Arthritis & Rheumatology
Vol. 77, No. 8, August 2025, pp 971–980
DOI 10.1002/art.43149

American College of Rheumatology
Empowering Rheumatology Professionals

Peripheral Biomarker Signatures in Rheumatoid Arthritis–Associated Interstitial Lung Disease

Austin M. Wheeler,[1] Joshua F. Baker,[2] Thomas R. Riley IV,[2] Tate M. Johnson,[1] Yangyuna Yang,[3] Punyasha Roul,[3] Katherine D. Wysham,[4] Grant W. Cannon,[5] Gary Kunkel,[5] Gail Kerr,[6] Dana P. Ascherman,[7] Paul Monach,[8] Andreas Reimold,[9] Scott Matson,[10] Jill A. Poole,[3] Michael J. Duryee,[1] Geoffrey M. Thiele,[1] Ted R. Mikuls,[1] and Bryant R. England[1]

Objective. Rheumatoid arthritis–associated interstitial lung disease (RA-ILD) is a significant cause of morbidity and mortality among patients with RA, yet effective risk stratification for RA-ILD is lacking. We sought to characterize unique peripheral blood biomarker signatures in RA that could improve RA-ILD discrimination beyond clinical and genetic risk factors.

Methods. We performed a cross-sectional study of participants in the Veterans Affairs Rheumatoid Arthritis Registry. ILD was validated through systematic record review. Autoantibodies (n = 14), proinflammatory cytokines/chemokines (n = 36), adipokines (n = 3), alarmins (n = 3), and matrix metalloproteinases (n = 9) were measured from banked serum or plasma. *MUC5B* rs35705950 genotyping was obtained via microarray. Principal component (PC) analysis was performed to derive unique biomarker signatures. Logistic regression was used to compare models predicting RA-ILD presence using combinations of clinical, peripheral biomarker, and genetic variables.

Results. Among 2,001 participants with RA (88.7% male, mean age 63.7 years), 121 (6.4%) had RA-ILD. A total of 15 PCs were identified, with 8 significantly associated with RA-ILD after adjusting for clinical factors. These eight included biomarker themes of innate and allergic responses, autoantibodies (anti–cyclic citrullinated peptide, rheumatoid factor, anti–malondialdehyde-acetaldehyde adduct), adipokines, alarmins, tissue remodeling, and neutrophil chemotaxis. Models with PCs (area under the receiver operating characteristic curve [AUC] 0.739) and PCs with *MUC5B* (AUC 0.751) outperformed clinical risk factors alone (AUC 0.630; $P < 0.001$).

Conclusion. Peripheral biomarker signatures are associated with RA-ILD and improve RA-ILD identification beyond clinical risk factors. In addition to demonstrating the potential for peripheral biomarkers to aid RA-ILD risk stratification, these findings suggest diverse pathways are involved in RA-ILD pathogenesis.

INTRODUCTION

Rheumatoid arthritis–associated interstitial lung disease (RA-ILD) is a significant cause of morbidity and mortality among patients with RA.[1,2] Despite the importance of this extra-articular manifestation, there are relatively few identified risk factors for RA-ILD. Clinical risk factors consistently identified include older age, male sex, cigarette smoking, and severe articular disease.[3,4]

Presented in part at 2021 and 2024 American College of Rheumatology Convergence.

The views expressed herein are those of the authors and do not reflect the official policy of the Department of Veterans Affairs or the United States government.

Dr Wheeler's work was supported by the Rheumatology Research Foundation. Dr Baker's work was supported by the Veteran Affairs (VA) Clinical Science Research and Development (CSR&D) (merit award CX001703) and the Rehabilitation Research and Development Service (merit award RX003644). Dr Riley's work was supported by the NIH Pharmacoepidemiology (grant 5T32GM075766-17). Dr Johnson's work was supported by the Rheumatology Research Foundation. Dr Wysham's work was supported by the VA CSR&D (award IK2 CX002351). Dr Poole's work was supported by the US Department of Defense (grant PR200793) and National Institute of Occupational Safety and Health (grant R01OH012045). Dr Mikuls' work was supported by the VA Biomedical Laboratory Research and Development (merit award I01 BX003635), NIH (grant 1U54-GM-115458), US Department of Defense (grant PR200793), and the Rheumatology Research Foundation. Dr England's work was supported by the VA CSR&D (award IK2 CX002203) and Rheumatology Research Foundation.

[1]Austin M. Wheeler, MD, Tate M. Johnson, MD, Michael J. Duryee, MS, Geoffrey M. Thiele, PhD, Ted R. Mikuls, MD, MSPH, Bryant R. England, MD, PhD: VA Nebraska-Western Iowa Health Care System and University of Nebraska Medical Center, Omaha; [2]Joshua F. Baker, MD, MSCE, Thomas R. Riley IV, MD: University of Pennsylvania and Corporal Michael J. Crescenz VA Medical Center, Philadelphia; [3]Yangyuna Yang, PhD, MBBS, Punyasha Roul, MS, Jill A. Poole, MD: University of Nebraska Medical Center, Omaha; [4]Katherine D. Wysham, MD: VA Puget Sound Health Care System and University of Washington, Seattle; [5]Grant W. Cannon, MD, Gary Kunkel, MD: VA Salt Lake City Healthcare System and University of Utah; [6]Gail Kerr, MD: Washington VA Medical Center, Howard University, and Georgetown University, Washington, District of Columbia; [7]Dana P. Ascherman, MD: Pittsburgh VA Medical Center-University Drive and University of Pittsburgh, Pittsburgh, Pennsylvania; [8]Paul Monach, MD, PhD: Jamaica Plain VA Medical Center, Boston, Massachusetts; [9]Andreas Reimold, MD; Dallas VA Medical Center and

There are also observed associations between prevalent RA-ILD and seropositivity with rheumatoid factor (RF), anti–cyclic citrullinated peptide (CCP) antibodies, and anti–malondialdehyde-acetaldehyde adduct (MAA) antibodies.[4,5] In addition to clinical measures and autoantibody status, genetic risk factors have been increasingly recognized for their contribution to RA-ILD development, specifically when it occurs in a usual interstitial pneumonia (UIP) pattern. The strongest genetic risk factor identified to date is the *MUC5B* rs35705950 promoter variant, a finding validated in multiple studies and independent cohorts.[6–10]

