Autoantibodies against citrullinated and native proteins and prediction of rheumatoid arthritis-associated interstitial lung disease: a nested case–control study

Vanessa L Kronzer, Keigo Hayashi, Kazuki Yoshida, John M Davis III, Gregory C McDermott, Weixing Huang, Paul F Dellaripa, Jing Cui, Vivi Feathers, Ritu R Gill, Hiroto Hatabu, Mizuki Nishino, Rachel Blaustein, Cynthia S Crowson, William H Robinson, Jeremy Sokolove, Katherine P Liao, Michael E Weinblatt, Nancy A Shadick, Tracy J Doyle, Jeffrey A Sparks**

Summary

Background Rheumatoid arthritis-associated interstitial lung disease (ILD) is one of the leading causes of premature death among patients with rheumatoid arthritis. Improving prediction of rheumatoid arthritis-associated ILD is crucial to allow for earlier diagnosis and treatment. We aimed to identify fine-specificity anti-citrullinated protein antibodies (ACPAs) associated with incident rheumatoid arthritis-associated ILD.

Methods In this nested case–control study within the prospective Brigham Rheumatoid Arthritis Sequential Study (BRASS), we matched cases of incident rheumatoid arthritis-associated ILD diagnosed between March 1, 2003, and April 14, 2016, to control patients with rheumatoid arthritis without ILD on the following characteristics: time of blood collection, age, sex, rheumatoid arthritis duration, and rheumatoid factor status. We measured ACPA and anti-native protein antibodies using a multiplex assay on stored serum collected before onset of rheumatoid arthritis-associated ILD. We used logistic regression models to calculate odds ratios (ORs) with 95% CIs for rheumatoid arthritis-associated ILD, adjusting for prospectively collected covariates. We estimated the optimism-corrected area under the curves (AUCs) using internal validation. We used model coefficients to generate a risk score for rheumatoid arthritis-associated ILD.

Findings We identified 84 incident rheumatoid arthritis-associated ILD cases (mean age 67 [SD 10] years, 65 [77%] female and 19 [23%] male, 76 [90%] White) and 233 rheumatoid arthritis controls without ILD (mean age 66 [11] years, 186 [80%] female and 47 [20%] male, 219 [94%] White). We identified six fine-specificity antibodies that were associated with rheumatoid arthritis-associated ILD. The antibody isotypes and targeted proteins were IgA2 to citrullinated histone 4 (adjusted OR 0·08 [95% CI 0·03–0·22] per log-transformed unit), IgA2 to citrullinated histone 2A (4·03 [2·03–8·00]), IgG to cyclic citrullinated filaggrin (3·47 [1·71–7·01]), IgA2 to native cyclic histone 2A (5·52 [2·38–12·78]), IgA2 to native histone 2A (4·60 [2·18–9·74]), and IgG to native cyclic filaggrin (2·53 [1·47–4·34]). These six antibodies predicted the risk of rheumatoid arthritis-associated ILD better than did all clinical factors combined (optimism-corrected AUC 0·84 versus 0·73). We developed a risk score for rheumatoid arthritis-associated ILD by combining these antibodies with clinical factors (smoking, disease activity, glucocorticoid use, and obesity). At 50% predicted probability of developing rheumatoid arthritis-associated ILD, the risk scores both without (2·6) and with (5·9) antibody biomarkers achieved a specificity of 93% or higher for rheumatoid arthritis-associated ILD.

Interpretation Specific ACPAs and anti-native protein antibodies improve prediction of rheumatoid arthritis-associated ILD. These findings implicate synovial protein antibodies in the pathogenesis of rheumatoid arthritis-associated ILD and, once validated in external studies, suggest that these antibodies might have clinical utility in predicting the development of ILD in patients with rheumatoid arthritis.

Funding US National Institutes of Health/National Institute of Arthritis and Musculoskeletal and Skin Diseases.

Copyright © 2023 Elsevier Ltd. All rights reserved.

Lancet Rheumatol 2023; 5: e77–87

See **Comment** page e61

*Co-senior authors

Division of Rheumatology (V L Kronzer MD, Prof J M Davis III MD, Prof C S Crowson PhD), **and Department of Quantitative Health Sciences** (Prof C S Crowson), **Mayo Clinic, Rochester, MN, USA; Division of Rheumatology, Inflammation, and Immunity** (K Hayashi PhD, K Yoshida MD, G C McDermott MD, W Huang MSPH, P F Dellaripa MD, J Cui PhD, V Feathers MS, R Blaustein BS, K P Liao MD, Prof M E Weinblatt MD, N A Shadick MD, J A Sparks MD), **Department of Radiology** (Prof H Hatabu MD, M Nishino MD), **and Division of Pulmonary and Critical Care** (T J Doyle MD), **Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; Department of Radiology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA** (R R Gill MD); **Department of Imaging, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA** (M Nishino); **Stanford University School of Medicine, Palo Alto, CA, USA** (Prof W H Robinson MD, J Sokolove MD); **VA Palo Alto Health Care System, Palo Alto, CA, USA** (Prof W H Robinson, J Sokolove)

Correspondence to: Dr Jeffrey A Sparks, Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Boston, MA 02115, USA **jsparks@bwh.harvard.edu**

Introduction

Approximately 5–10% of individuals with rheumatoid arthritis have clinically significant rheumatoid arthritis-associated interstitial lung disease (ILD).[1] Having rheumatoid arthritis-associated ILD confers two to ten times the odds of mortality compared with patients with rheumatoid arthritis without ILD,[2] with median survival ranging between 3 and 8 years after diagnosis in previous studies.[1,2] As a result, rheumatoid arthritis-associated ILD is one of the leading causes of premature death among patients with rheumatoid arthritis.[1] Improving prediction of rheumatoid arthritis-associated ILD is therefore crucial to allow for earlier diagnosis and treatment before worsening lung damage and death.

Biomarkers might hold the key to predicting the development of rheumatoid arthritis-associated ILD. For example, rheumatoid factor and anti-cyclic citrullinated peptide (anti-CCP) have been associated with the

Research in context

Evidence before this study
To appraise the previously published literature on rheumatoid arthritis-associated interstitial lung disease (ILD), before undertaking the study we searched PubMed for studies in English from database inception to Sept 6, 2022, using the following search terms: "rheumatoid arthritis[MeSH Terms] AND interstitial lung disease[MeSH Terms]". Rheumatoid arthritis-associated ILD is a leading cause of the excess mortality burden in individuals with rheumatoid arthritis. However, the predictive value of clinical factors is modest, and previous studies have focused on detecting prevalent subclinical rheumatoid arthritis-associated ILD rather than predicting incident rheumatoid arthritis-associated ILD.

Added value of this study
We identified several fine-specificity anti-citrullinated protein antibodies and anti-native protein antibodies, especially to histone 2A and filaggrin, that were associated with an increased risk of developing incident rheumatoid arthritis-associated ILD. Furthermore, the risk score developed in this nested case–control study had a high sensitivity and specificity for rheumatoid arthritis-associated ILD.

Implications of all the available evidence
Clinical practice and policy guidelines in rheumatology should consider implementing screening for rheumatoid arthritis-associated ILD in high-risk patients with rheumatoid arthritis by use of scoring systems such as those outlined in this study. Future research should validate this rheumatoid arthritis-associated ILD risk score in other populations. Additionally, future studies should continue to discover and validate biomarkers, including fine-specificity protein antibodies, that influence the risk of rheumatoid arthritis-associated ILD.

development of rheumatoid arthritis-associated ILD.[3] More recently, the gain-of-function *MUC5B* promoter variant was found to increase the risk of rheumatoid arthritis-associated ILD by three times.[4] Studies have also shown that anti-modified peptide antibodies (AMPAs)[5] and fine-specificity anti-citrullinated protein antibodies (ACPAs) such as anti-citrullinated alpha enolase (anti-CEP1),[6] anti-citrullinated heat shock protein 90 (anti-HSP90),[7] and anti-fibrinogen[8] are associated with rheumatoid arthritis-associated ILD. Identifying and testing additional fine-specificity ACPAs might be important, as anti-CCP testing does not encompass all ACPAs. Furthermore, certain fine-specificity ACPAs could be more associated with rheumatoid arthritis-associated ILD than others. One cross-sectional study suggested an association between the number of fine-specificity ACPAs and subclinical rheumatoid arthritis-associated ILD.[3] However, prospective studies examining each fine-specificity ACPA separately in rheumatoid arthritis-associated ILD are needed.

