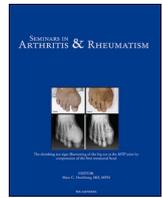




Contents lists available at ScienceDirect

Seminars in Arthritis and Rheumatism

journal homepage: www.elsevier.com/locate/semarthrit

Genetic, social, and environmental risk factors in rheumatoid arthritis-associated interstitial lung disease

Austin M. Wheeler^{a,b}, Joshua F. Baker^c, Jill A. Poole^b, Dana P. Ascherman^d, Yangyuna Yang^b, Gail S. Kerr^e, Andreas Reimold^f, Gary Kunkel^g, Grant W. Cannon^g, Katherine D. Wysham^h, Namrata Singh^h, Deana Lazaroⁱ, Paul Monach^j, S. Louis Bridges Jr.^k, Ted R. Mikuls^{a,b}, Bryant R. England^{a,b,*}

^a VA Nebraska-Western Iowa Health Care System, Omaha, NE, USA^b University of Nebraska Medical Center, Omaha, NE, USA^c University of Pennsylvania and VA Medical Center, Philadelphia, PA, USA^d Pittsburgh VA & University of Pittsburgh, Pittsburgh, PA, USA^e Washington D.C. VA, Howard University, & Georgetown University, Washington D.C., USA^f Dallas VA & University of Texas Southwestern, Dallas, TX, USA^g VA Salt Lake City & University of Utah, Salt Lake City, UT, USA^h VA Puget Sound Health Care System & University of Washington, Seattle, WA, USAⁱ Brooklyn VA, Brooklyn, NY, USA^j Boston VA, Boston, MA, USA^k Hospital for Special Surgery, New York, NY, USA

ARTICLE INFO

Keywords:

Rheumatoid arthritis

Interstitial lung disease

Single nucleotide polymorphism

ABSTRACT

Objective: *MUC5B* and *TOLLIP* single nucleotide polymorphisms (SNPs) and cigarette smoking were associated with rheumatoid arthritis-interstitial lung disease (RA-ILD) in a predominantly Northern European population. We evaluated whether RA-ILD is associated with these genetic variants and HLA-DRB1 shared epitope (SE) alleles in a large RA cohort stratified by race and smoking history.

Methods: HLA-DRB1 SE alleles and *MUC5B* rs35705950 and *TOLLIP* rs5743890 SNPs were genotyped in U.S. veterans with RA. ILD was validated through medical record review. Genetic associations with ILD were assessed in logistic regression models overall and in subgroups defined by race and smoking status, with additive interactions assessed by the relative excess risk of interaction (RERI).

Results: Of 2,556 participants (88% male, 77% White), 238 (9.3%) had ILD. The *MUC5B* variant was associated with ILD (OR 2.25 [95% CI 1.69, 3.02]), whereas *TOLLIP* and HLA-DRB1 SE were not. The *MUC5B* variant was less frequent among Black/African American participants (5.8% vs. 22.6%), though its association with RA-ILD was numerically stronger (OR 4.23 [1.65, 10.86]) compared to all other participants (OR 2.32 [1.70, 3.16]). Those with the *MUC5B* variant and a smoking history had numerically higher odds of ILD (OR 4.18 [2.53, 6.93]) than non-smokers (OR 2.41 [1.16, 5.04]). Additive interactions between *MUC5B*-race and *MUC5B*-smoking were not statistically significant.

Conclusion: In this large RA cohort, the *MUC5B* promoter variant was associated with >2-fold higher odds of RA-ILD. While this variant is less common among Black/African American patients, its presence in this population carried >4-fold higher odds of RA-ILD.

Rheumatoid arthritis-associated interstitial lung disease (RA-ILD) affects approximately 10% of people with RA and causes substantial morbidity and premature mortality [1,2]. Despite the contribution of ILD to reduced quality of life and survival in RA, very few risk factors

have been identified. To date, these include older age, male sex, cigarette smoking, and indicators of severe articular disease [3,4].

Unlike ILD that accompanies other connective tissue diseases, usual interstitial pneumonia (UIP) is the most common ILD pattern in RA [2,

* Corresponding author at: 986270 Nebraska Medical Center, Omaha, NE 68198-6270.

E-mail address: bryant.england@unmc.edu (B.R. England).

<https://doi.org/10.1016/j.semarthrit.2022.152098>

5]. This pattern is shared with idiopathic pulmonary fibrosis (IPF), a chronic, progressive, fibrotic ILD of unknown cause [6]. Given histo-radiologic overlap between RA-ILD and IPF, recent studies have investigated whether IPF genetic risk variants are also associated with RA-ILD. A large international study demonstrated a striking overlap of genetic risk factors, specifically the *MUC5B* promoter variant rs35705950, between IPF and RA-ILD [7,8]. A prospective study of a smaller, Northern European population showed a similar *MUC5B* and RA-ILD association [9]. The *MUC5B* gene encodes for the mucin 5B protein, which is normally expressed in the bronchial epithelium and contributes to the viscosity of mucus for mucociliary clearance. This variant is associated with increased expression of the mucin 5B protein in lung parenchyma, specifically in the metaplastic epithelia lining honeycomb cysts and mucus within these cysts in RA-ILD patients [7]. The frequency of the *MUC5B* promoter variant is highly variable across racial and ethnic groups, although the differential risk of ILD related to this variant among these groups, specifically among the Black/African American population, has not been comprehensively assessed. As race represents a social construct, a differential impact among racial groups may reveal the influence of social determinants of health and indicate important health disparities. Among other genetic variants associated with IPF, the *TOLLIP* rs5743890 variant has also been associated with the presence of ILD in RA patients [7,10–12].