Beyond clinically available autoantibodies and these genetic risk factors, there are a number of peripheral biomarkers associated with the presence of RA-ILD. A recent systematic review by our group identified 104 unique peripheral biomarkers associated with RA-ILD among 70 different studies, including a variety of cytokines/chemokines, growth factors, extracellular matrix proteins, tumor markers, lung epithelial or surfactant proteins, and others.[11] Many of these peripheral biomarkers are also recognized to be associated with RA, regardless of ILD status.[12,13] Most studies have focused on a single or select number of peripheral biomarkers for RA-ILD. The few studies that have applied more broad peripheral biomarker analyses in RA-ILD have been performed among small, selected samples.[14,15]

Given the vast range of recognized and candidate biomarkers for RA-ILD, a multidimensional evaluation of composite biomarkers for RA-ILD risk is needed, accounting for the interrelated nature across peripheral biomarker measures. Machine learning approaches have previously been leveraged to identify unique biomarker profiles in RA and other disease states. In this study, we aimed to derive unique peripheral biomarker signatures using an unsupervised machine learning approach among a large, multicenter RA cohort and test whether these biomarker signatures would improve RA-ILD prediction beyond clinical risk factors.

PATIENTS AND METHODS

Study design, population, and demographics. This cross-sectional study was conducted within the Veterans Affairs Rheumatoid Arthritis (VARA) Registry, which is a prospective, multicenter cohort of US veterans with RA fulfilling the 1987 American College of Rheumatology classification criteria that was initiated in 2003.[16,17] This study was approved by the Veterans Affairs (VA) Nebraska-Western Iowa Health Care System Institutional Review Board (IRB; 1576193), with each site receiving IRB approval and all participants providing written informed consent. Study participants and the public were not involved in the design, reporting, or dissemination plans of this research.

Clinical and demographic data, including age, sex, race and ethnicity, body mass index (BMI; in kilograms per square meter), RA disease duration, and cigarette smoking history (categorized as ever or never smoker), were recorded at the time of registry enrollment. Race and ethnicity was self-reported from a fixed set of categories and subsequently categorized for this study as non-Hispanic White or other. The Disease Activity Score With 28-Joint Count (DAS28) and Multidimensional Health Assessment Questionnaire were documented by treating rheumatologists at enrollment. Disease-modifying antirheumatic drug (DMARD) use at enrollment was determined by linkage with administrative pharmacy dispensing data obtained through the VA Corporate Data Warehouse. Participants with missing demographic or biomarker data were excluded from the analysis (ie, complete case analysis).

ILD assessment. ILD was assessed in the cohort using screening and validation steps, as previously reported.[7,8,18] All participants were screened for possible ILD based on diagnostic codes for ILD or ILD-related terms in chest computed tomography (CT) reports.[19,20] ILD diagnoses were subsequently validated by a board-certified rheumatologist based on documentation of a diagnosis of ILD in the medical record by the treating provider with supporting findings from either imaging or lung biopsy per the radiology or pathology report (chest CT in 97% of patients with ILD).[19] Diagnostic evaluation (CT, pulmonary function testing, and lung biopsy) was performed according to clinical indication at the discretion of treating providers. Given disease latency, prevalent ILD status was defined as the presence of ILD at any time before and up to two years after registry enrollment. Participants who later developed RA-ILD or who had indeterminate ILD were excluded from analyses. When available, ILD pattern was extracted from clinical radiology reports, biopsy reports, and clinical notes. The presence of honeycombing on chest CT reports was considered additional evidence of a UIP pattern.[18] Forced vital capacity (FVC) percentage predicted was extracted from medical records at the most proximate time to enrollment.

Peripheral biomarker assessment. Biomarker assessment was performed using serum and plasma that were collected at enrollment on all participants and stored at −70°C. Biomarkers were selected based on prior associations with RA and/or RA-ILD as part of previous studies within the VARA cohort.[5,15,18,21–23] Biomarker measurement was performed via either the Mesoscale

University of Texas Southwestern, Dallas; [10]Scott Matson, MD; University of Kansas, Kansas City.

Additional supplementary information cited in this article can be found online in the Supporting Information section (https://acrjournals.onlinelibrary.wiley.com/doi/10.1002/art.43149).

Author disclosures and graphical abstract are available at https://onlinelibrary.wiley.com/doi/10.1002/art.43149.

Address correspondence via email to Bryant R. England, MD, PhD, at bryant.england@unmc.edu.

Submitted for publication October 30, 2024; accepted in revised form February 20, 2025.

Table 1. Peripheral biomarkers included in principal component analysis*

Variable	Measurement method
Adipocytokine	
Adiponectin	Mesoscale discovery
FGF-21	Mesoscale discovery
Leptin	Mesoscale discovery
Alarmin/cytokine	
IL-17e–IL-25	Mesoscale discovery
IL-33	Mesoscale discovery
TSLP	Mesoscale discovery
Antibody	
Anti-MAA-albumin IgA	ELISA
Anti-MAA-albumin IgG	ELISA
Anti-MAA-albumin IgM	ELISA
Anti-MAA-collagen IgA	ELISA
Anti-MAA-collagen IgG	ELISA
Anti-MAA-collagen IgM	ELISA
Anti-MAA-fibrinogen IgA	ELISA
Anti-MAA-fibrinogen IgG	ELISA
Anti-MAA-fibrinogen IgM	ELISA
Anti-MAA-vimentin IgA	ELISA
Anti-MAA-vimentin IgG	ELISA
Anti-MAA-vimentin IgM	ELISA
IgG-CCP	ELISA
RF	Nephelometry
Chemokine	
Eotaxin	Mesoscale discovery
MCP-1	Mesoscale discovery
MCP-4	Mesoscale discovery
MDC	Mesoscale discovery
MIP-1α	Mesoscale discovery
MIP-1β	Mesoscale discovery
MIP-3α	Mesoscale discovery
TARC	Mesoscale discovery
Collagenase	
MMP1	Mesoscale discovery
MMP3	Mesoscale discovery
MMP7	Mesoscale discovery
MMP9	Mesoscale discovery
Cytokine	
FL	Mesoscale discovery
Fractalkine	Mesoscale discovery
G-CSF	Mesoscale discovery
GM-CSF	Mesoscale discovery
IFNα2a	Mesoscale discovery
IL-1RA	Mesoscale discovery
IL-1α	Mesoscale discovery
IL-1β	Mesoscale discovery
IL-2	Mesoscale discovery
IL-3	Mesoscale discovery
IL-4	Mesoscale discovery
IL-5	Mesoscale discovery
IL-6	Mesoscale discovery
IL-7	Mesoscale discovery
IL-8	Mesoscale discovery
IL-9	Mesoscale discovery
IL-10	Mesoscale discovery
IL-12–IL-23-p40	Mesoscale discovery
IL-12-p70	Mesoscale discovery
IL-13	Mesoscale discovery
IL-15	Mesoscale discovery
IL-16	Mesoscale discovery
IL-17A	Mesoscale discovery
IL-23	Mesoscale discovery
IL-27	Mesoscale discovery
IFN	Mesoscale discovery
TNFα	Mesoscale discovery
VEGF	Mesoscale discovery