In addition to identification of novel biomarkers, another important unmet need is the development of an accurate rheumatoid arthritis-associated ILD risk prediction score based on such biomarkers. Our previous work showed that combining established clinical risk factors achieves only a modest area under the curve (AUC) in predicting rheumatoid arthritis-associated ILD.[9,10] However, adding biomarkers such as matrix metalloproteinase 7, pulmonary and activation-regulated chemokine, and surfactant protein D significantly improved prediction of rheumatoid arthritis-associated ILD.[10] Juge and colleagues[11] also showed that adding the *MUC5B* promoter variant improved the AUC for prediction of rheumatoid arthritis-associated ILD to 0·82, although this study detected the presence of subclinical rheumatoid arthritis-associated ILD (rather than future development of rheumatoid arthritis-associated ILD) and was based on cross-sectional data. Another risk score recently developed for detection of subclinical rheumatoid arthritis-associated ILD based on cross-sectional data highlighted the urgent need for both additional biomarkers and prospective studies to improve prediction of clinical rheumatoid arthritis-associated ILD.[12] Therefore, a risk score for rheumatoid arthritis-associated ILD is needed that incorporates more biomarkers and is based on prospective data collection and clinical rheumatoid arthritis-associated ILD (for improved clinical utility).

To address these two gaps, we used data from a prospective cohort of patients with incident rheumatoid arthritis-associated ILD. We aimed to identify fine-specificity ACPAs or anti-native protein antibodies associated with incident rheumatoid arthritis-associated ILD and to develop a risk score for predicting the incidence of rheumatoid arthritis-associated ILD. We hypothesised that several specific ACPAs such as anti-citrullinated fibrinogen[8] would be associated with an increased risk of incident rheumatoid arthritis-associated ILD beyond the effects of the *MUC5B* promotor variant, and that a risk score for rheumatoid arthritis-associated ILD that incorporates such biomarkers would improve prediction of this life-threatening disease.

Methods

Study population and design

This nested case–control study took place within the prospective Brigham Rheumatoid Arthritis Sequential Study (BRASS). BRASS is an ongoing prospective registry of more than 1500 adult patients aged 18 years or older with rheumatoid arthritis, containing data (collected semi-annually) on rheumatoid arthritis characteristics, physician-reported and patient-reported measures, and blood and DNA samples.[13] BRASS recruited participants from Brigham and Women's

Hospital outpatient clinics from March 1, 2003, to present, with an overall participation rate of approximately 18%. All participants met 1987 American College of Rheumatology (ACR) or 2010 ACR–European Alliance of Associations for Rheumatology (EULAR) criteria for rheumatoid arthritis as verified by a rheumatologist. Criteria for inclusion in this substudy consisted of the availability of banked serum collected during a BRASS study visit for the ACPA assay. We defined the index date of rheumatoid arthritis-associated ILD onset as the initial chest CT scan showing rheumatoid arthritis-associated ILD, or the matched date for controls. For each participant, we also chose the earliest study visit before the index date, during which blood was obtained for both antibody measurements and covariate assessments. This study received institutional review board approval (2016P000226) from Mass General Brigham and complies with the Declaration of Helsinki. All participants provided informed consent.

Rheumatoid arthritis-associated ILD cases

We identified patients with rheumatoid arthritis-associated ILD (cases) diagnosed between March 1, 2003, and April 14, 2016. All patients had incident rheumatoid arthritis-associated ILD, as outlined previously.[9] We excluded patients with prevalent rheumatoid arthritis-associated ILD at baseline of their enrollment into the study. All patients with incident rheumatoid arthritis-associated ILD after baseline had blood banked for these studies. In brief, we defined rheumatoid arthritis-associated ILD as interstitial changes on clinically indicated chest CT after BRASS study enrolment, confirmed by two attending thoracic radiologists and one attending pulmonologist by use of a validated sequential reading method.[14] Previous studies confirmed this definition to be clinically significant, with more than 75% of cases having additional evaluations or follow-up for rheumatoid arthritis-associated ILD.[15] We also assessed for rheumatoid arthritis-associated ILD subtypes and severity, and collected data on clinically performed pulmonary function tests within 6 months of rheumatoid arthritis-associated ILD onset.

Rheumatoid arthritis controls without ILD

We matched each rheumatoid arthritis-associated ILD case with up to three rheumatoid arthritis controls without ILD, on the following characteristics: the duration between the study visit blood draw date and index date (±6 months), age (±5 years), sex, rheumatoid arthritis duration (±5 years), and rheumatoid factor status (positive *vs* negative). Eligible control patients never developed rheumatoid arthritis-associated ILD throughout the entire follow-up. We required controls to have banked blood available at the study visit blood draw date. To ensure controls did not have ILD, we also required the absence of International Classification of Diseases, Ninth Revision (ICD-9) codes for ILD (515, 516.30, and 516.31), the absence of patient-reported ILD, and the absence of physician-reported history of ILD, asthma, chronic obstructive pulmonary disease, pulmonary fibrosis, bronchiectasis, bronchiolitis obliterans-organising pneumonia, drug-induced pneumonitis, or tuberculosis.

Research multiplex ACPA assay

We measured antibody concentrations with a research multiplex flow cytometry assay performed at Stanford University (Palo Alto, CA, USA). Previous work identified key fine-specificity proteins and peptides from rheumatoid arthritis synovium and cartilage, and these were combined into a validated, bead-based multiplex assay.[16] In brief, protein antigens were conjugated to beads with the Bio-Plex multiplex assay platform (Bio-Rad Laboratories, Hercules, CA, USA) and analysed with Luminex 200 (Luminex, Austin, TX, USA). Each plate used established samples with no, low, medium, or high reactivity as internal controls. The specific amino acid sequences, citrullinated residues, and reproducibility of the targeted antigens for each antibody have been previously published.[17] The assay has previously passed validation standards that included blinded remeasurements from the same samples,[18] and the amino acid sequences for each peptide are publicly available.

For the **NCBI Protein database** see https://www.ncbi.nlm.nih.gov/protein

We did this research multiplex assay for all cases and controls in a single batch at Stanford University. Serum from each participant was added to the bead mix, and antibody reactivity was measured in raw fluorescent intensity units. Citrullinated antigens tested for this study included beta-actin, biglycan, collagen type II, enolase-1-alpha, histone 2A, histone 2B, histone 4, fibrinogen A, fibrinogen B, fibronectin, filaggrin, and vimentin. Native (non-citrullinated) antigens tested included histone 2A, histone 2B, fibrinogen B, filaggrin, tenascin C1, tenascin C5, type II collagen, and vimentin. Some proteins had multiple epitope targets, and we also studied three immunoglobulin isotypes for each target. Thus, we measured a total of 78 fine-specificity ACPAs and 39 anti-native protein antibodies. The assay did not measure antibodies to some native antigens because previous studies showed no appreciable reactivity to the native antigens in patients with rheumatoid arthritis nor in healthy controls.[17] A previous study that used this assay found that 2% of healthy controls were ACPA positive.[19] We also examined possible correlations between anti-CCP titre and each fine-specificity ACPA and found modest correlations.