Cigarette smoking has consistently been demonstrated to be the strongest environmental risk factor for the development of seropositive RA, particularly among those with HLA-DRB1 alleles encoding the shared epitope (SE) [13–15]. In addition to associations with RA susceptibility, smoking appears to contribute to the development of extra-articular disease, including ILD [1,16]. Among RA-related characteristics, higher disease activity portends a higher risk of developing incident RA-ILD [17]. Whether interactions between the aforementioned genetic and environmental factors influence risk of the development of RA-ILD is unknown.

In this study, we sought to determine the associations of the *MUC5B* rs35705950 and *TOLLIP* rs5743890 SNPs, and HLA-DRB1 SE alleles with ILD in a large RA cohort. Additionally, we aimed to examine whether these genetic factors are differentially associated with ILD in the Black/African American population with RA, a group under-represented in prior work. Finally, we aimed to determine whether there was evidence of gene-environment interactions between these variants and other established ILD risk factors, including cigarette smoking and RA disease activity. We hypothesized that both the *MUC5B* and *TOLLIP* variants are associated with RA-ILD and influenced by other risk factors.

Methods

Study design and population

We studied participants with RA enrolled in the Veterans Affairs Rheumatoid Arthritis (VARA) Registry, a multicenter, longitudinal, observational cohort of U.S. veterans [18], all meeting the 1987 American College of Rheumatology classification criteria for RA [19]. All participants provided written informed consent prior to enrollment. Each site received institutional review board approval, and this study was approved by the VA Nebraska-Western Iowa Health Care System IRB (1,576,193). This study was also approved by the VARA Scientific Ethics Advisory Committee.

Demographic and clinical data

Demographic variables collected at registry enrollment include sex, self-reported race and ethnicity, and cigarette smoking history (categorized as ever vs. never smoking history). Race and ethnicity data were collected for this study due to the differences in frequency of the genetic variants of interest among various populations [7]. Participants self-identified as White, Black/African American, Asian American,

American Indian/Pacific Islander, or Hispanic. These terms were not defined, and participants were given the option of not identifying with a particular race or ethnicity. A small proportion of patients ($n = 118$; 4.6%) reported race and ethnicity as unknown or chose not to respond. RA disease duration was collected at the time of registry enrollment and subsequently calculated at the time of ILD diagnosis. Body mass index (BMI) was calculated as the mean value over all clinical encounters with available measurements to minimize variability and was categorized as underweight ($<20 \text{ kg/m}^2$), normal (≥ 20 and $<25 \text{ kg/m}^2$), overweight (≥ 25 and $<30 \text{ kg/m}^2$), and obese ($\geq 30 \text{ kg/m}^2$). Anti-cyclic citrullinated peptide antibody (anti-CCP, positivity $>5 \text{ U/mL}$) was measured using banked serum collected at enrollment, as previously described [18]. Disease Activity Score in 28 joints with C-reactive protein (DAS28-CRP) was assessed and recorded at routine rheumatology care visits following VARA enrollment. For clinical disease activity assessments with missing CRP values, erythrocyte sedimentation rate (ESR) was used to calculate DAS28-ESR. The mean DAS28 (CRP or ESR) score over all visits was calculated for each patient to account for cumulative effects and then categorized as remission or low disease activity (≤ 3.2) vs. moderate to high disease activity (>3.2) [20].

Genotyping

HLA-DRB1 SE genotyping was performed on all study participants using banked enrollment samples via DNA sequencing of exon 2 using the AlleleSEQR HLA-DRB1 reagent kit and protocol (Abbott Molecular, Abbott Park, IL, USA) or a PCR based, sequence specific oligonucleotide probe system, as previously described [21]. HLA-DRB1 SE status was analyzed based on the presence of one or two copies of the SE allele versus none.

Genotyping for an additional 665,608 SNPs was performed using the Infinium Global Screening Array-24 v2.0 microarray (Illumina, Inc.; San Diego, CA, USA) [22]. From this array, the *MUC5B* rs35705950 and *TOLLIP* rs5743890 SNPs were extracted and analyzed assuming autosomal dominant inheritance. The wild type allele for *MUC5B* rs35705950 is defined by the guanine (G) nucleotide residue, while the variant allele is defined by the nucleotide thymine (T) [7]. For *TOLLIP* rs5743890, the wildtype and variant alleles were defined by the T and cytosine (C) nucleotide residues, respectively.