* CCP, cyclic citrullinated peptide; ELISA, enzyme-linked immunosorbent assay; FGF, fibroblast growth factor; FL, Fms-like tyrosine kinase 3 ligand; G-CSF, granulocyte colony-stimulating factor; GM-CSF, granulocyte–macrophage colony-stimulating factor; IFN, interferon; IL, interleukin; IL-1RA, IL-1 receptor antagonist; MAA, malondialdehyde-acetaldehyde adduct; MCP, monocyte chemoattractant protein; MDC, macrophage-derived chemokine; MIP, macrophage inflammatory protein; MMP, matrix metalloproteinase; RF, rheumatoid factor; TARC, thymus and activation-regulated chemokine; TNF, tumor necrosis factor; TSLP, thymic stromal lymphopoietin; VEGF, vascular endothelial growth factor.

Discovery platform, enzyme-linked immunosorbent assay (ELISA), or nephelometry (Table 1). Adipokines included adiponectin, fibroblast growth factor (FGF)-21, and leptin. Alarmins included interleukin (IL)-17e–IL-25, IL-33, and thymic stromal lymphopoietin (TSLP). RA-related antibodies included anti-MAA (IgA, IgM, and IgG against MAA-adducted albumin, collagen, fibrinogen, and vimentin; ELISA), anti-CCP (IgG; ELISA), and RF (nephelometry). Chemokines included eotaxin, monocyte chemoattractant proteins, macrophage-derived chemokine, macrophage inflammatory proteins, and thymus and activation-regulated chemokine. Collagenases included matrix metalloproteinases (MMPs). Cytokines included Fms-like tyrosine kinase 3 ligand (FL), fractalkine, granulocyte colony-stimulating factor (CSF), granulocyte–macrophage CSF, interferons (IFNs), ILs, tumor necrosis factor (TNF) α, and vascular endothelial growth factor. All biomarker concentrations were log transformed and standardized for analyses, with a mean ± SD of 0 ± 1.

Genotyping. Genotyping was performed using the Infinium Global Screening Array-24 version 2.0 microarray (Illumina Inc). *MUC5B* rs35705950 was directly measured and passed quality control measures including Hardy–Weinberg equilibrium. The *MUC5B* rs35705950 promoter variant was analyzed via an additive model according to the number of thymine copies as the variant allele.[7]

Statistical analysis. Principal component analysis (PCA) was used to identify unique peripheral biomarker signatures for RA-ILD. PCA is a statistical technique that has the powerful ability to compress large amounts of predictor variables into more accessible groups as well as illustrate patterns in the data based on these variables.[24,25] All standardized, log-transformed biomarkers (Table 1) were included in the PCA. As part of PCA, a correlation matrix was constructed for the aforementioned biomarkers, eigenvalues were calculated from this matrix, and principal components (PCs) were selected based on eigenvalue (>1,

per Kaiser–Guttman rule), scree plots, and cumulative variance.[25–27] An orthogonal varimax rotation was used to improve the interpretability of extracted PCs, forcing variables to align strongly with a single component.[24] Biomarkers with loadings >0.3 were considered to meaningfully represent the respective PC.[26] Each PC was descriptively characterized based on the function of the biomarkers with a meaningful loading. The relationship of each PC with the outcome of RA-ILD was initially evaluated as a PC score within a multivariable logistic regression model that adjusted for age, sex, race and ethnicity, and smoking status.

Clinical variables, including age, sex, race and ethnicity, and smoking history, were included in a multivariable logistic regression model, evaluating the outcome of RA-ILD among the entire cohort with RA. The 15 PCs were included together in a multivariable model assessing the predictive capacity of biomarkers alone. Subsequently, a combined clinical and biomarker logistic regression model was assessed, including the clinical factors and all 15 PCs extracted from the PCA. A final logistic regression model included the combination of clinical factors, PCs, and the *MUC5B* promoter variant. The area under the receiver operating characteristic curve (AUC) was compared among the clinical model, clinical model with PCA, and clinical model with PCA and *MUC5B*, and statistical comparisons were made via the nested likelihood ratio test to evaluate for improvement in model fit. Performance of a clinical model with *MUC5B* was assessed and compared with the clinical model with PCA and *MUC5B*. As a sensitivity analysis, this process was then repeated using an expanded clinical model that included BMI, DAS28, prednisone use, and DMARD use. Additional sensitivity analysis was performed with the inclusion of only those PCs that were individually associated with RA-ILD within the models. Internal validation was performed for all models using bootstrapping with 50 different random samples, and optimism-adjusted measures were obtained.[28]

We performed subgroup analyses of individual PC associations with RA-ILD subtypes (UIP and non-UIP) and FVC severity (FVC <80% and FVC ≥80% predicted) versus without ILD via multivariable logistic regression with adjustment for age, sex, race and ethnicity, and smoking history. Logistic regression model performance was also evaluated on the subgroup with UIP. All analyses were completed with Stata version 18.