MUC5B promoter variant

We assessed *MUC5B* promoter variant status (rs35705950 G>T) from genetic testing done through the Mass General Brigham Biobank with the genotyping array (Illumina, San Diego, CA, USA), as previously detailed.[20] For participants not included or tested in the Mass

	Rheumatoid arthritis-associated ILD cases (n=84)	Rheumatoid arthritis controls without ILD (n=233)	p value
Years from blood draw to index or matched date	1·5 (0·6–2·5)	1·9 (1·6–5·6)	0·84*
Age, years	67 (10)	66 (11)	0·41*
Sex			
Female	65 (77%)	186 (80%)	0·64*
Male	19 (23%)	47 (20%)	..
White, non-Hispanic	76 (90%)	219 (94%)	0·28
Rheumatoid arthritis duration, years	20 (12)	20 (11)	0·99*
Rheumatoid factor positive	73 (87%)	195 (84%)	0·48*
Rheumatoid factor level			
Negative (≤ULN)	11 (13%)	38 (16%)	0·53
Low positive (>1× to 3× ULN)	21 (25%)	67 (29%)	..
High positive (>3× ULN)	52 (62%)	128 (55%)	..
Anti-CCP positive	70 (83%)	190 (82%)	0·71
Anti-CCP concentration			
Negative (≤ULN)	14 (17%)	43 (18%)	0·86
Low positive (>1× to 3× ULN)	9 (11%)	21 (9%)	..
High positive (>3× ULN)	61 (73%)	169 (73%)	..
Education less than university level	30 (36%)	98 (42%)	0·31
BMI, kg/m²	28 (6)	27 (6)	0·029
Smoking status			
Never smoked	33 (39%)	124 (53%)	0·030
Past smoker	43 (51%)	100 (43%)	..
Current smoker	8 (10%)	9 (4%)	..
Smoking pack-years	7 (0–31)	0 (0–10)	0·0008
DAS28-CRP	3·6 (1·6)	3·0 (1·3)	0·0010
CRP, mg/L	3·6 (1·3–9·9)	2·1 (0·8–5·1)	0·0034
MDHAQ score	0·8 (0·6)	0·5 (0·5)	<0·0001
Bone erosions	49 (58%)	144 (62%)	0·58
Rheumatoid nodules	42 (50%)	80 (34%)	0·011
Biologic DMARD use			
Never	29 (35%)	76 (33%)	0·027
Past	19 (23%)	27 (12%)	..
Current	36 (43%)	130 (56%)	..
Methotrexate use			
Never	18 (21%)	24 (10%)	0·031
Past	32 (38%)	93 (40%)	..
Current	34 (40%)	116 (50%)	..
Glucocorticoid use			
Never	5 (6%)	31 (13%)	0·0066
Past	42 (50%)	140 (60%)	..
Current	37 (44%)	62 (27%)	..
Copies of *MUC5B* promoter variant (ie, genotype)			
None (GG)	48/69 (70%)	184/231 (80%)	0·16
One (GT)	19/69 (28%)	43/231 (19%)	..
Two (TT)	2/69 (3%)	4/231 (2%)	..

Data are n (%), n/N (%), mean (SD), or median (IQR). ILD=interstitial lung disease. ULN=upper limit of normal. CCP=cyclic citrullinated peptide. DAS28-CRP=Disease Activity Score for 28 joints with CRP. CRP=C-reactive protein. MDHAQ=Multi-Dimensional Health Assessment Questionnaire. DMARD=disease-modifying anti-rheumatic drug. *Matched factor.

Table 1: **Characteristics of 84 rheumatoid arthritis-associated ILD cases and 233 rheumatoid arthritis controls without ILD at the time of blood draw**

General Brigham Biobank but with DNA available for genotyping, we directly measured the single nucleotide polymorphism using a custom TaqMan assay (Thermo Fisher, Waltham, MA, USA) done at the Mass General Brigham HealthCare Dana-Farber Brigham Cancer Center/Harvard Cancer Center (Boston, MA, USA).

Covariates

We obtained covariates known to be associated with rheumatoid arthritis-associated ILD[9] at the time of the earliest available study visit before the index date during which blood was collected. These included age (continuous), sex (male *vs* female, self-reported), rheumatoid arthritis duration (continuous), rheumatoid factor positivity (present *vs* absent), anti-CCP positivity (present *vs* absent), race and ethnicity (White non-Hispanic *vs* other), education (less than university level *vs* completed university education), BMI (continuous), smoking status (never, past, or current), smoking pack-years (continuous), C-reactive protein (CRP; continuous), Disease Activity Score for 28 joints with CRP (DAS28-CRP; continuous), Multi-Dimensional Health Assessment Questionnaire (MDHAQ; continuous), bone erosions (present *vs* absent), rheumatoid nodules (present *vs* absent), biologic disease modifying anti-rheumatic drug (DMARD) use (never, past, or current), methotrexate use (never, past, or current), and glucocorticoid use (never, past, or current). We also categorised rheumatoid factor and anti-CCP as negative (≤upper limit of normal [ULN]), low-positive (>1× to 3× ULN), and high-positive (>3× ULN).

Notably, if covariates were not available at the date of the earliest blood draw, we took them from the next closest BRASS study visit that preceded the index date. All covariates were measured for research rather than clinical care. Third-generation anti-CCP measurements were done with validated enzyme-linked immunosorbent assays from Inova Diagnostics (San Diego, CA, USA) and Euro Diagnostica (Minneapolis, MN, USA) with a ULN of 19·9. Rheumatoid factor measurements were done by an immunoturbidimetric technique on the Cobas Integra 700 analyser (Roche Diagnostics; Indianapolis, IN, USA), with reagents and calibrators from Roche with a ULN of 15.

Statistical analysis

For demographic characteristics, we compared normally distributed continuous variables using *t* tests, non-normally distributed continuous variables using the Wilcoxon rank sum test, and categorical variables using the χ^2 or Fisher's exact tests (for variables with low cell sizes). Medians were reported with IQRs.

To identify specific ACPAs and anti-native protein antibodies associated with rheumatoid arthritis-associated ILD, we log-transformed these biomarkers due to skewed distributions. We then did multivariable conditional logistic regression models for each

log-transformed unit antibody to obtain adjusted odds ratios (aORs) for rheumatoid arthritis-associated ILD. These models adjusted for all covariates that were statistically significant in univariate analyses, including BMI, smoking pack-years, DAS28-CRP, MDHAQ, rheumatoid nodules, biologic DMARD use, methotrexate use, and glucocorticoid use. We also included *MUC5B* genotype to determine the associations of these novel biomarkers beyond *MUC5B,* the best known genetic risk factor for rheumatoid arthritis-associated ILD. Because of the number of antibodies tested, we adjusted p values for multiple comparisons by use of Benjamini and Hochberg's procedure with a false discovery rate (FDR) of 5%. To assess the added benefit of significant biomarkers (FDR p<0·05) on prediction of rheumatoid arthritis-associated ILD, we fit an unconditional logistic regression including both the predictors of interest and the matching factors to obtain receiver operator characteristic (ROC) curves and associated AUCs. We also plotted ROC curves and reported AUCs for each antibody identified in the association analyses as well as two groups: citrullinated and native proteins. We did internal validation using a bootstrapping optimism correction[21] and then reported the optimism-corrected AUC. In brief, we did 1000 bootstrap replications to obtain resamples with AUCs and 95% CI metrics. Since the pathogenesis of rheumatoid arthritis-associated ILD subtypes might differ, we also investigated antibody associations for usual interstitial pneumonia or fibrotic non-specific interstitial pneumonia and cellular non-specific interstitial pneumonia subtypes since these subtypes were the most frequent. We also did analyses stratified by seropositive and seronegative rheumatoid arthritis, and anti-CCP-positive and anti-CCP-negative rheumatoid arthritis. To investigate the potential for cross-reactivity of antibodies against citrullinated and non-citrullinated antibodies, we did a sensitivity analysis subtracting the results of non-citrullinated biomarkers from those of citrullinated biomarkers for each specific antigen.