Lung disease assessment

ILD was identified through a screening and validation process that included detailed, systematic medical record confirmation of ILD diagnoses [23–25]. Participants were classified as having RA-ILD if, through medical record review, they had a provider diagnosis of ILD and either imaging findings or a lung biopsy consistent with ILD. In addition to confirming ILD classification based on provider diagnoses in the medical records and imaging findings consistent with ILD, the date of clinical diagnosis and ILD pattern based on CT or pathology reports, when available, were extracted. Patterns identified included usual interstitial pneumonia (UIP), non-specific interstitial pneumonia (NSIP), and other. Chronic obstructive pulmonary disease (COPD) or bronchiectasis diagnoses were identified via the presence of at least two International Classification of Diseases (ICD) diagnostic codes from inpatient or outpatient encounters using linked administrative data from the VA Corporate Data Warehouse according to Healthcare Cost and Utilization Project Clinical Classification Software categorization, as used in a prior study [26].

Statistical analysis

The frequency of each SNP genotype and HLA-DRB1 SE positivity was calculated for the overall cohort and in subgroups defined by ILD status and self-reported race and ethnicity. Associations of *MUC5B* rs35705950, *TOLLIP* rs5743890, and HLA-DRB1 SE alleles with RA-ILD

were quantified via logistic regression in the overall study population and then among subgroups. In subgroup analyses, race and ethnicity were categorized as Black/African American or all other racial and ethnic groups (summarized as “non-Black/African American”), since the frequency of the genetic variants was similar across White, Asian American, and American Indian/Pacific Islander, and Hispanic participants with ILD. The combined association of race and genetic variants with ILD was assessed using logistic regression with a referent group of non-Black/African American participants without the genetic variant. To assess for the presence of additive interaction between race and genotypes, we calculated the relative excess risk of interaction (RERI) and 95% confidence interval (CI) by bootstrapping [27]. This approach was additionally used to examine additive interactions between the genotypes and cigarette smoking history (ever vs. never), as well as the genotypes and RA disease activity (mean DAS28-CRP of moderate-to-high vs. remission/low).

A secondary analysis was performed to assess for absence of an association between *MUC5B* and other types of lung disease (COPD/bronchiectasis). In this analysis, COPD/bronchiectasis was evaluated as the outcome and patients with RA-ILD were excluded, as COPD/bronchiectasis and ILD may co-occur. Additional secondary analyses were performed among the entire study cohort to examine the associations of *MUC5B* with specific ILD patterns and RA duration at ILD onset (dichotomized at the median value, 7 years). RA duration at the time of ILD onset was also compared between RA-ILD patients with and without the *MUC5B* variant using an independent *t*-test. Data analyses were performed using Stata version 17 (StataCorp LLC., College Station, TX).

Results

Patient characteristics

Among 2561 patients with RA, 238 (9.3%) had ILD. ILD pattern was available for 102 patients, with UIP most frequent (68.6%). Enrollment characteristics of the study population by ILD status are summarized in Table 1. Consistent with the VA population, patients were predominantly male (89%), and older (mean 69.4 years). A majority of the study population self-identified as White ($n = 1962$, 77%), with 417 (16%) self-identifying as Black/African American. Compared to RA patients without ILD, RA patients with ILD were slightly older (72.0 vs. 69.2 years), more often male (95% vs. 88%), and were more likely to have a smoking history (86% vs. 77%), anti-CCP antibody seropositivity (84% vs. 77%), and a higher DAS28 score (3.48 vs. 3.21).

Variant allele frequencies and association with RA-ILD

Frequencies of the *MUC5B* rs35705950 and *TOLLIP* rs5743890 variants, as well as the presence of HLA-DRB1 SE alleles among ILD and non-ILD RA patients are shown in Table 2. The *MUC5B* rs35705950 promoter variant was significantly more frequent in patients with RA-ILD as compared to those without ILD (32.7% vs. 17.8%; OR 2.25 [95% CI 1.69, 3.02]). In contrast, frequencies of the *TOLLIP* rs5743890 variant (21.0% vs. 22.0%; OR 0.94 [95% CI 0.68, 1.31]) and HLA-DRB1 SE positivity (68.3% vs. 68.2%; OR 1.01 [95% CI 0.75, 1.36]) were similar in those with and without ILD.

Genetic associations with RA-ILD by race

The frequency of the *MUC5B* rs35705950 variant was lower in Black/African American participants (5.8%) overall compared to White participants (22.6%) or other racial and ethnic groups (13.0%). There was a similar variability in the frequency of the *TOLLIP* rs5743890 variant among the Black/African American (3.4%), White (26.3%), and other (17.5%) subgroups. While the HLA-DRB1 SE allele was much more common in the overall population, the frequency again differed between Black/African American (38.9%), White (75.3%), and other (58.6%)

Table 1

Characteristics of study population at registry enrollment by interstitial lung disease (ILD) status.

	N	Total (N = 2561)	RA-ILD (N = 238)	RA no- ILD (N = 2323)	p value
Age, years	2560	69.4 (10.8)	72.0 (9.0)	69.2 (10.9)	<0.001
Male sex, n (%)	2557	2275 (89.0)	225 (94.5)	2050 (88.4)	0.004
Race and ethnicity, n (%)	2556				0.36
Black/African American		417 (16.3)	44 (18.5) (16.1)	373 (16.1)	
White		1962 (76.7)	174 (73.1)	1788 (77.1)	
Other*		177 (6.9)	20 (8.4)	157 (6.8)	
BMI category, n (%)	2545				0.73
Underweight (<20 kg/m ²)		84 (3.3)	6 (2.5)	78 (3.4)	
Normal (20–24.9 kg/m ²)		479 (18.8)	50 (21.2) (18.6)	429 (18.6)	
Overweight (25–29.9 kg/m ²)		920 (36.1)	84 (35.6) (36.2)	836 (36.2)	
Obese (>30 kg/m ²)		1062 (41.7)	96 (40.7) (41.8)	966 (41.8)	
Smoking history (ever), n (%)	2560	1989 (77.7)	204 (85.7)	1785 (76.9)	0.002
RA disease duration, years	2507	10.1 (11.2)	11.1 (12.1)	10.0 (11.2)	0.42
Anti-CCP positive, n (%)	2419	1874 (77.5)	192 (84.2)	1682 (76.8)	0.01
DAS28	2522	3.23 (1.08)	3.48 (1.10)	3.21 (1.07)	<0.001