RESULTS

Patient characteristics. Among 2,001 participants with RA, 121 (6.4%) were classified as having prevalent ILD. Participants with RA-ILD were older (average age 67.4 vs 63.5 years; $P < 0.001$) and more frequently male (94.2% vs 88.3%; $P = 0.047$) compared to patients with RA without ILD (Table 2). A cigarette smoking history (84.3% vs 77.4%; $P = 0.07$) as well as positivity for RF (86.8% vs 76.3%; $P = 0.01$) and anti-CCP (83.5% vs 76.6%; $P = 0.08$) were more frequent in those with ILD. In patients with RA-ILD, there was more frequent use of biologic or targeted synthetic DMARDs (33.9% vs 25.1%; $P = 0.03$) and prednisone (33.9% vs 17.7%; $P < 0.001$). Among those with RA-ILD, evidence of a UIP pattern was present in 62.8% of patients, and the mean FVC predicted was 77.6%. The *MUC5B* promoter variant occurred more frequently among those with RA-ILD (36.8% vs 16.8%).

PCA. A total of 15 PCs were extracted based on an eigenvalue >1 according to the Kaiser–Guttman rule.[27] These PCs explained 66% of the cumulative variance (Supplementary Table 1), with the first two PCs accounting for the largest amount of variance (PC1 15.5% and PC2 11.9%, respectively). The peripheral biomarkers that loaded onto each respective PC, a descriptive characterization of each PC, and the PC score association with RA-ILD are provided in Table 3. The first two PCs predominantly comprised proinflammatory cytokines and chemokines and were not associated with RA-ILD. Eight PCs were individually associated with RA-ILD. PC3 (adjusted odds ratio [aOR] 1.16, 95% confidence interval [CI] 1.05–1.29) was composed of cytokines involved in the innate immune response (IL-3, IFNα-2a, IL-15, and IL-5). PC4 (aOR 1.28, 95% CI 1.11–1.47), PC7 (aOR 1.18, 95% CI 1.01–1.37), and PC15 (aOR 1.26, 95% CI 1.07–1.49) were composed of distinct anti-MAA antibodies to albumin, fibrinogen, collagen, and/or vimentin. PC6 (aOR 1.17, 95% CI 1.02–1.34) was composed of anti-CCP antibodies and RF. PC8 (aOR 1.43, 95% CI 1.23–1.66) was composed of adipokines and cytokines (FL, leptin, fractalkine, and FGF-21). PC10 (aOR 1.21, 95% CI 1.03–1.41) was composed of alarmins (IL-17e–IL-25 and TSLP). PC12 (aOR 1.20, 95% CI 1.03–1.40) was composed of biomarkers related to tissue remodeling and neutrophil chemotaxis (MMP9 and IL-8). Other PCs were not significantly associated with RA-ILD.

Predictive models for RA-ILD. A logistic regression model including only clinical risk factors was able to discriminate RA-ILD with an AUC of 0.630 (95% CI 0.582–0.679). A logistic regression model composed of all 15 PCs was able to discriminate RA-ILD with an AUC of 0.713 (95% CI 0.662–0.764). When all PC scores were included with the clinical risk factors, the predictive ability of the model was significantly increased over the clinical model (AUC 0.739, 95% CI 0.693–0.786; Figure 1A). The addition of the MUC5B promoter variant to the clinical model with PC further improved performance (AUC 0.751, 95% CI 0.707–0.802; Figure 1A). This was also improved over the clinical model with *MUC5B* (AUC 0.681, 95% CI 0.637–0.726; Table 4). All model comparisons were significant via the nested likelihood ratio test ($P < 0.001$).

Sensitivity analyses with BMI, DAS28, and DMARD use included as additional clinical variables yielded slightly improved performance in clinical model (0.706, 95% CI 0.659–0.752), clinical model with PC (AUC 0.767, 95% CI 0.723–0.811), and clinical

Table 2. Characteristics of participants by RA-ILD status*

Variable	RA no-ILD (N = 1,880), n (%)	RA-ILD (N = 121), n (%)	*P* value
Age at enrollment, mean (±SD), y	63.5 (±11.2)	67.4 (±9.2)	<0.001
Male	1,660 (88.3)	114 (94.2)	0.047
Race and ethnicity			0.36
Non-Hispanic White	1,466 (78.0)	90 (74.4)	
Other	414 (22.0)	31 (25.6)	
Smoking history			0.07
Never smoker	386 (20.5)	15 (12.4)	
Ever smoker	1,456 (77.4)	102 (84.3)	
Unknown	38 (2.0)	4 (3.3)	
Body mass index, mean (±SD)	28.9 (±5.6)	28.5 (±6.1)	0.5
RA duration, mean (±SD), y	11.3 (±11.2)	12.9 (±12.7)	0.15
Anti-CCP seropositivity	1,441 (76.6)	101 (83.5)	0.08
RF seropositivity	1,435 (76.3)	105 (86.8)	0.01
DAS28-CRP, mean (±SD)	3.6 (±1.4)	3.7 (±1.3)	0.21
MD-HAQ, mean (±SD)	0.9 (±0.6)	1.0 (±0.5)	0.11
Prednisone use	332 (17.7)	41 (33.9)	<0.001
cDMARD use	1,379 (73.4)	94 (77.7)	0.29
b/tsDMARD use	471 (25.1)	41 (33.9)	0.03
MUC5B promoter variant			<0.001
0	1,471 (83.2)	74 (63.2)	
1	280 (15.8)	42 (35.9)	
2	18 (1.0)	1 (0.9)	
FVC predicted, mean (±SD), %	–	77.6 (±15.9)	–
ILD pattern			
UIP	–	76 (62.8)	–
Other ILD pattern (NSIP, etc)	–	36 (29.8)	–
Unknown/missing	–	9 (7.4)	–

* b/tsDMARD, biologic or targeted synthetic disease-modifying antirheumatic drug; CCP, cyclic citrullinated peptide; cDMARD, conventional disease-modifying antirheumatic drug; DAS28-CRP, Disease Activity Score With 28-Joint Count and C-reactive protein; FVC, forced vital capacity; MD-HAQ, Multidimensional Health Assessment Questionnaire; NSIP, nonspecific interstitial pneumonia; RA-ILD, rheumatoid arthritis–associated interstitial lung disease; RA no-ILD, rheumatoid arthritis without interstitial lung disease; RF, rheumatoid factor; UIP, usual interstitial pneumonia.

model with PC and *MUC5B* (AUC 0.776, 95% CI 0.733–0.823; Figure 1B). Again, all model comparisons were significant via the nested likelihood ratio test ($P < 0.001$). Sensitivity analysis with PCs selected based on individual associations (PC3, PC4, PC6, PC7, PC8, PC10, PC12, and PC15) did not significantly impact the results (Table 4). Internal validation was performed on all models using bootstrapping. Model performance was attenuated with optimism correction, but the overall trend in performance among the various models was preserved, as shown in Table 4.