	Adjusted* odds ratio (95% CI)	FDR p value
Citrullinated antigens		
H4 33–48 citrullinated 39 (IgA2)	0·08 (0·03–0·22)	<0·0001
H2A/a-2 1–20 citrullinated (IgA2)	4·03 (2·03–8·00)	0·0027
Filaggrin 48–65 cit2 cyclic (IgG)	3·47 (1·71–7·01)	0·014
FibrinogenB 36–52 citrullinated (IgA2)	0·37 (0·16–0·86)	0·397
Clusterin 231–250 citrullinated cyclic (IgG)	1·27 (1·03–1·56)	0·397
H4 33–48 citrullinated 39–40 (IgG)	1·25 (1·02–1·53)	0·397
Filaggrin 48–65 cit2 cyclic (IgA2)	5·04 (1·10–23·1)	0·419
H2A/a 1–20 citrullinated cyclic (IgA2)	1·60 (1·00–2·57)	0·455
Fibronectin citrullinated 1035–36 (IgG)	1·22 (0·99–1·50)	0·455
Filaggrin 48–65 cit2 cyclic (IgA1)	1·87 (0·98–3·59)	0·455
Non-citrullinated (native) antigens		
H2A/a 1–20 cyclic (IgA2)	5·52 (2·38–12·78)	0·0013
H2A/a-2 1–20 (IgA2)	4·60 (2·18–9·74)	0·0013
Filaggrin 48–65 cyclic (IgG)	2·53 (1·47–4·34)	0·010
H2A/a 1–20 cyclic (IgG)	1·91 (1·18–3·10)	0·088
H2A/a-2 1–20 (IgG)	1·86 (1·14–3·05)	0·105
Tenascin C 1 (IgG)	0·69 (0·50–0·95)	0·148
Tenascin C 1 (IgA1)	0·73 (0·54–0·98)	0·195
Vimentin 58–77 cyclic (IgG)	1·51 (1·02–2·24)	0·195
Tenascin C 1 (IgA2)	0·50 (0·22–1·15)	0·455
Vimentin 58–77 cyclic (IgA2)	1·75 (0·84–3·63)	0·530

Additional results are shown in the appendix (pp 4–5). Findings are ordered by level of significance. Cit2 refers to two separate citrulline amino acids within the peptide sequence. ILD=interstitial lung disease. FDR=false discovery rate. H=histone. *Per log-transformed unit. Bold values indicate FDR p<0·05 (and nominal p<0·001). Adjusted for significant covariates including BMI, smoking pack-years, Disease Activity Score for 28 joints with C-reactive protein, Multi-Dimensional Health Assessment Questionnaire, rheumatoid nodules, biologic disease-modifying anti-rheumatic drug use, methotrexate use, glucocorticoid use, and *MUC5B* genotype.

***Table 2:* Top associations between autoantibodies to citrullinated and native antigens and incidence of rheumatoid arthritis-associated ILD in 84 rheumatoid arthritis-associated ILD cases and 233 rheumatoid arthritis controls without ILD**

To develop a rheumatoid arthritis-associated ILD risk score, we also generated multivariable conditional logistic regression models for rheumatoid arthritis-associated ILD. However, this time we included only the top predictors that were independently associated with rheumatoid arthritis-associated ILD. To facilitate easy clinical application, we dichotomised clinical predictors and divided antibody biomarkers into tertiles based on the distribution of controls. We dichotomised each antibody as the highest tertile versus the lowest two tertiles (reference group). For biomarkers inversely associated with rheumatoid arthritis-associated ILD, we considered the lowest tertile as the risk group, and the top two tertiles as the reference group. We also constructed a score without biomarkers for clinical encounters where biomarkers are not available. Although the *MUC5B* promoter variant is not available clinically, it is already an established risk factor for rheumatoid arthritis-associated ILD. Thus, we included the established rheumatoid arthritis-associated ILD risk factors model that also included the *MUC5B* promoter variant. Unconditional logistic regression was used for AUCs since some of the model components were also matching factors. We compared the AUCs of individual models using DeLong's statistic. Once constructed, we calculated predicted probabilities of developing rheumatoid arthritis-associated ILD via unconditional logistic regression with the offset function, using the full BRASS cohort as the reference population (n=1581).[22] We chose rheumatoid arthritis-associated ILD probability cutoffs of 30% (expected low sensitivity), 50%, and 80% (expected high specificity) as different thresholds that could be implemented clinically.

There were no missing data for any of the biomarkers or covariates except for 15 cases and two controls missing

the *MUC5B* genotype. We imputed these values using multiple imputation with all covariates and case or control status as predictors. For the antibody biomarker analysis, an FDR-corrected p value less than 0·05 was considered significant. The threshold for statistical significance in the rest of the study was a two-sided p value less than 0·05. We prespecified all analyses in our protocol and did them in SAS (version 9.4).

Role of the funding source

The funder had no role in study design, data collection, data analysis, or data interpretation; writing the report; or the decision to submit this manuscript for publication.

See Online for appendix

Results

We identified 84 incident cases of rheumatoid arthritis-associated ILD (mean age 67 [SD 10] years, 65 [77%] female and 19 [23%] male, 76 [90%] White) and matched them to 233 controls (mean age 66 [11] years, 186 [80%] female and 47 [20%] male, 219 [94%] White; table 1). Blood was obtained at study visits a median of 1·5 (IQR 0·6–2·5) years before the index date of rheumatoid arthritis-associated ILD onset for cases and a median of 1·9 (IQR 1·6–5·6) years before the assigned index date for controls. In univariate analyses, cases had a higher BMI, smoking history, disease activity, rheumatoid nodules, and current glucocorticoid use, along with lower current biologic DMARD and methotrexate use, than controls (table 1). There were no differences in rheumatoid factor or anti-CCP status or levels between cases and controls. Rheumatoid arthritis-associated ILD subtypes and clinical characteristics are summarised in the appendix (p 3).

After adjusting for the above covariates plus *MUC5B* genotype, three fine-specificity ACPAs were associated with the risk of rheumatoid arthritis-associated ILD, with an FDR of 5% (table 2). The top two were IgA2 antibodies to citrullinated histone antigens (histone 4 33–48: aOR 0·08 [95% CI 0·03–0·22] per log-transformed unit; histone 2A/a-2 1–20: 4·03 [2·03–8·00]). The third was IgG to citrullinated cyclic filaggrin 48–65 (aOR 3·47 [95% CI 1·71–7·01]), although IgA antibodies to citrullinated cyclic filaggrin also showed an association (table 2). The associations between the remaining 68 ACPAs and rheumatoid arthritis-associated ILD are shown in the appendix (pp 4–5).

Three anti-native (non-citrullinated) antibodies were also associated with rheumatoid arthritis-associated ILD (table 2). Once again, the top two were IgA2 antibodies to histone antigens (histone 2A/a 1–20 cyclic: aOR 5·52 [95% CI 2·38–12·78]; histone 2A/a-2 1–20: 4·60 [2·18–9·74]), with the third representing IgG to cyclic filaggrin 48–65 (2·53 [1·47–4·34]). Antibodies to these three antigens were associated with rheumatoid arthritis-associated ILD in both their native and citrullinated forms. The association of the remaining 29 antibodies to native antigens is shown in the appendix (pp 6). Unadjusted associations are shown in the appendix (pp 7–9). The results of autoantibody associations for cellular non-specific interstitial pneumonia (n=44) and usual interstitial pneumonia or fibrotic non-specific interstitial pneumonia (n=29) rheumatoid arthritis-associated ILD subtypes are shown in the appendix pp 10–19). Results of ROC curves for individual antibodies and those grouped as citrullinated or native proteins are shown in the appendix (p 37). Results of analyses stratified by seropositive or seronegative rheumatoid arthritis and anti-CCP-positive or anti-CCP-negative rheumatoid arthritis are shown in the appendix (pp 20–34). Results of the analysis subtracting non-citrullinated from citrullinated biomarkers are shown in the appendix (pp 35–36).