Values mean (SD) unless otherwise noted. P values by chi-square or independent *t*-test.

BMI = body mass index; Anti-CCP = anti-cyclic citrullinated protein antibody; DAS-28 = disease activity score (28 joints).

* Other includes Asian American, American Indian/Pacific Islander, and Hispanic.

subgroups. When limiting to patients with ILD, the frequency of the *MUC5B* variant was similar in the White (37.3%) and other racial and ethnic subgroups (35.0%) but was notably lower in the Black/African American subgroup (13.7%).

Associations of genetic variants and race with ILD are shown in Table 3. Among those without the *MUC5B* variant, the odds of ILD were not significantly higher among Black/African American participants compared to non-Black/African American participants (OR 1.36 [95% CI 0.93, 1.99]). The odds of ILD were increased among both non-Black/African American (OR 2.32 [95% CI 1.70, 3.16]) and Black/African American (OR 4.23 [95% CI 1.65, 10.86]) individuals with the *MUC5B* variant compared to non-Black/African American individuals without the *MUC5B* variant. While the RERI suggested the possibility of an additive interaction between the *MUC5B* variant and race, this was not statistically significant (RERI 1.56 [95% CI -1.34, 7.54]). The odds of ILD did not appear to differ by race based on the presence of either the *TOLLIP* variant or HLA-DRB1 SE positivity (Table 3).

Genetic associations with RA-ILD by smoking status

The frequency of ILD according to the combinations of smoking status and genotypes are shown in Table 4. Referent to never smokers without the *MUC5B* rs35705950 variant, the presence of this variant was associated with ILD among never smokers (OR 2.41 [95% CI 1.16, 5.04]). Using the same referent group, ever smoking was associated with ILD in *MUC5B* negative participants (OR 1.87 [95% CI 1.18, 2.97]). The highest odds of ILD occurred in participants with the *MUC5B* variant and smoking history (OR 4.18 [95% CI 2.53, 6.93]). An additive interaction between *MUC5B* and smoking was not statistically significant (RERI

Table 2
Genotype frequency in patients with rheumatoid arthritis according to interstitial lung disease (ILD) status.

	RA-ILD (%)	RA no-ILD (%)	OR (95% CI) for RA-ILD vs RA no-ILD
SNPs (n=2561)			
MUC5B rs35705950 (G>T)			
GG	160 (67.2)	1910 (82.2)	2.25 (1.69, 3.02)
GT,TT	78 (32.7)	413 (17.8)	
TOLLIP rs5743890 (T>C)			
TT	188 (79.0)	1812 (78.0)	0.94 (0.68, 1.31)
TC,CC	50 (21.0)	511 (22.0)	
HLA-DRB1 (n=2335)			
HLA-DRB1 SE (copies)			
0	70 (31.7)	673 (31.8)	1.01 (0.75, 1.36)
1,2	151 (68.3)	1441 (68.2)	

Odds ratio (OR) and 95% confidence interval (CI) calculated using unadjusted logistic regression. All genetic variants were analyzed assuming autosomal dominant inheritance. Abbreviations: ILD (interstitial lung disease), OR (odds ratio), CI (confidence interval), SNP (single nucleotide polymorphism), SE (shared epitope)

Table 3
Associations of genetic risk variants and race and ethnicity with interstitial lung disease (ILD) in patients with rheumatoid arthritis.

	RA-ILD	RA no-ILD	OR (95% CI)	p value
SNPs (n=2561)				
Number (%)				
MUC5B rs35705950				
Non-Black/ <i>MUC5B</i> neg.	122 (51.3)	1550 (66.9)	1 (Ref)	-
Non-Black/ <i>MUC5B</i> pos.	72 (30.2)	395 (17.0)	2.32 (1.70, 3.16)	<0.001
Black/ <i>MUC5B</i> neg.	38 (16.0)	355 (15.3)	1.36 (0.93, 1.99)	0.12
Black/ <i>MUC5B</i> pos.	6 (2.5)	18 (0.8)	4.23 (1.65, 10.86)	0.003
RERI 1.56 (-1.34, 7.54)^a				
TOLLIP rs5743890				
Non-Black/ <i>TOLLIP</i> neg.	144 (60.5)	1448 (62.5)	1 (Ref)	-
Non-Black/ <i>TOLLIP</i> pos.	50 (21.0)	497 (21.4)	1.01 (0.72, 1.42)	0.95
Black/ <i>TOLLIP</i> neg.	44 (18.5)	359 (15.5)	1.23 (0.86, 1.76)	0.25
Black/ <i>TOLLIP</i> pos.	0	14 (0.6)	N/A	-
RERI N/A^b				
HLA-DRB1 (n=2334)				
HLA-DRB1 SE				
Non-Black/SE neg.	46 (20.8)	462 (21.9)	1 (Ref)	-
Non-Black/SE pos.	136 (61.5)	1307 (61.9)	1.05 (0.74, 1.48)	0.81
Black/SE neg.	24 (10.9)	210 (9.9)	1.15 (0.68, 1.93)	0.60
Black/SE pos.	15 (6.8)	134 (6.3)	1.12 (0.61, 2.08)	0.71
RERI -0.07 (-0.93, 0.79)^a				