Subgroup analyses. Subgroup analyses were performed to evaluate the association of each individual PC with RA-ILD based on ILD subtype and FVC impairment. Most PC associations with RA-ILD were consistent across subgroups (Supplementary Table 2). The exception was PC10 (IL-17e–IL-25 and TSLP), which was positively associated with UIP and negatively associated with non-UIP ILD. Additionally, PC13 (IgA/IgM antibodies to MAA-adducted collagen and adiponectin) was more closely associated with non-UIP ILD, and PC7 (IgA/IgG/IgM antibodies to MAA-adducted fibrinogen, collagen, and vimentin) was more closely associated with severe ILD (FVC < 80%), though 95% CIs for estimates were overlapping. Although a clinical model performed similarly in the subgroup with UIP (AUC 0.635 vs 0.630), there was a substantial improvement in performance of the PC model (AUC 0.768 vs 0.713) and combined models (clinical model with PC and *MUC5B* AUC 0.834 vs 0.755; Supplementary Table 3).

DISCUSSION

In this study, we identified unique peripheral biomarker signatures occurring in a large, diverse RA population and evaluated their association with the presence of RA-ILD. PCs that were positively associated with RA-ILD were characterized by biomarkers that encompassed inflammatory cytokines (IL-3, IL-5, IL-8, IL-15, INFα-2a), autoantibodies (anti-MAA, anti-CCP, RF), adipokines (FL, leptin, fractalkine, FGF-21), alarmins (IL-17e–IL-25, TSLP), and MMPs (MMP9). The variety of immune pathways and inflammatory processes associated with RA-ILD highlight the diversity of pathways that likely contribute to RA-ILD risk and pathogenesis or that are impacted by the presence of RA-ILD. Furthermore, combining these biomarker signatures with clinical risk and genetic factors improved RA-ILD risk stratification, illustrating the potential role of peripheral blood biomarkers in facilitating RA-ILD identification.

Table 3. Extracted peripheral biomarker principal components and their association with RA-ILD*

Principal component	Loading	Theme	aOR[a] for RA-ILD (95% CI)
1		Macrophage-related chemokines	0.96 (0.91–1.02)
MCP4	0.319		
MDC	0.311		
IL-16	0.310		
MCP1	0.306		
IL-27	0.304		
IL-7	0.302		
2		Proinflammatory cytokines	1.04 (0.98–1.10)
IL-1β	0.352		
IL-4	0.344		
IL-1α	0.308		
IL-6	0.304		
3		Myeloid differentiation; innate immune response; allergic response	1.16 (1.05–1.29)
IL-3	0.478		
IFNα-2a	0.407		
IL-15	0.375		
IL-5	0.309		
4		Autoantibodies; oxidative stress	1.28 (1.11–1.47)
IgA-MAA-Alb	0.565		
IgG-MAA-Alb	0.533		
IgM-MAA-Alb	0.560		
5		Adaptive immune response	0.97 (0.85–1.10)
IL-12–IL-23-p40	0.649		
IL-17A	0.532		
6		Autoantibodies	1.17 (1.02–1.34)
IgG-CCP	0.600		
RF	0.546		
7		Autoantibodies; oxidative stress	1.18 (1.01–1.37)
IgM-MAA-Fib	0.612		
IgM-MAA-Col	0.527		
IgA-MAA-Fib	0.314		
IgG-MAA-Vim	0.302		
8		Adipocytokines; metabolism regulation	1.43 (1.23–1.66)
FL	0.482		
Leptin	0.476		
Fractalkine	0.460		
FGF-21	0.391		
9		Autoantibodies; oxidative stress	1.10 (0.95–1.27)
IgG-MAA-Fib	0.588		
IgG-MAA-Col	0.468		
IgG-MAA-Vim	0.448		
10		Alarmins	1.21 (1.03–1.41)
IL-17e–IL-25	0.665		
TSLP	0.427		
11		Tissue remodeling	1.01 (0.88–1.17)
MMP3	0.652		
MMP7	0.521		
12		Tissue remodeling; neutrophil chemotaxis	1.20 (1.03–1.40)
MMP9	0.580		
IL-8	0.318		
13		Autoantibodies; oxidative stress; adipokines	1.10 (0.95–1.27)
IgA-MAA-Col	0.576		
Adiponectin	0.563		
IgG-MAA-Col	0.340		
14		Innate immune response	1.02 (0.87–1.18)
IL-33	0.641		
15		–	1.26 (1.07–1.49)
IgA-MAA-Vim	−0.439		

* Biomarkers with absolute loading value >0.3 are shown. Alb, albumin; aOR, adjusted odds ratio; CCP, cyclic citrullinated peptide; CI, confidence interval; Col, collagen; FGF, fibroblast growth factor; Fib, fibrinogen; FL, Fms-like tyrosine kinase 3 ligand; IFN, interferon; IL, interleukin; MAA, malondialdehyde-acetaldehyde adduct; MCP, monocyte chemoattractant protein; MDC, macrophage-derived chemokine; MMP, matrix metalloproteinase; RA-ILD, RA-ILD, rheumatoid arthritis–associated interstitial lung disease; RF, rheumatoid factor; TSLP, thymic stromal lymphopoietin; Vim, vimentin.
[a] Odds ratio values were adjusted for age, sex, race and ethnicity, and smoking history.