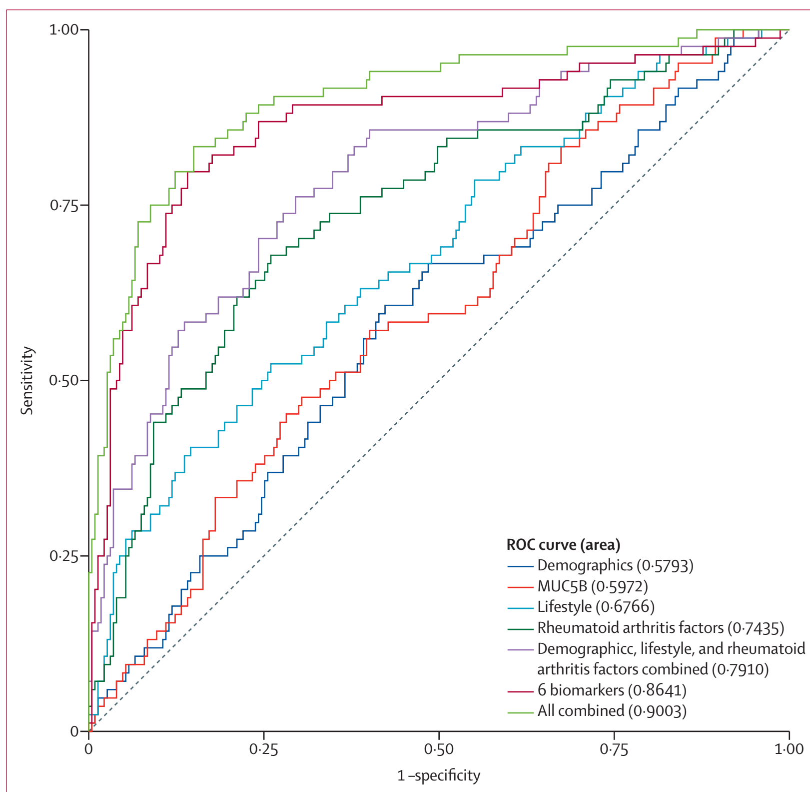

***Figure:* Comparison of ROC curves for incident rheumatoid arthritis-associated ILD risk among 84 cases and 233 rheumatoid arthritis controls without ILD, adding novel biomarkers and *MUC5B* to clinical factors**
ROC=receiver operating characteristic. ILD=interstitial lung disease. The area under the curve is shown in parentheses next to each model name. Demographic refers to age, race, and ethnicity, and education. *MUC5B* refers to the *MUC5B* promoter variant genotype. Lifestyle factors refer to BMI and smoking pack-years. Rheumatoid arthritis (RA) factors correspond to rheumatoid arthritis duration, rheumatoid factor, anti-cyclic citrullinated peptide (CCP) status, bone erosion, rheumatoid nodules, methotrexate use, glucocorticoid use, biologic disease-modifying anti-rheumatic drug use, Disease Activity Score for 28 joints with CRP, and the Multi-Dimensional Health Assessment Questionnaire. Clinical factors combined refer to demographic plus lifestyle plus rheumatoid arthritis factors. Six novel biomarkers refer to H4 33–48 citrullinated 39 (IgA2), H2A/a-2 1–20 citrullinated (IgA2), filaggrin 48–65 cit2 cyclic (IgG), H2A/a 1–20 cyclic (IgA2), H2A/a-2 1–20 (IgA2), and filaggrin 48–65 cyclic (IgG).

The figure shows ROC curves for predicting the risk of incident rheumatoid arthritis-associated ILD when adding the six novel fine-specificity protein antibody biomarkers (three ACPAs and three anti-native protein

antibodies) described above. We then calculated the AUC for these curves, including AUC corrected for optimism (table 3). 95% CIs between the clinical factors model and all factors combined model were non-overlapping. Combining all clinical factors including demographic, lifestyle, and rheumatoid arthritis factors (including anti-CCP) together achieved an optimism-corrected AUC of 0·73 for the risk of rheumatoid arthritis-associated ILD. After adding biomarkers including *MUC5B* and the six novel biomarkers from this study, the optimism-corrected AUC increased to 0·84. However, nearly all the predictive capacity came from the six novel biomarkers alone (AUC 0·84; table 3). Comparisons between AUCs of individual models are shown in the appendix (p 39).

Using the above novel biomarkers plus known clinical risk factors for rheumatoid arthritis-associated ILD, we developed a risk score for development of rheumatoid arthritis-associated ILD (table 4). Although we showed above that biomarkers alone are predictive of the risk of rheumatoid arthritis-associated ILD, we also included a rheumatoid arthritis-associated ILD risk score without biomarkers to facilitate use in clinical settings where such testing is not possible (table 4). Model coefficients and odds ratios used for the development of this risk score are shown in the appendix (p 40). Applying the clinical score (without biomarkers) in the BRASS dataset, cases scored a median of 1·6 (IQR 1·2–4·3) and controls scored a median of 1·0 (0·0–1·2). Applying the biomarker score to the BRASS dataset, cases scored a median of 6·8 (IQR 5·7–7·7) and controls scored a median of 3·0 (1·3–4·3), indicating good separation between cases and controls.

Finally, we calculated the predicted probability of developing rheumatoid arthritis-associated ILD by risk score level (table 5). The score with biomarkers showed a higher sensitivity and specificity for rheumatoid arthritis-associated ILD than did the score without biomarkers for all predicted probabilities of rheumatoid arthritis-associated ILD (table 5). However, both scores showed high specificity (≥93%) for developing rheumatoid arthritis-associated ILD at the threshold of 50% predicted probability (2·6 for the clinical score and 5·9 for the biomarker score; table 5).

	Apparent AUC (95% CI)	Optimism	Optimism-corrected AUC (95% CI)
Demographic*	0·58 (0·51–0·65)	0·07	0·51 (0·45–0·57)
MUC5B promoter variant	0·60 (0·53–0·67)	0·05	0·55 (0·48–0·61)
Lifestyle†	0·68 (0·62–0·75)	0·04	0·64 (0·58–0·70)
Rheumatoid arthritis factors‡	0·74 (0·68–0·81)	0·06	0·68 (0·62–0·74)
Clinical factors§ combined	0·79 (0·73–0·85)	0·06	0·73 (0·67–0·78)
Established rheumatoid arthritis-associated ILD factors¶	0·80 (0·74–0·85)	0·07	0·73 (0·67–0·78)
Novel biomarkers‖ (n=6)	0·86 (0·81–0·92)	0·03	0·84 (0·79–0·89)
All factors combined	0·90 (0·85–0·94)	0·06	0·84 (0·80–0·88)

AUC=area under the curve. ILD=interstitial lung disease. *Age, race and ethnicity, and education. †BMI and smoking pack-years. ‡Rheumatoid arthritis duration, rheumatoid factor, anti-cyclic citrullinated peptide, bone erosion, rheumatoid nodules, methotrexate use, glucocorticoid use, biologic disease-modifying anti-rheumatic drug use, Disease Activity Score for 28 joints with C-reactive protein, and Multi-Dimensional Health Assessment Questionnaire. §Demographic plus lifestyle plus rheumatoid arthritis factors. ¶Demographic plus lifestyle plus rheumatoid arthritis factors plus *MUC5B*. ‖H4 33–48 citrullinated 39 (IgA2), H2A/a-2 1–20 citrullinated (IgA2), filaggrin 48–65 cit2 cyclic (IgG), H2A/a 1–20 cyclic (IgA2), H2A/a-2 1–20 (IgA2), filaggrin 48–65 cyclic (IgG); cit2 refers to two separate citrulline amino acids within the peptide sequence.

***Table 3:* Optimism-corrected AUC for predicting risk of rheumatoid arthritis-associated ILD by adding novel biomarkers and the *MUC5B* promoter variant to clinical factors**

	Score points without biomarkers	Score points with biomarkers
Smoking pack-years ≥30	1·9	2·3
DAS28-CRP ≥3·2	1·2	1·2
Current glucocorticoid use	1·0	1·2
BMI ≥30 kg/m²	0·4	0·8
Filaggrin 48–65 citrullinated 2 cyclic (IgG), highest tertile	··	2·0
H4 33–48 citrullinated 39 (IgA2), lowest tertile	··	1·9
H2A/a-2 1–20 (IgA2), highest tertile	··	1·8
MUC5B promoter variant present	··	0·1

ILD=interstitial lung disease. DAS28-CRP=Disease Activity Score for 28 joints with CRP.

***Table 4:* Risk score for developing rheumatoid arthritis-associated ILD with and without biomarkers**

	Without biomarkers			With biomarkers		
	Score	Sensitivity (%)	Specificity (%)	Score	Sensitivity (%)	Specificity (%)
30% probability of rheumatoid arthritis-associated ILD	1·6	62%	79%	5·0	83%	87%
50% probability of rheumatoid arthritis-associated ILD	2·6	25%	93%	5·9	67%	94%
80% probability of rheumatoid arthritis-associated ILD	4·1	7%	99%	7·3	34%	98%

ILD=interstitial lung disease.

