-value	C-statistic (95% CI)	p-value	C-statistic (95% CI)	p-value	
AGES-Reykjavik							
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	0.69 [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.

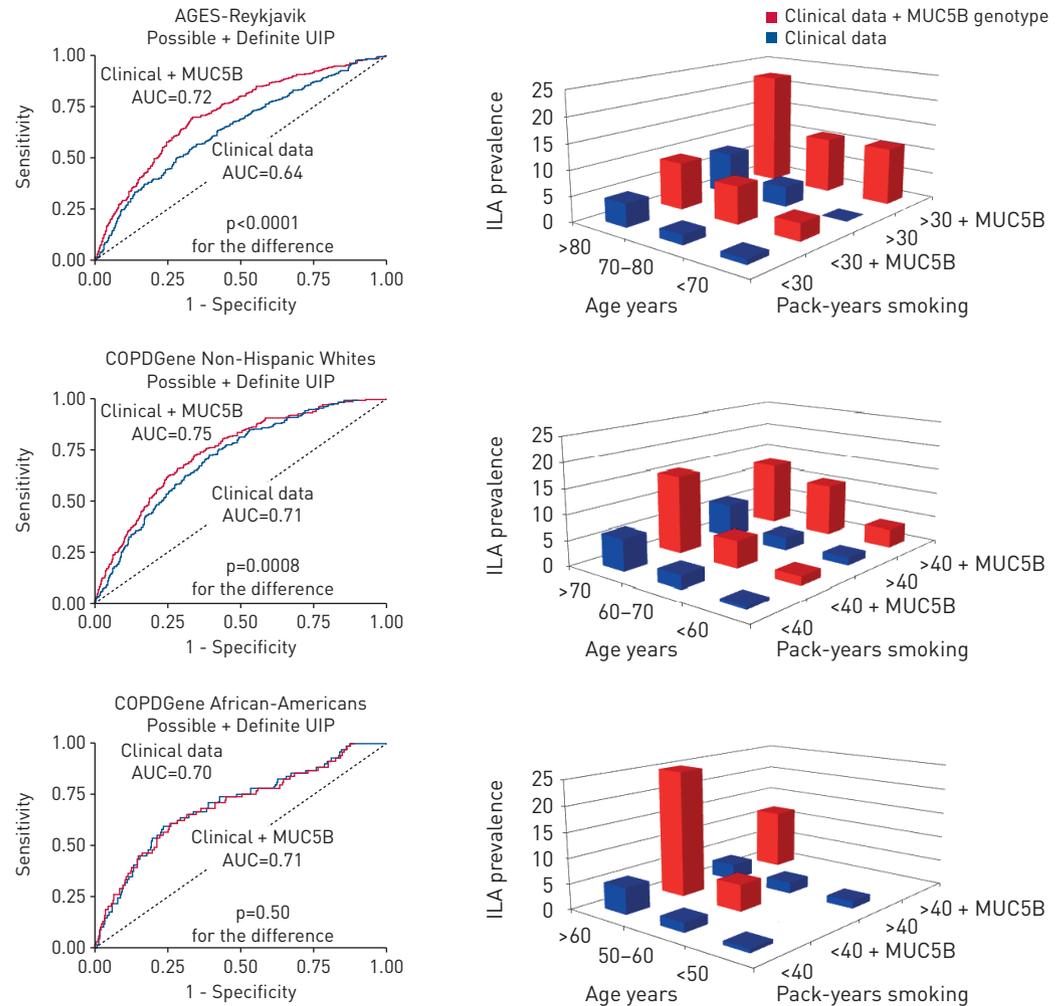


FIGURE 2 Receiver operating characteristic curves and bar graphs depicting interstitial lung abnormality (ILA), specifically possible and definite usual interstitial pneumonia (UIP) prediction, using baseline clinical information (age, sex and pack-years of smoking) and then with the addition of the *MUC5B* promoter polymorphism in each cohort. In the bar graphs, the addition of the *MUC5B* minor allele applies for at least one copy of the minor allele.

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

Author contributions are as follows. Study design: G. Gudmundsson, V. Gudnason, H. Hatabu, G.M. Hunninghake, I.O. Rosas, D.A. Schwartz, E.K. Silverman, G.R. Washko. Acquisition, analysis or interpretation of the data: T. Araki, T. Aspelund, M.H. Cho, S. El-Chemaly, G. Eiriksdottir, E. Gudmundsson, G. Gudmundsson, V. Gudnason, H. Hatabu, G.M. Hunninghake, M. Nishino, R.K. Putman, J.C. Ross, R. San José Estépar, D.A. Schwartz, S. Sigurdsson, E.K. Silverman, G.R. Washko, Y. Yamada, M. Yanagawa. Critical revision of the manuscript for important intellectual content: T. Araki, T. Aspelund, M.H. Cho, S. El-Chemaly, G. Eiriksdottir, E. Gudmundsson, G. Gudmundsson, V. Gudnason, T.B. Harris, H. Hatabu, G.M. Hunninghake, L.J. Launer, E.R. Miller, M. Nishino, R.K. Putman, B.A. Raby, I.O. Rosas, J.C. Ross, R. San José Estépar, D.A. Schwartz, S. Sigurdsson, E.K. Silverman, N. Tomiyama, G.R. Washko, Y. Yamada, M. Yanagawa. Statistical analysis: G.M. Hunninghake and R.K. Putman. Obtained funding: G. Gudmundsson, V. Gudnason, G.M. Hunninghake, E.K. Silverman