^aRelative excess risk due to interaction (RERI) is represented as relative risk with 95% CI.
^bRERI could not be calculated due to inability to estimate odds in Black/*TOLLIP* positive patients. SNPs were analyzed assuming autosomal dominant inheritance. Black refers to all participants self-reporting as Black/African American. Non-Black refers to all participants self-reporting White, Asian American, American Indian/Pacific Islander, or Hispanic. Abbreviations: ILD (interstitial lung disease), OR (odds ratio), CI (confidence interval), SNP (single nucleotide polymorphism), RERI (relative excess risk of interaction), SE (shared epitope)

0.90 [95% CI -1.35, 2.81]. In similar analyses assessing associations of the *TOLLIP* rs5743890 variant and HLA-DRB1 SE with ILD by smoking history; smoking history, but not the genotypes, was associated with ILD. There was no evidence of additive interactions between smoking and these genotypes (Table 4).

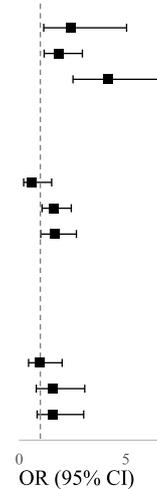
Association between the *MUC5B* variant and RA-ILD by disease activity status

The *MUC5B* rs35705950 variant was associated with ILD in both the DAS28 remission/low (OR 2.41 [95% CI 1.59, 3.65]) and moderate/high disease activity groups (OR 3.03 [95% CI 1.99, 4.61]) without

Table 4

Associations of genetic risk variants and smoking with interstitial lung disease (ILD) in patients with rheumatoid arthritis.

	RA-ILD	RA no-ILD	OR (95% CI)	p value	
SNPs (n=2561)		Number (%)			
MUC5B rs35705950					
Never smoker/MUC5B neg.	22 (9.2)	438 (18.9)	1 (Ref)	-	
Never smoker/MUC5B pos.	12 (5.0)	99 (4.3)	2.41 (1.16, 5.04)	0.02	
Ever smoker/MUC5B neg.	138 (58.0)	1471 (63.4)	1.87 (1.18, 2.97)	0.008	
Ever smoker/MUC5B pos.	66 (27.7)	314 (13.5)	4.18 (2.53, 6.93)	<0.001	
			RERI 0.90 (-1.35, 2.81)^a		
TOLLIP rs5743890					
Never smoker/TOLLIP neg.	29 (12.2)	414 (17.8)	1 (Ref)	-	
Never smoker/TOLLIP pos.	5 (2.1)	123 (5.3)	0.58 (0.22, 1.53)	0.27	
Ever smoker/TOLLIP neg.	159 (66.8)	1397 (60.2)	1.62 (1.08, 2.45)	0.02	
Ever smoker/TOLLIP pos.	45 (18.9)	388 (16.7)	1.66 (1.02, 2.69)	0.04	
			RERI 0.45 (-0.55, 1.21)^a		
HLA-DRB1 (n=2335)					
HLA-DRB1 SE					
Never smoker/SE neg.	11 (5.0)	153 (7.2)	1 (Ref)	-	
Never smoker/SE pos.	22 (10.0)	320 (15.1)	0.96 (0.45, 2.02)	0.91	
Ever smoker/SE neg.	59 (26.7)	520 (24.6)	1.58 (0.81, 3.08)	0.18	
Ever smoker/SE pos.	129 (58.4)	1121 (53.0)	1.60 (0.85, 3.03)	0.15	
			RERI 0.07 (-0.81, 0.94)^a		

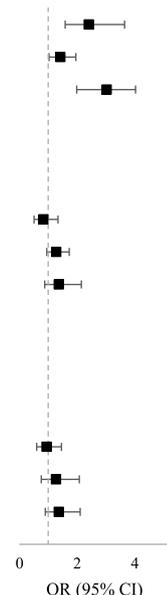


^aRelative excess risk due to interaction (RERI) is represented as relative risk with 95% CI. SNPs were analyzed assuming autosomal dominant inheritance. Abbreviations: ILD (interstitial lung disease), OR (odds ratio), CI (confidence interval), SNP (single nucleotide polymorphism), RERI (relative excess risk of interaction), SE (shared epitope)

Table 5

Association of genetic risk variants and disease activity with interstitial lung disease (ILD) in patients with rheumatoid arthritis.