***Table 5:* Performance characteristics of incident rheumatoid arthritis-associated ILD risk score by thresholds of predicted probability**

Discussion

This study identified several fine-specificity ACPAs and anti-native protein antibodies associated with the development of rheumatoid arthritis-associated ILD. For example, antibodies against cyclic filaggrin and histone 2A were associated with a higher risk of rheumatoid arthritis-associated ILD. By contrast, antibodies against histone 4 were inversely associated with rheumatoid arthritis-associated ILD, suggesting that ACPA profiles might identify distinct clinical trajectories for patients with rheumatoid arthritis. Many of the associations were with IgA, suggesting a link between mucosal autoantibody production and rheumatoid arthritis-associated ILD pathogenesis. These biomarkers substantially improved prediction of rheumatoid arthritis-associated ILD beyond clinical factors (including anti-CCP) and *MUC5B*. Second, we used these novel biomarkers to develop a risk score for incident rheumatoid arthritis-associated ILD, which showed high sensitivity and specificity at several thresholds. These findings can be used to help predict

future development of clinically significant rheumatoid arthritis-associated ILD, thus potentially allowing for earlier recognition and treatment of this common and life-threatening disease among patients with rheumatoid arthritis.

Several fine-specificity ACPAs were associated with the risk of incident rheumatoid arthritis-associated ILD. Although one previous study showed that the number of fine-specificity ACPAs is associated with the risk of rheumatoid arthritis-associated ILD, this study exclusively examined subclinical ILD in a cross-sectional manner and did not test each ACPA separately.[3] The present study, by contrast, shows that the strength of the association varies significantly for each fine-specificity ACPA, with an association observed for antibodies to histone 4, histone 2A, and filaggrin. Anti-citrullinated histone is not part of the traditional anti-CCP assay. Nevertheless, citrullinated histones are associated with rheumatoid arthritis[23] and are arthritogenic in mice.[24] The protective effect seen for the antibody to citrullinated histone 4 might result from a genetic polymorphism in the peptidylarginine deaminase 4 (*PADI4*) gene responsible for citrullinating histones, leading to decreased anti-native and anti-citrullinated histone antibodies as well as an increased risk of rheumatoid arthritis.[25] Meanwhile, the association between anti-citrullinated filaggrin and rheumatoid arthritis-associated ILD also has biological plausibility given its sensitivity and specificity for diagnosis of rheumatoid arthritis.[26] Although filaggrin is a protein that binds keratin fibres in epithelial cells primarily in skin, it is also expressed in the salivary glands, explaining the higher anti-citrullinated filaggrin concentrations in individuals with subgingival *Porphyromonas gingivalis* and thereby its connection to rheumatoid arthritis.[27] Genetic variants resulting in filaggrin deficiency are also associated with asthma,[28] a disease implicated in the risk of rheumatoid arthritis.[29] Furthermore, filaggrin-deficient mice have spontaneous pulmonary inflammation.[30] As hypothesised and suggested previously,[8] anti-citrullinated fibrinogen was also associated with rheumatoid arthritis-associated ILD. Autophagy generates citrullinated peptides in human synoviocytes and has also been implicated in idiopathic pulmonary fibrosis.[31,32] Shortened telomeres are associated with an increased risk of fibrotic lung diseases including rheumatoid arthritis-associated ILD,[33] perhaps through accelerated biologic ageing. Histones are important components of telomeres, which might explain some of the findings implicating antibodies against histones in rheumatoid arthritis-associated ILD. Another mechanism relates to netosis, a form of cell death occurring in inflammatory conditions where nuclear contents, including histones, are externally exposed. Netosis has been linked to cardiac and pulmonary fibrosis.[34] Thus, mechanistic studies should be done to link autophagy, telomere shortening, and netosis in the pathogenesis of rheumatoid arthritis-associated ILD. The improved prediction of rheumatoid arthritis-associated ILD with such antibody biomarkers demonstrates the importance of incorporating these biomarkers into methods for prediction of this disease.

Using the top biomarkers for rheumatoid arthritis-associated ILD discovered in this study, we developed a well-performing risk score for this disease. To the best of our knowledge, this is the first risk score for predicting future development of rheumatoid arthritis-associated ILD based on prospectively collected data, as other scores have used cross-sectional data to predict an outcome of subclinical rheumatoid arthritis-associated ILD.[11,12] The score we developed with biomarkers had high sensitivity (>80%) at the threshold of 30% predicted probability of rheumatoid arthritis-associated ILD, whereas the score without biomarkers did not. However, both scores showed high specificity (≥93%) for predicting rheumatoid arthritis-associated ILD at the score thresholds corresponding to a predicted probability of 50% and higher for rheumatoid arthritis-associated ILD. Notably, 50% of patients with scleroderma also have ILD,[35] and guidelines already recommend baseline high-resolution chest CT screening for ILD in this population. The combination of the high mortality of rheumatoid arthritis-associated ILD,[1,2] the high specificity of the risk score, and the recommendation for screening in comparable populations underscores the importance of validating and then implementing such risk scoring for obtaining chest CT screening in clinical practice for detection of rheumatoid arthritis-associated ILD.

Although we hypothesised the association of fine-specificity ACPAs with rheumatoid arthritis-associated ILD and their ability to improve disease prediction, we also observed a few unanticipated findings. First, IgA antibodies generally showed an association with rheumatoid arthritis-associated ILD rather than most IgG antibodies. However, previous work has also shown an association of IgA rather than IgG antibodies for both rheumatoid arthritis and rheumatoid arthritis-associated ILD,[5] supporting the growing recognition of the importance of mucosal surfaces in rheumatoid arthritis pathogenesis and autoantibody production. Second, anti-citrullinated enolase was not associated with rheumatoid arthritis-associated ILD as it had been in previous studies,[6] although those studies did not adjust for some potential confounders. Finally, antibodies to several non-citrullinated (native) antigens were also associated with rheumatoid arthritis-associated ILD. Although not studied as often as ACPAs, antibodies to native antigens have previously shown an association with rheumatoid arthritis, including anti-filaggrin (with rheumatoid arthritis)[36] and anti-fibrinogen (with rheumatoid arthritis-associated ILD).[8] These studies support our findings and encourage future exploration of biomarkers to non-citrullinated antigens.

Strengths of this study include the use of prospectively collected exposure data and prediction of incident rheumatoid arthritis-associated ILD, an area of considerable unmet need. Additionally, this study verified both rheumatoid arthritis and rheumatoid arthritis-associated ILD cases, included a comprehensive panel of antibodies, and adjusted for many important confounders. There are also several important limitations to consider. The study population was neither population-based nor diverse, which is important as ACPA fine specificities vary in different populations.[37] Most participants were female with longstanding rheumatoid arthritis, so the results might not be generalisable to male patients or those with newly diagnosed rheumatoid arthritis. The number of incident cases was relatively low, which might have hampered the ability to detect true associations, such as with the *MUC5B* genotype[4] or anti-citrullinated enolase.[6] Larger studies are needed to identify antibody signatures for specific rheumatoid arthritis-associated ILD subtypes as well as among patients with classically seropositive and seronegative rheumatoid arthritis. It is possible that some of the controls might have had subclinical rheumatoid arthritis-associated ILD. However, this would bias the results towards the null. Alternatively, we could have required controls to have had a clinically indicated chest CT without ILD, but this might have introduced bias due to the indication of the chest CT, and these patients might not be representative of the general rheumatoid arthritis population. We chose controls without chronic lung disease from the general rheumatoid arthritis population so that findings could be relevant to this population and to lower the chances of confounding by indication of imaging and misclassification. Additionally, this study included rheumatoid arthritis meeting classification criteria and rheumatoid arthritis-associated ILD cases independently verified by review of CT images by at least two experts to avoid the risk of misclassification. Future studies should investigate whether antibody profiling might be helpful in identifying patients at risk of rheumatoid arthritis-associated ILD by incorporating other control groups such as healthy individuals, patients with rheumatoid arthritis who had chest CT imaging done, patients with rheumatoid arthritis and other forms of chronic lung disease, and serial chest CT imaging in a prospective research protocol.