References

- 1 Washko GR, Hunninghake GM, Fernandez IE, *et al.* Lung volumes and emphysema in smokers with interstitial lung abnormalities. *N Engl J Med* 2011; 364: 897–906.
- 2 Hunninghake GM, Hatabu H, Okajima Y, *et al.* *MUC5B* promoter polymorphism and interstitial lung abnormalities. *N Engl J Med* 2013; 368: 2192–2200.
- 3 Lederer DJ, Enright PL, Kawut SM, *et al.* Cigarette smoking is associated with subclinical parenchymal lung disease: the Multi-Ethnic Study of Atherosclerosis (MESA)-lung study. *Am J Respir Crit Care Med* 2009; 180: 407–414.
- 4 Tsushima K, Sone S, Yoshikawa S, *et al.* The radiological patterns of interstitial change at an early phase: over a 4-year follow-up. *Respir Med* 2010; 104: 1712–1721.
- 5 Araki T, Putman RK, Hatabu H, *et al.* Development and progression of interstitial lung abnormalities in the Framingham Heart Study. *Am J Respir Crit Care Med* 2016; 194: 1514–1522.
- 6 Doyle TJ, Washko GR, Fernandez IE, *et al.* Interstitial lung abnormalities and reduced exercise capacity. *Am J Respir Crit Care Med* 2012; 185: 756–762.
- 7 Putman RK, Hatabu H, Araki T, *et al.* Association between interstitial lung abnormalities and all-cause mortality. *JAMA* 2016; 315: 672–681.
- 8 Seibold MA, Wise AL, Speer MC, *et al.* A common *MUC5B* promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011; 364: 1503–1512.
- 9 Harris TB, Launer LJ, Eiriksdottir G, *et al.* Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol* 2007; 165: 1076–1087.
- 10 Regan EA, Hokanson JE, Murphy JR, *et al.* Genetic epidemiology of COPD (COPDGene) study design. *COPD* 2010; 7: 32–43.
- 11 Washko GR, Lynch DA, Matsuo S, *et al.* Identification of early interstitial lung disease in smokers from the COPDGene Study. *Acad Radiol* 2010; 17: 48–53.
- 12 Raghu G, Collard HR, Egan JJ, *et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; 183: 788–824.
- 13 Ross JC, Estepar RS, Diaz A, *et al.* Lung extraction, lobe segmentation and hierarchical region assessment for quantitative analysis on high resolution computed tomography images. *Med Image Comput Comput Assist Interv* 2009; 12: 690–698.
- 14 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177–188.
- 15 Noth I, Zhang Y, Ma SF, *et al.* Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013; 1: 309–317.
- 16 Putman RK, Rosas IO, Hunninghake GM. Genetics and early detection in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; 189: 770–778.
- 17 Fingerlin TE, Murphy E, Zhang W, *et al.* Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013; 45: 613–620.
- 18 Doyle TW, Washko GR, Fernandez IE, *et al.* Interstitial lung abnormalities and reduced exercise capacity. *Am J Respir Crit Care Med* 2012; 185: 756–762.
- 19 Podolanczuk AJ, Oelsner EC, Barr RG, *et al.* High attenuation areas on chest computed tomography in community-dwelling adults: the MESA study. *Eur Respir J* 2016; 48: 1442–1452.
- 20 Ho JE, Gao W, Levy D, *et al.* Galectin-3 is associated with restrictive lung disease and interstitial lung abnormalities. *Am J Respir Crit Care Med* 2016; 194: 77–83.
- 21 Nakano Y, Yang IV, Walts AD, *et al.* *MUC5B* Promoter Variant rs35705950 affects *MUC5B* expression in the distal airways in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016; 193: 464–466.
- 22 Kropski JA, Pritchett JM, Zoz DF, *et al.* Extensive phenotyping of individuals at risk for familial interstitial pneumonia reveals clues to the pathogenesis of interstitial lung disease. *Am J Respir Crit Care Med* 2015; 191: 417–426.
- 23 Yang IV, Coldren CD, Leach SM, *et al.* Expression of cilium-associated genes defines novel molecular subtypes of idiopathic pulmonary fibrosis. *Thorax* 2013; 68: 1114–1121.

- 24 Chung JH, Peljto AL, Chawla A, *et al.* CT imaging phenotypes of pulmonary fibrosis in the MUC5B promoter site polymorphism. *Chest* 2016; 149: 1215–1222.
- 25 Taylor R, Bryant J, Gudnason V, *et al.* A study of familial hypercholesterolaemia in Iceland using RFLPs. *J Med Genet* 1989; 26: 494–498.
- 26 Heutink P, Oostra BA. Gene finding in genetically isolated populations. *Hum Mol Genet* 2002; 11: 2507–2515.
- 27 Online Mendelian Inheritance in Man. Mucin 5 sB, tracheobronchial: MUC5B. OMIM no. 600770. http://useastensembl.org/Homo_sapiens/Variation/Population?db=core;r=11:1219491-1220491;v=rs35705950;vdb=variation;vf=10314659 Date last accessed: April 30, 2017. Date last updated: February 4, 2014.
- 28 Wacholder S, Hartge P, Prentice R, *et al.* Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 2010; 362: 986–993.
- 29 Kathiresan S, Melander O, Anevski D, *et al.* Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 2008; 358: 1240–1249.
- 30 Hernesniemi JA, Seppala I, Lyytikainen LP, *et al.* Genetic profiling using genome-wide significant coronary artery disease risk variants does not improve the prediction of subclinical atherosclerosis: the Cardiovascular Risk in Young Finns Study, the Bogalusa Heart Study and the Health 2000 Survey--a meta-analysis of three independent studies. *PLoS One* 2012; 7: e28931.
- 31 Peljto AL, Zhang Y, Fingerlin TE, *et al.* Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA* 2013; 309: 2232–2239.