	RA-ILD	RA no-ILD	OR (95% CI)	p value	
SNPs (n=2561)		Number (%)			
MUC5B rs35705950					
DAS28 REM-LDA/MUC5B neg.	73 (30.7)	1038 (44.7)	1 (Ref)	-	
DAS28 REM-LDA/MUC5B pos.	39 (16.4)	230 (9.9)	2.41 (1.59, 3.65)	<0.001	
DAS28 MDA-HDA/MUC5B neg.	87 (36.6)	872 (37.5)	1.42 (1.03, 1.96)	0.03	
DAS28 MDA-HDA/MUC5B pos.	39 (16.4)	183 (7.9)	3.03 (1.99, 4.61)	<0.001	
			RERI 0.20 (-1.24, 1.64)^a		
TOLLIP rs5743890					
DAS28 REM-LDA/TOLLIP neg.	90 (37.8)	979 (42.1)	1 (Ref)	-	
DAS28 REM-LDA/TOLLIP pos.	22 (9.2)	289 (12.4)	0.83 (0.51, 1.34)	0.45	
DAS28 MDA-HDA/TOLLIP neg.	98 (41.2)	833 (35.9)	1.28 (0.95, 1.73)	0.11	
DAS28 MDA-HDA/TOLLIP pos.	28 (11.8)	222 (9.6)	1.37 (0.88, 2.15)	0.17	
			RERI 0.26 (-0.49, 1.01)^a		
HLA-DRB1 (n=2335)					
HLA-DRB1 SE					
DAS28 REM-LDA/SE neg.	32 (14.5)	348 (16.5)	1 (Ref)	-	
DAS28 REM-LDA/SE pos.	67 (30.3)	776 (36.7)	0.94 (0.60, 1.46)	0.78	
DAS28 MDA-HDA/SE neg.	38 (17.2)	325 (15.4)	1.27 (0.76, 2.08)	0.34	
DAS28 MDA-HDA/SE pos.	84 (38.0)	665 (31.5)	1.37 (0.90, 2.11)	0.15	
			RERI 0.16 (-0.65, 0.73)^a		



^aRelative excess risk due to interaction (RERI) is represented as relative risk with 95% CI. Abbreviations: ILD (interstitial lung disease), OR (odds ratio), CI (confidence interval), SNP (single nucleotide polymorphism), DAS28 (28-joint disease activity score), REM (remission), LDA (low disease activity), MDA (moderate disease activity), HDA (high disease activity), SE (shared epitope)

evidence of an interaction between *MUC5B* and disease activity (RERI -0.20 [95% CI $-1.24, 1.64$]) (Table 5). Those with moderate/high disease activity were more likely to have ILD.

Secondary analyses

In secondary analysis evaluating RA-ILD pattern, the *MUC5B* rs35705950 variant was significantly associated with RA-ILD in a UIP pattern (OR 3.09 [95% CI 1.89, 5.04]), but not with non-UIP patterns (OR 1.54 [95% CI 0.69, 3.46]) (Table S1). There was no significant difference in RA duration at the time of ILD diagnosis between *MUC5B* variant positive (mean 10.7 \pm 11.3 years) vs. *MUC5B* variant negative (mean 11.3 \pm 12.1 years) RA-ILD patients ($p = 0.76$). Estimated associations between the *MUC5B* variant and ILD were similar between those with earlier (OR 2.00 [95% CI 1.30, 3.06]) and later ILD diagnosis (OR 2.51 [95% CI 1.69, 3.75]), dichotomizing at the median RA duration (Table S2).

To assess the specificity of associations of *MUC5B* with ILD, we also assessed its association with COPD/bronchiectasis by smoking history (Table S3). In contrast to findings for ILD, the odds of COPD/bronchiectasis were driven by smoking status (OR 3.31 [95% CI 2.53, 4.32] for *MUC5B*-negative smokers) rather than the presence of the variant (OR 0.77 [95% CI 0.42, 1.43] for *MUC5B*-positive nonsmokers). There was no additive interaction between *MUC5B* and smoking history for COPD/bronchiectasis (RERI 0.82 [95% CI $-0.23, 1.93$]).

Discussion

In this large cohort of U.S. veterans with RA, the *MUC5B* rs35705950 promoter variant was associated with >2 -fold odds of ILD. This association was similar in magnitude to that observed in a prior study [7] and was restricted to RA-ILD of a UIP pattern. Conversely, there was no association between the *TOLLIP* rs5743890 variant, another genetic variant previously associated with IPF [7,10-12], or HLA-DRB1 SE with RA-ILD. Expanding upon prior work, we found that *MUC5B* associations with ILD were similar (or potentially magnified) in patient sub-groups defined by Black/African American race and smoking history, although estimated additive interactions were not statistically significant. These novel findings advance our understanding of the genetic and environmental risks of RA-ILD and begin to elucidate the potential interactions among distinct RA-ILD risk factors.