It is possible that participants in the BRASS cohort might not be reflective of the general rheumatoid arthritis population.[38] However, the nested case–control study design protects against possible differences in attrition since even those who only participated at baseline could be included in the study as the outcome was determined through data from routine clinical care. Because of the matched case–control design, the rheumatoid arthritis prediction curves and scores in reported this study could not fully incorporate the predictive performance of the matching factors, potentially reducing model performance. It is also possible that such performance metrics might have been artificially elevated because of overfitting, although we did optimism-corrected analyses to account for this overfitting. We did not update covariates after the date of blood collection since this might have mediated the associations between circulating antibodies and the risk of rheumatoid arthritis-associated ILD. Thus, covariates do not reflect changes that could have occurred between the blood draw date and onset of rheumatoid arthritis-associated ILD. However, using the date before disease onset made it unlikely that rheumatoid arthritis-associated ILD influenced variables such as medications that can be either prescribed or avoided in rheumatoid arthritis-associated ILD.

There were some overlapping findings of autoantibodies to citrullinated and native antibodies that could be explained by cross-reactivity of the assay or they could be due to separate antibodies. However, analyses comparing individual targets and those grouped by citrullinated and native targets suggested that most biomarkers provided a unique prediction ability, which was generally higher for antibodies to citrullinated antigens. We found that several antibodies that were associated with rheumatoid arthritis-associated ILD were of the IgA2 isotype. Although Sokolova and colleagues[39] also found this isotype to be important in rheumatoid arthritis, other reports have been inconclusive, and it might be more difficult to detect circulating concentrations of this isotype.[40] The assay did not include other antibodies such as anti-CEP1 and anti-HSP90 that might have an important role in rheumatoid arthritis-associated ILD.[6,7] Future studies should incorporate other anti-modified antibodies into other processes, such as carbamylation and secretory ACPAs, to more accurately predict the risk of rheumatoid arthritis-associated ILD.[41,42] We did many tests, so some findings could have been due to chance. However, all analyses were prespecified in our protocol, and we used FDR for the antibody analysis. At the same time, we did not test all antibodies to citrullinated antigens that might have an important role in rheumatoid arthritis-associated ILD, such as anti-HSP90.[7] Since we had a relatively low sample size, we relied on optimism-correction to perform internal validation rather than other methods such as splitting the sample into derivation and validation sets. Future work should validate the findings from this study in external populations and determine the duration of validity before ILD onset. Studies are also needed to implement screening strategies and to test whether early treatment could improve clinical outcomes.

In conclusion, both fine-specificity ACPAs and anti-native protein antibodies are associated with the risk of rheumatoid arthritis-associated ILD and improve its prediction. Given the high mortality associated with rheumatoid arthritis-associated ILD and the high specificity of the risk score developed in this study, the clinical utility of screening high-risk patients with rheumatoid arthritis for ILD should be urgently investigated.

Contributors

VLK, TJD, and JAS contributed to the conception and design of the study. VLK, RRG, HH, MN, RB, WHR, JS, MEW, NAS, TJD, and JAS prepared material and collected data. Data were analysed by VLK, KH, KY, WH, and JAS. Data were interpreted by VLK, KH, KY, JMD, GCM, WH, PFD, JC, VF, RRG, HH, MN, RB, CSC, WHR, JS, KPL, MEW, NAS, TJD, and JAS. KH and JAS directly accessed and verified the underlying data. The first draft of the manuscript was written by VLK and JAS. All authors had full access to all the data in the study, read and approved the final version of the manuscript, and had final responsibility for the decision to submit for publication.

Declaration of interests

KY has received consulting fees from OM1 unrelated to this work. JMD has received research support from Pfizer, has a patent pending for assessing and treating arthritis and serves on a data safety monitoring board for a rheumatoid arthritis clinical trial sponsored by the US National Institutes of Health, all unrelated to this work. PFD has received consulting fees from Boehringer Ingelheim, Bristol Myers Squibb, and Genentech, and receives royalties from UpToDate unrelated to this work. HH has received research support from Canon Medical Systems and Konica-Minolta as well as consulting fees from Canon Medical Systems and the Mitsubishi Chemical Company unrelated to this work. MN has received research support from AstraZeneca, Canon Medical Systems, and Daiichi Sankyo, and consulting fees from AstraZeneca and Daiichi Sankyo unrelated to this work. JS is currently an employee at GlaxoSmithKline and owns shares in GlaxoSmithKline, and his work on this project preceded this employment. MEW has received research support from Amgen, Bristol Myers Squibb, Eli Lilly, Aqtual, and Janssen; consultancy fees from AbbVie, Aclaris, Amgen, Aqtual, Bayer, Bristol Myers Squibb, CorEvitas, EqRX, Genosco, GlaxoSmithKline, Gilead, Johnson & Johnson, Kyvrena, Eli Lilly, Pfizer, Rani, Revolo, Sanofi, Scipher, SciRom, SetPoint, and Tremeau; and stock options from Canfite, Inmedex, and Scipher, all unrelated to this work. NAS has received research grants from AbbVie, Amgen, Aqtual, Bristol Myers Squibb, Eli Lilly, and Mallinckrodt unrelated to this work. TJD has received research support from Bristol Myers Squibb; consulting fees from Boehringer Ingelheim and L.E.K. consulting; speaking fees and travel support from Aura; and has been part of a clinical trial funded by Genentech, unrelated to this work. JAS has received research support from Bristol Myers Squibb and done consultancy work for AbbVie, Amgen, Boehringer Ingelheim, Bristol Myers Squibb, Gilead, Inova Diagnostics, Janssen, Optum, and Pfizer, all unrelated to this work. All other authors declare no competing interests.

Data sharing

Data are not available without approval from the Brigham Rheumatoid Arthritis Sequential Study Scientific Advisory Board. Requests should be sent via email to the corresponding author.

Acknowledgments

This work was funded by the US National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant number K23 AR069688 to JAS). TJD is supported by the US National Heart, Lung, and Blood Institute (grant number R01 HL155522). GCM is supported by the US National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant number T32 AR007530). HH is supported by the US National Cancer Institute (grant numbers R01 CA203636 and U01 CA209414) and the US National Heart, Lung, and Blood Institute (grant numbers R01 HL111024, R01 HL135142, and R01 HL130974). JAS is also supported by the US National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant numbers R01 AR077607, P30 AR070253, and P30 AR072577), the R. Bruce and Joan M. Mickey Research Scholar Fund, and the Llura Gund Award for Rheumatoid Arthritis Research and Care. This content is solely the responsibility of the authors and does not necessarily represent the official views of Harvard University, its affiliated academic health care centres, or the US National Institutes of Health.

References

1 Bongartz T, Nannini C, Medina-Velasquez YF, et al. Incidence and mortality of interstitial lung disease in rheumatoid arthritis: a population-based study. *Arthritis Rheum* 2010; **62:** 1583–91.

2 Hyldgaard C, Hilberg O, Pedersen AB, et al. A population-based cohort study of rheumatoid arthritis-associated interstitial lung disease: comorbidity and mortality. *Ann Rheum Dis* 2017; **76:** 1700–06.

3 Giles JT, Danoff SK, Sokolove J, et al. Association of fine specificity and repertoire expansion of anticitrullinated peptide antibodies with rheumatoid arthritis associated interstitial lung disease. *Ann Rheum Dis* 2014; **73:** 1487–94.

4 Juge PA, Lee JS, Ebstein E, et al. *MUC5B* promoter variant and rheumatoid arthritis with interstitial lung disease. *N Engl J Med* 2018; **379:** 2209–19.

5 García-Moreno C, Gómara MJ, Castellanos-Moreira R, Sanmartí R, Haro I. Peptides bearing multiple post-translational modifications as antigenic targets for severe rheumatoid arthritis patients. *Int J Mol Sci* 2021; **22:** 13290.

6 Alunno A, Bistoni O, Pratesi F, et al. Anti-citrullinated alpha enolase antibodies, interstitial lung disease and bone erosion in rheumatoid arthritis. *Rheumatology* 2018; **57:** 850–55.

7 Harlow L, Rosas IO, Gochuico BR, et al. Identification of citrullinated hsp90 isoforms as novel autoantigens in rheumatoid arthritis-associated interstitial lung disease. *Arthritis Rheum* 2013; **65:** 869–79.