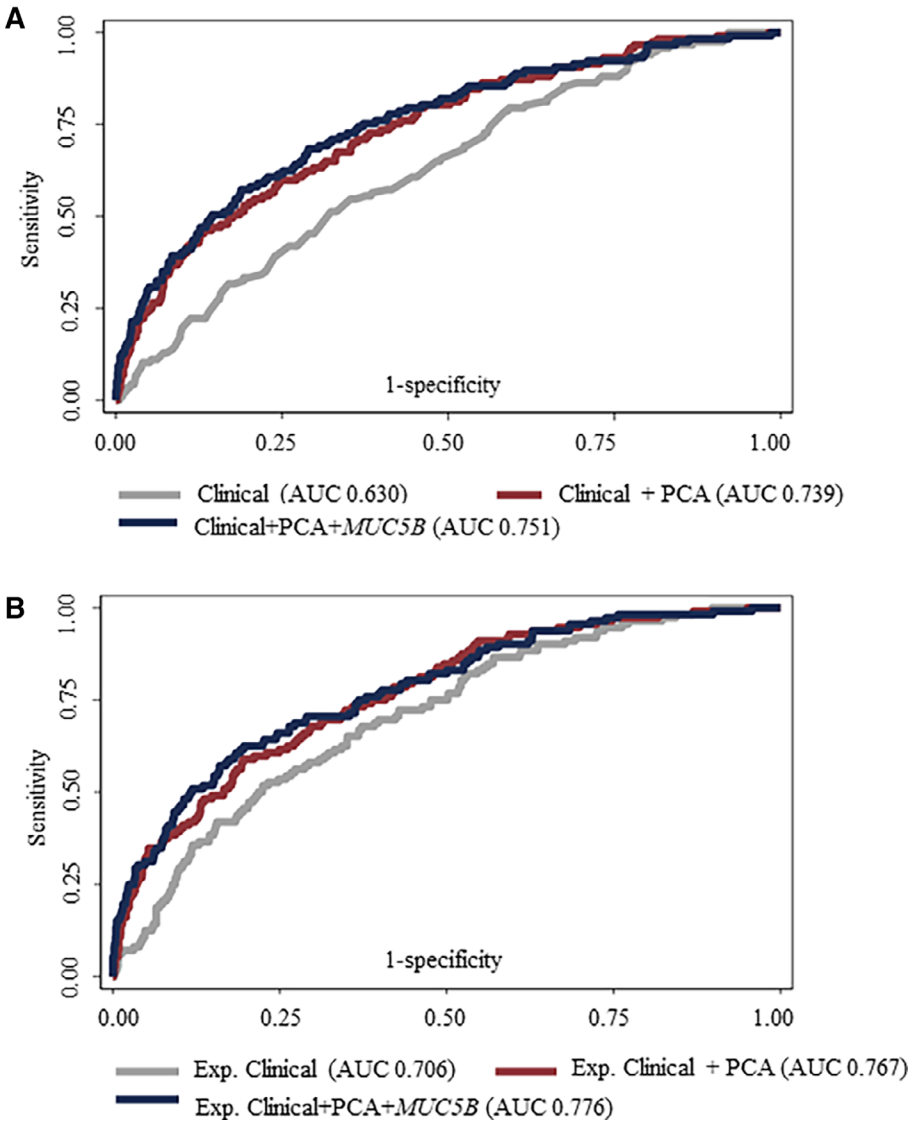

Figure 1. Receiver operating characteristic curves for logistic regression models for rheumatoid arthritis–associated interstitial lung disease including clinical risk factors, principal components, and *MUC5B* promoter variant. Receiver operating characteristic curves, along with their respective area under the receiver operating characteristic curve (AUC), are shown in each panel for clinical model alone, clinical model with principal components analysis (clinical + PCA), and clinical model with PCA and *MUC5B* promoter variant (clinical+PCA+*MUC5B*). (A) Clinical risk factors include age, sex, race and ethnicity, and smoking history. (B) Expanded (Exp.) clinical risk factors include the above plus body mass index, Disease Activity Score With 28-Joint Count, prednisone use, conventional disease-modifying antirheumatic drug (DMARD) use, and biologic or targeted synthetic DMARD use. All model comparisons were significantly different by nested likelihood ratio test ($P < 0.001$).

The current lack of RA-ILD screening strategies is due in part to a paucity of informative clinical risk factors for RA-ILD. This was evident in our study, in which models with only clinical risk factors yielded modest AUC values. Adding the biomarker PCs significantly improved the predictive performance of a model for the identification of RA-ILD over clinical risk factors alone. This improvement suggests that using peripheral blood biomarkers could help overcome the limited predictive value of clinical risk factors for RA-ILD. Although the performance of a model with clinical risk factors and peripheral biomarker signatures was statistically improved with the further addition of the *MUC5B* promoter variant, the absolute improvement in AUC was relatively modest. However, this degree of improvement was similar to the effect observed by adding the *MUC5B* promoter variant to clinical prediction models of RA-ILD,[8] suggesting that this genetic variant identifies features of susceptibility not captured through peripheral biomarkers. Additional research investigating the utility of integrating more comprehensive genetic and proteomic measurements into such models is warranted.

Biomarker themes that emerged through the PCA included myeloid differentiation, the innate immune and allergic responses, autoantibodies (specifically anti-MAA, anti-CCP, and RF),

Table 4. Performance of regression models for RA-ILD before and after internal validation*

Model	AUC (95% CI)	Optimism-adjusted AUC (95% CI)[a]
Primary analyses		
Clinical factors	0.630 (0.583–0.679)	0.609 (0.558–0.669)
PCs	0.713 (0.662–0.764)	0.677 (0.628–0.723)
Clinical with *MUC5B*	0.681 (0.637–0.726)	0.663 (0.625–0.702)
Clinical with PCs	0.739 (0.693–0.786)	0.693 (0.647–0.740)
Clinical with PCs and *MUC5B*	0.755 (0.707–0.802)	0.713 (0.667–0.758)
Sensitivity analyses		
Expanded clinical[b]	0.706 (0.659–0.752)	0.671 (0.626–0.716)
Expanded clinical with PCs	0.767 (0.723–0.811)	0.712 (0.665–0.757)
Expanded clinical with PCs and *MUC5B*	0.776 (0.731–0.821)	0.724 (0.676–0.782)
Selected PCs[c]	0.702 (0.651–0.753)	0.677 (0.621–0.732)
Clinical with selected PCs	0.731 (0.684–0.778)	0.695 (0.659–0.748)
Clinical with selected PCs and *MUC5B*	0.749 (0.701–0.797)	0.716 (0.675–0.770)