The *MUC5B* rs35705950 variant was previously shown to be associated with RA-ILD in international and Northern European cohorts that were predominantly composed of White and Asian individuals [7,9]. For the first time, we estimated the odds of RA-ILD based on the presence of the *MUC5B* variant among specific racial groups. While the frequency of the *MUC5B* rs35705950 variant was lower in Black/African American participants (5.8% vs. 21.8%), the estimated odds ratios for ILD were highest in Black/African American participants with the *MUC5B* variant (OR 4.23). The point estimate for the RERI (1.56) indicated the possibility of an additive interaction, though this estimate was imprecise and not statistically significant. This imprecision was the result of limited numbers of Black/African American participants with the *MUC5B* variant and prohibits definitive conclusions regarding differential risk of ILD with the *MUC5B* variant in this group. In contrast, associations between the *TOLLIP* variant and HLA-DRB1 SE alleles with ILD were null, and unchanged in subgroups defined by race and ethnicity. This study involves one of the largest and most comprehensive examinations of RA-ILD in the Black/African American population, which has historically been an understudied group among patients with RA-ILD. Despite our large cohort, the low frequency of the *MUC5B* variant in Black/African American participants resulted in imprecision when assessing for an additive interaction that will require further investigation in studies with a greater number of Black/African American participants. Such studies are needed to improve ILD risk assessment, identify social determinants of health and health inequities that may underlie these

differences, and reduce the burden of RA-ILD.

Previously identified risk factors for ILD in RA have consisted of older age, cigarette smoking, and more severe RA [3,4]. How these may affect ILD risk in the context of various genetic backgrounds has not been elucidated. Smoking history was associated with ILD, and the odds of ILD further increased in the setting of the *MUC5B* variant, but not the *TOLLIP* variant or HLA-DRB1 SE. *MUC5B*-positive smokers had >4 -fold higher odds of ILD, which was independent of race and ethnicity. While this indicates a group at high risk of RA-ILD, we did not detect a significant additive interaction between these two factors, which is consistent with previous findings [7]. The lack of an interaction between HLA-DRB1 SE alleles and smoking in ILD risk contrasts with the gene-smoking interaction noted for seropositive RA risk [13]. There was also no evidence of interaction between any of the genetic risk factors and RA disease activity on ILD risk. Our results therefore suggest that the *MUC5B* rs35705950 variant impacts the risk of RA-ILD without modification from these known environmental RA-ILD risk factors. These novel gene-environment analyses further our understanding of how genetic risk factors ultimately influence the development of ILD.

Limitations to this study include the imprecision of some estimates of lung disease risk, particularly in subgroup analyses. In the Black/African American population specifically, the absolute numbers within each strata were limited due to the low frequency of the variant, which led to imprecision in the risk estimates. While external validation of the associations of the *MUC5B* promoter variant with RA-ILD by race should be completed in additional cohorts, the cohort used in this study recruited from 12 VA sites and is highly representative of the national VA population. Characterization of ILD was retrospective, which may result in misclassification, such as under-recognition of subclinical ILD, which would bias findings towards the null. Similarly, assessment of COPD/bronchiectasis in secondary analyses was based on administrative classification. There was no quantitative data on number of pack-years of smoking, and we were unable to assess for other inhalant exposures that could also impact ILD risk. This cohort is predominantly composed of male veterans, which may limit generalizability to other populations but does provide important information specific to an otherwise understudied population in RA. There may be additional genetic risk factors that contribute to RA-ILD risk which were not evaluated in this study. Strengths of this study include the relatively large number of Black/African American participants with RA that allowed racial and ethnic subgroup analysis, and robust clinical/demographic data on study participants to allow assessment of gene-environment interactions.

In conclusion, the *MUC5B* rs35705950 promoter variant was demonstrated to be associated with the presence of ILD, specifically UIP ILD, in a cohort of U.S. veterans with RA. Neither *TOLLIP* rs5743890 or HLA-DRB1 SE alleles were associated with the presence of ILD. While there was a lower frequency of the *MUC5B* variant allele in Black/African American individuals; when present, the association with ILD was of similar or greater magnitude than in other racial and ethnic groups. Individuals possessing *MUC5B* variant alleles and with a smoking history were also at particularly higher risk of RA-ILD. These findings characterize how genetic, social, and environmental factors affect RA-ILD risk and can inform precision risk stratification in RA-ILD.

Key messages

- Participants with rheumatoid arthritis (RA) who have the *MUC5B* rs35705950 promoter variant and a smoking history are at higher risk of interstitial lung disease (ILD), but there was no evidence of a significant additive gene-smoking interaction.
- The *MUC5B* rs35705950 promoter variant is less common among Black/African American patients with RA, but it appears to portend a risk of ILD that is similar, or greater, in magnitude than among other racial and ethnic groups.

- Neither the *TOLLIP* rs5743890 variant nor HLA-DRB1 shared epitope alleles were associated with RA-ILD risk.
- Risk of ILD among people with RA appears to vary by social and environmental factors, but larger studies in diverse populations are needed to elucidate interactions between these and genetic risk factors.

Funding

AMW is supported by the Rheumatology Research Foundation. BRE is supported by a VA CSR&D (IK2 CX002203) and the Rheumatology Research Foundation. TRM is supported by grants from the VA (BLR&D Merit I01 BX004660), National Institutes of Health (2U54GM115458), U.S. Department of Defense (PR200793), and the Rheumatology Research Foundation. JFB is supported by a VA CSR&D Merit (CX0001703) and an RR&D Merit (RX0003644). KDW is supported by VA CSR&D (IK2CX002351) and by the Rheumatology Research Foundation. JAP receives funding from the Department of Defense (PR200793) and National Institute for Occupational Safety and Health (R01OH012045 and U54OH010162). NS is supported by the Rheumatology Research Foundation, American Heart Association, and National Institute for Arthritis and Musculoskeletal and Skin Diseases (K23AR079588).