8 Liao KP, Sparks JA, Hejblum BP, et al. Phenome-wide association study of autoantibodies to citrullinated and noncitrullinated epitopes in rheumatoid arthritis. *Arthritis Rheumatol* 2017; **69:** 742–49.

9 Kronzer VL, Huang W, Dellaripa PF, et al. Lifestyle and clinical risk factors for incident rheumatoid arthritis-associated interstitial lung disease. *J Rheumatol* 2021; **48:** 656–63.

10 Doyle TJ, Patel AS, Hatabu H, et al. Detection of rheumatoid arthritis-interstitial lung disease is enhanced by serum biomarkers. *Am J Respir Crit Care Med* 2015; **191:** 1403–12.

11 Juge PA, Granger B, Debray MP, et al. A risk score to detect subclinical rheumatoid arthritis-associated interstitial lung disease. *Arthritis Rheumatol* 2022; **74:** 1755–65.

12 Qin Y, Wang Y, Meng F, et al. Identification of biomarkers by machine learning classifiers to assist diagnose rheumatoid arthritis-associated interstitial lung disease. *Arthritis Res Ther* 2022; **24:** 115.

13 Iannaccone CK, Lee YC, Cui J, et al. Using genetic and clinical data to understand response to disease-modifying anti-rheumatic drug therapy: data from the Brigham and Women's Hospital Rheumatoid Arthritis Sequential Study. *Rheumatology* 2011; **50:** 40–46.

14 Doyle TJ, Dellaripa PF, Batra K, et al. Functional impact of a spectrum of interstitial lung abnormalities in rheumatoid arthritis. *Chest* 2014; **146:** 41–50.

15 Sparks JA, He X, Huang J, et al. Rheumatoid arthritis disease activity predicting incident clinically-apparent RA-associated interstitial lung disease: a prospective cohort study. *Arthritis Rheumatol* 2019; **71:** 1472–82.

16 Monach PA, Hueber W, Kessler B, et al. A broad screen for targets of immune complexes decorating arthritic joints highlights deposition of nucleosomes in rheumatoid arthritis. *Proc Natl Acad Sci USA* 2009; **106:** 15867–72.

17 Sokolove J, Bromberg R, Deane KD, et al. Autoantibody epitope spreading in the pre-clinical phase predicts progression to rheumatoid arthritis. *PLoS One* 2012; **7:** e35296.

18 Arkema EV, Goldstein BL, Robinson W, et al. Anti-citrullinated peptide autoantibodies, human leukocyte antigen shared epitope and risk of future rheumatoid arthritis: a nested case-control study. *Arthritis Res Ther* 2013; **15:** R159.

19 Wagner CA, Sokolove J, Lahey LJ, et al. Identification of anticitrullinated protein antibody reactivities in a subset of anti-CCP-negative rheumatoid arthritis: association with cigarette smoking and HLA-DRB1 'shared epitope' alleles. *Ann Rheum Dis* 2015; **74:** 579–86.

20 McDermott G, Gill R, Gagne S, et al. Associations of the *MUC5B* promoter variant with timing of interstitial lung disease and rheumatoid arthritis onset. *Rheumatology* 2022; **61:** 4915–23.

21 Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001; **54:** 774–81.

22 Qian J, Payabvash S, Kemmling A, Lev MH, Schwamm LH, Betensky RA. Variable selection and prediction using a nested, matched case-control study: application to hospital acquired pneumonia in stroke patients. *Biometrics* 2014; **70:** 153–63.

23 Meng X, Ezzati P, Smolik I, Bernstein CN, Hitchon CA, El-Gabalawy HS. Characterization of autoantigens targeted by anti-citrullinated protein antibodies in vivo: prominent role for epitopes derived from histone 4 proteins. *PLoS One* 2016; **11:** e0165501.

24 Sohn DH, Rhodes C, Onuma K, et al. Local Joint inflammation and histone citrullination in a murine model of the transition from preclinical autoimmunity to inflammatory arthritis. *Arthritis Rheumatol* 2015; **67:** 2877–87.

25 Mergaert AM, Bawadekar M, Nguyen TQ, et al. Reduced anti-histone antibodies and increased risk of rheumatoid arthritis associated with a single nucleotide polymorphism in PADI4 in North Americans. *Int J Mol Sci* 2019; **20:** 3093.

26 Won P, Kim Y, Jung H, et al. Pathogenic role of circulating citrullinated antigens and anti-cyclic monoclonal citrullinated peptide antibodies in rheumatoid arthritis. *Front Immunol* 2021; **12:** 692242.

27 Mikuls TR, Payne JB, Yu F, et al. Periodontitis and *Porphyromonas gingivalis* in patients with rheumatoid arthritis. *Arthritis Rheumatol* 2014; **66:** 1090–100.

28 Rice NE, Patel BD, Lang IA, et al. Filaggrin gene mutations are associated with asthma and eczema in later life. *J Allergy Clin Immunol* 2008; **122:** 834–36.

29 Kronzer VL, Westerlind H, Alfredsson L, et al. Respiratory diseases as risk factors for seropositive and seronegative rheumatoid arthritis and in relation to smoking. *Arthritis Rheumatol* 2021; **73:** 61–68.

30 Saunders SP, Moran T, Floudas A, et al. Spontaneous atopic dermatitis is mediated by innate immunity, with the secondary lung inflammation of the atopic march requiring adaptive immunity. *J Allergy Clin Immunol* 2016; **137:** 482–91.

31 Sharma P, Alizadeh J, Juarez M, et al. Autophagy, apoptosis, the unfolded protein response, and lung function in idiopathic pulmonary fibrosis. *Cells* 2021; **10:** 1642.

32 Celia AI, Colafrancesco S, Barbati C, Alessandri C, Conti F. Autophagy in rheumatic diseases: role in the pathogenesis and therapeutic approaches. *Cells* 2022; **11:** 1359.

33 Natalini JG, England BR, Baker JF, et al. Associations between shortened telomeres and rheumatoid arthritis-associated interstitial lung disease among U.S. Veterans. *Respir Med* 2022; **201:** 106943.

34 Dimmeler S, Zeiher AM. Netting insights into fibrosis. *N Engl J Med* 2017; **376:** 1475–77.

35 Bergamasco A, Hartmann N, Wallace L, Verpillat P. Epidemiology of systemic sclerosis and systemic sclerosis-associated interstitial lung disease. *Clin Epidemiol* 2019; **11:** 257–73.

36 Choi KH, Lee EB, Yoo CD, et al. Clinical significance of anti-filaggrin antibody recognizing uncitrullinated filaggrin in rheumatoid arthritis. *Exp Mol Med* 2005; **37:** 546–52.

37 Too CL, Murad S, Hansson M, et al. Differences in the spectrum of anti-citrullinated protein antibody fine specificities between Malaysian and Swedish patients with rheumatoid arthritis: implications for disease pathogenesis. *Arthritis Rheumatol* 2017; **69:** 58–69.

38 Iannaccone CK, Fossel A, Tsao H, Cui J, Weinblatt M, Shadick N. Factors associated with attrition in a longitudinal rheumatoid arthritis registry. *Arthritis Care Res* 2013; **65:** 1183–89.

39 Sokolova MV, Hagen M, Bang H, et al. IgA anti-citrullinated protein antibodies are associated with flares during DMARD tapering in rheumatoid arthritis. *Rheumatology* 2022; **61:** 2124–31.

40 Derksen V, Allaart CF, Van der Helm-Van Mil AHM, Huizinga TWJ, Toes REM, van der Woude D. In rheumatoid arthritis patients, total IgA1 and IgA2 levels are elevated: Implications for the mucosal origin hypothesis. *Rheumatology* 2022; **62:** 407–16.

41 Castellanos-Moreira R, Rodríguez-García SC, Gomara MJ, et al. Anti-carbamylated proteins antibody repertoire in rheumatoid arthritis: evidence of a new autoantibody linked to interstitial lung disease. *Ann Rheum Dis* 2020; **79:** 587–94.

42 Roos Ljungberg K, Joshua V, Skogh T, et al. Secretory anti-citrullinated protein antibodies in serum associate with lung involvement in early rheumatoid arthritis. *Rheumatology* 2020; **59:** 852–59.