* AUC, area under the receiver operating characteristic curve; CI, confidence interval; DMARD, disease-modifying antirheumatic drug; *MUC5B*, *MUC5B* rs35705950 promoter variant; PC, principal component; RA-ILD, rheumatoid arthritis–associated interstitial lung disease.
[a] Internal validation by was achieved through bootstrapping.
[b] Expanded refers to the addition of body mass index, Disease Activity Score With 28-Joint Count, prednisone use, conventional DMARD use, and biologic or targeted synthetic DMARD use to the clinical model.
[c] Selected refers to the inclusion of only the PCs that were individually associated with RA-ILD.

adipokines/metabolism, alarmins, tissue remodeling, and neutrophil chemotaxis. These findings may hint toward the diverse pathophysiologic pathways involved in RA-ILD, many of which are supported by prior literature. It has been demonstrated that serum anti–citrullinated protein antibodies (ACPAs) correlate with local ACPA production in the lung mucosa and mucosal inflammation, leading to the "mucosal origins" hypothesis of RA.[29,30] Inflammatory cytokines such as TNF, IL-1, IL-4, IL-5, IL-13, and several chemokines are thought to likely play a role in the transition from epithelial injury to the development of lung fibrosis.[29] MAA expression in lung tissue has been shown to be higher in RA-ILD than in other forms of ILD, emphysema, or controls, and autoreactive B cells to MAA have been found in the lung and joint tissue of patients with RA.[5,31] Adipokines, specifically adiponectin and leptin, have been shown to play a regulatory role in the inflammatory response, which has been suspected to have a potential role in ILD pathogenesis.[32] MMPs (including MMP9) have been shown to be overexpressed by alveolar epithelial cells within fibroblastic foci of the lungs in idiopathic pulmonary fibrosis and drive lung fibrosis in a highly similar disease process to UIP RA-ILD.[29,33] Thus, our findings are in line with these prior studies and emphasize the need for further translational and mechanistic efforts to determine how these processes are involved in RA-ILD pathogenesis.

In the subgroup analyses, we found that the associations of PCs with UIP RA-ILD were generally consistent with those for all RA-ILD, likely reflecting the preponderance of UIP in RA-ILD. Interestingly, PC10 (IL-17e–IL-25 and TSLP) was positively associated with a UIP pattern and negatively associated with other ILD patterns. This has not been previously demonstrated, and this hypothesis-generating finding warrants further investigation into factors that differentiate specific ILD patterns and severity. Notably, because of the limited number of patients with ILD in subgroup analyses, we were unable to detect modest differences in PC associations with subgroups with RA-ILD. There was improvement in clinical/biomarker/MUC5B model performance for the outcome of UIP over all RA-ILD. This was limited by power but supports the idea that risk stratification efforts in RA-ILD may need to be pattern specific.

Strengths of this study include a large, diverse, multicenter US cohort of well-phenotyped patients with RA with comprehensive RA-ILD identification and peripheral biomarker measurements.[19] Limitations include a lack of chest CT imaging for the entirety of the cohort given that RA-ILD testing was performed as part of routine clinical care. This may result in misclassification of subclinical ILD as non-ILD. However, our results illustrate the utility of peripheral biomarker signatures for RA-ILD risk stratification in a real-world practice setting. ILD pattern (UIP, nonspecific interstitial pneumonia, etc) was unavailable for some patients (n = 9) and was based on clinical determinations rather than a separate research evaluation. Although this cohort is predominantly male, it is reflective of the national VA population, and RA-ILD preferentially affects males. Although the lack of external validation is a limitation, internal validation was performed via bootstrapping, and models underwent optimism correction, which did not impact the interpretation of findings. Given the cross-sectional nature of this study, there is potential for reverse causality between RA-ILD and peripheral biomarkers, although with the goal of identifying prevalent disease, this should not impact the interpretation of our results. We included a broad range of peripheral blood biomarkers implicated in RA and RA-ILD pathogenesis but were unable to include all previously reported biomarkers associated with ILD. This remains an evolving area of research, and future studies will need to continually reassess other novel biomarkers that would benefit from inclusion.[34–36] Last, the PCA method used allowed for the

identification of distinct biomarker profiles within a broad range of potential peripheral biomarkers and their association with RA-ILD. However, alternative approaches would be needed to identify individual biomarker cutoffs or evaluate the clinical application of predictive models, including the use of other machine learning techniques to guide efficient biomarker selection for clinical prediction models.

In conclusion, PCA identified distinct peripheral biomarker signatures that implicate a wide range of biologic pathways associated with RA-ILD. These peripheral biomarker signatures significantly improved ILD discrimination in a large cohort with RA beyond clinical risk factors alone. The addition of the *MUC5B* promoter variant resulted in a modest improvement over clinical and peripheral biomarker models. Further investigation is needed regarding the effective use of peripheral biomarkers in the identification of RA-ILD as well as applications toward ILD incidence and progression. This includes their additive contribution to other clinical and genetic risk factors and their relationship to the various pathways involved in RA-ILD pathogenesis that could impact therapeutic targeting.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr England confirms that all authors have provided the final approval of the version to be published, and takes responsibility for the affirmations regarding article submission (eg, not under consideration by another journal), the integrity of the data presented, and the statements regarding compliance with institutional review board/Declaration of Helsinki requirements.