Disclosures

BRE has consulted for Boehringer-Ingelheim and UpToDate. TRM has consulted for Horizon Therapeutics, Pfizer, Gilead, and Sanofi. GSK has consulted with Janssen, Horizon, UCB, Sanofi, BMS, and Amgen. JAP has received research grant funding from AstraZeneca and clinical research study support from Takeda and GlaxoSmithKline. JFB has received consulting fees from Bristol-Myers Squibb, Pfizer, CorEvitas, and Burns-White, LLC. NS has received consulting fees from Pfizer.

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.semarthrit.2022.152098](https://doi.org/10.1016/j.semarthrit.2022.152098).

References

- [1] Spagnolo P, et al. The lung in rheumatoid arthritis: focus on interstitial lung disease. *Arthritis Rheumatol* 2018;70(10):1544–54.
- [2] Marigliano B, et al. Lung involvement in connective tissue diseases: a comprehensive review and a focus on rheumatoid arthritis. *Autoimmun Rev* 2013;12(11):1076–84.
- [3] Chartrand S, et al. Clinical characteristics and natural history of autoimmune forms of interstitial lung disease: a single-center experience. *Lung* 2019;197(6):709–13.
- [4] Kelly CA, et al. Rheumatoid arthritis-related interstitial lung disease: associations, prognostic factors and physiological and radiological characteristics—a large multicentre UK study. *Rheumatology (Oxford)* 2014;53(9):1676–82.
- [5] Lee HK, et al. Histopathologic pattern and clinical features of rheumatoid arthritis-associated interstitial lung disease. *Chest* 2005;127(6):2019–27.
- [6] Lederer DJ, Martinez FJ. Idiopathic pulmonary fibrosis. *N Engl J Med* 2018;378(19):1811–23.
- [7] Juge PA, et al. MUC5B promoter variant and rheumatoid arthritis with interstitial lung disease. *N Engl J Med* 2018;379(23):2209–19.
- [8] Seibold MA, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011;364(16):1503–12.
- [9] Jönsson E, et al. Pulmonary fibrosis in relation to genetic loci in an inception cohort of patients with early rheumatoid arthritis from northern Sweden. *Rheumatology (Oxford)* 2022;61(3):943–52.
- [10] Fingerlin TE, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013;45(6):613–20.
- [11] Fingerlin TE, et al. Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet* 2016;17(1):74.
- [12] Noth I, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013;1(4):309–17.
- [13] Källberg H, et al. Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. *Ann Rheum Dis* 2011;70(3):508–11.
- [14] Hedström AK, et al. Complex relationships of smoking, HLA-DRB1 genes, and serologic profiles in patients with early rheumatoid arthritis: update from a Swedish population-based case-control study. *Arthritis Rheumatol* 2019;71(9):1504–11.
- [15] Karlson EW, et al. Gene-environment interaction between HLA-DRB1 shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis. *Ann Rheum Dis* 2010;69(1):54–60.
- [16] Kronzer VL, et al. Lifestyle and clinical risk factors for incident rheumatoid arthritis-associated interstitial lung disease. *J Rheumatol* 2021;48(5):656–63.
- [17] Sparks JA, et al. Rheumatoid arthritis disease activity predicting incident clinically apparent rheumatoid arthritis-associated interstitial lung disease: a prospective cohort study. *Arthritis Rheumatol* 2019;71(9):1472–82.
- [18] Mikuls TR, et al. Insights and implications of the VA rheumatoid arthritis registry. *Fed Pract* 2015;32(5):24–9.
- [19] Arnett FC, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31(3):315–24.
- [20] van Riel PL, Fransen J. Established rheumatoid arthritis: clinical assessments. *Best Pract Res Clin Rheumatol* 2007;21(5):807–25.
- [21] Mikuls TR, et al. Anticitrullinated protein antibody (ACPA) in rheumatoid arthritis: influence of an interaction between HLA-DRB1 shared epitope and a deletion polymorphism in glutathione S-transferase in a cross-sectional study. *Arthritis Res Ther* 2010;12(6):R213.
- [22] Illumina, I. *Infinium global screening Array v2.0* locus report. 2018.
- [23] England BR, et al. Performance of Administrative algorithms to identify interstitial lung disease in rheumatoid arthritis. *Arthritis Care Res (Hoboken)* 2020;72(10):1392–403.
- [24] England BR, et al. Malondialdehyde-acetaldehyde adducts and antibody responses in rheumatoid arthritis-associated interstitial lung disease. *Arthritis Rheumatol* 2019;71(9):1483–93.
- [25] Natalini JG, et al. Autoantibody seropositivity and risk for interstitial lung disease in a prospective male-predominant rheumatoid arthritis cohort of U.S. veterans. *Ann Am Thorac Soc* 2021;18(4):598–605.
- [26] England BR, et al. Chronic lung disease in U.S. Veterans with rheumatoid arthritis and the impact on survival. *Clin Rheumatol* 2018;37(11):2907–15.
- [27] Richardson DB, Kaufman JS. Estimation of the relative excess risk due to interaction and associated confidence bounds. *Am J Epidemiol* 2009;169(6):756–60.