REFERENCES

1. Marigliano B, Soriano A, Margiotta D, et al. Lung involvement in connective tissue diseases: a comprehensive review and a focus on rheumatoid arthritis. Autoimmun Rev 2013;12(11):1076–1084.
2. Spagnolo P, Lee JS, Sverzellati N, et al. The lung in rheumatoid arthritis: focus on interstitial lung disease. Arthritis Rheumatol 2018;70(10): 1544–1554.
3. Chartrand S, Lee JS, Swigris JJ, et al. Clinical characteristics and natural history of autoimmune forms of interstitial lung disease: a single-center experience. Lung 2019;197(6):709–713.
4. Kelly CA, Saravanan V, Nisar M, 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):1676–1682.
5. England BR, Duryee MJ, Roul P, et al. Malondialdehyde-acetaldehyde adducts and antibody responses in rheumatoid arthritis-associated interstitial lung disease. Arthritis Rheumatol 2019;71(9):1483–1493.
6. Juge PA, Lee JS, Ebstein E, et al. MUC5B promoter variant and rheumatoid arthritis with interstitial lung disease. N Engl J Med 2018; 379(23):2209–2219.
7. Wheeler AM, Baker JF, Poole JA, et al. Genetic, social, and environmental risk factors in rheumatoid arthritis-associated interstitial lung disease. Semin Arthritis Rheum 2022;57:152098.
8. Wheeler AM, Baker JF, Riley T, et al. Development and internal validation of a clinical and genetic risk score for rheumatoid arthritis-associated interstitial lung disease. Rheumatology (Oxford) 2025;64:268–275.
9. Juge PA, Granger B, Debray MP, et al. A risk score to detect subclinical rheumatoid arthritis-associated interstitial lung disease. Arthritis Rheumatol 2022;74(11):1755–1765.
10. Matson SM, Deane KD, Peljto AL, et al. Prospective identification of subclinical interstitial lung disease in a rheumatoid arthritis cohort is associated with the MUC5B promoter variant. Am J Respir Crit Care Med 2022;205(4):473–476.
11. Van Kalsbeek D, Brooks R, Shaver D, et al. Peripheral blood biomarkers for rheumatoid arthritis-associated interstitial lung disease: a systematic review. ACR Open Rheumatol 2023;5:201–226.
12. Centola M, Cavet G, Shen Y, et al. Development of a multi-biomarker disease activity test for rheumatoid arthritis. PLoS One 2013;8(4): e60635.
13. Abdelhafiz D, Baker T, Glascow DA, et al. Biomarkers for the diagnosis and treatment of rheumatoid arthritis - a systematic review. Postgrad Med 2023;135(3):214–223.
14. Wu X, Jeong Y, Poli de Frías S, et al. Serum proteomic profiling of rheumatoid arthritis-interstitial lung disease with a comparison to idiopathic pulmonary fibrosis. Thorax 2022;77(10):1041–1044.
15. Kass DJ, Nouraie M, Glassberg MK, et al. Comparative profiling of serum protein biomarkers in rheumatoid arthritis-associated interstitial lung disease and idiopathic pulmonary fibrosis. Arthritis Rheumatol 2020;72(3):409–419.
16. Mikuls TR, Reimold A, Kerr GS, et al. Insights and implications of the VA Rheumatoid Arthritis Registry. Fed Pract 2015;32(5):24–29.
17. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988;31(3):315–324.
18. Luedders BA, Wheeler AM, Ascherman DP, et al. Plasma matrix metalloproteinase concentrations and risk of interstitial lung disease in a prospective rheumatoid arthritis cohort. Arthritis Rheumatol 2024;76:1013–1022.
19. England BR, Roul P, Mahajan TD, et al. Performance of administrative algorithms to identify interstitial lung disease in rheumatoid arthritis. Arthritis Care Res (Hoboken) 2020;72(10):1392–1403.
20. Luedders BA, Cope BJ, Hershberger D, et al. Enhancing the identification of rheumatoid arthritis-associated interstitial lung disease through text mining of chest computerized tomography reports. Semin Arthritis Rheum 2023;60:152204.
21. Gould PW, Zemel BS, Taratuta EG, et al. Circulating fibroblast growth factor-21 levels in rheumatoid arthritis: associations with disease characteristics, body composition, and physical functioning. J Rheumatol 2021;48(4):504–512.
22. Baker JF, England BR, George MD, et al. Elevations in adipocytokines and mortality in rheumatoid arthritis. Rheumatology (Oxford) 2022; 61(12):4924–4934.
23. Poole JA, England BR, Sayles H, et al. Serum alarmins and the risk of incident interstitial lung disease in rheumatoid arthritis. Rheumatology (Oxford) 2023;63:1998–2005.
24. Sainani KL. Introduction to principal components analysis. PM R 2014;6(3):275–278.
25. Giuliani A. The application of principal component analysis to drug discovery and biomedical data. Drug Discov Today 2017;22(7):1069–1076.
26. Field A. Discovering Statistics Using IBM SPSS Statistics. SAGE; 2013.
27. Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas 1960;20(1):141–151.
28. Fernandez-Felix BM, García-Esquinas E, Muriel A, et al. Bootstrap internal validation command for predictive logistic regression models. Stata J 2021;21(2):498–509.

29. Kadura S, Raghu G. Rheumatoid arthritis-interstitial lung disease: manifestations and current concepts in pathogenesis and management. Eur Respir Rev 2021;30(160):210011.
30. Holers VM, Demoruelle MK, Kuhn KA, et al. Rheumatoid arthritis and the mucosal origins hypothesis: protection turns to destruction. Nat Rev Rheumatol 2018;14(9):542–557.
31. Sahlström P, Joshua V, Valkovskaia V, et al. Autoreactive B cells against malondialdehyde-induced protein cross-links are present in the joint, lung, and bone marrow of rheumatoid arthritis patients. J Biol Chem 2023;299(11):105320.
32. Haley S, Shah D, Romero F, et al. Scleroderma-related lung disease: are adipokines involved pathogenically? Curr Rheumatol Rep 2013; 15(12):381.
33. Spagnolo P, Kropski JA, Jones MG, et al. Idiopathic pulmonary fibrosis: disease mechanisms and drug development. Pharmacol Ther 2021;222:107798.
34. Matsuo T, Hashimoto M, Ito I, et al. Interleukin-18 is associated with the presence of interstitial lung disease in rheumatoid arthritis: a cross-sectional study. Scand J Rheumatol 2019;48(2):87–94.
35. 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(12):1403–1412.
36. Hussein MS, El-Barbary AM, Nada DW, et al. Identification of serum interleukin-13 and interleukin-13 receptor subunit expressions: rheumatoid arthritis-associated interstitial lung disease. Int J Rheum Dis 2021;24(4):591–598.