The Open Rheumatology Journal
DOI: 10.2174/0118743129499823260727114010, 2026, 20, e18743129499823
ISSN: 1874-3129

RESEARCH ARTICLE **OPEN ACCESS**

Cluster Analysis Identifies Rheumatoid Arthritis Phenotypes with Distinct Features of Pulmonary Involvement

Minh Tri Duong[1,2,*], Bui Bao Hoang[1], Minh Tri Bui[3], Dinh Kieu Diem Truong[3] and Hoang Kim Tu Trinh[3]

[1]*Internal Medicine, Hue University of Medicine and Pharmacy, Hue, Vietnam*
[2]*Department of Rheumatology, Nhan dan Gia Dinh Hospital, Ho Chi Minh City, Vietnam*
[3]*Center for Molecular Biomedicine, University of Medicine and Pharmacy at Ho Chi Minh City, Ho Chi Minh City, Vietnam*

Abstract:

Introduction: Rheumatoid Arthritis (RA) is a heterogeneous disease with variable articular and extra-articular manifestations. Interstitial Lung Disease (ILD) is a serious complication contributing to morbidity and mortality. The study aims to identify distinct clinical phenotypes using a clustering approach and to determine the prevalence of RA-associated ILD (RA-ILD) in a Vietnamese cohort.

Methods: This prospective study included 306 patients fulfilling the 2010 ACR/EULAR RA criteria. ILD was diagnosed by high-resolution computed tomography according to ATS/ERS/JRS/ALAT standards. Demographic, clinical, serologic, and radiologic data were collected. Partitioning around medoids clustering with Gower's distance was applied to identify phenotypic subgroups.

Results: Among 306 patients (median age 59 years; 87.6% female), 101 (33.0%) had ILD. Cluster analysis identified four phenotypes. Clusters 1 and 2 were articular-dominant with minimal pulmonary involvement, whereas Clusters 3 and 4 were pulmonary-dominant with higher ILD prevalence and greater fibrosis severity, demonstrating a reciprocal pattern between joint and lung manifestations. Cluster 2 emerged as a unique clinical phenotype, marked by later onset of disease, minimal pulmonary lesions, and male predominance. Patients with RA-ILD had longer disease duration, higher CRP levels, and higher RF titers compared with those without ILD. Respiratory symptoms were significantly more frequent in the ILD group, whereas swollen joint counts and DAS28-CRP scores were lower. The usual interstitial pneumonia pattern predominated (69.3%).

Discussion: The study comprehensively characterized RA phenotypes in a Vietnamese cohort of 306 patients and identified divergent articular and pulmonary patterns, including a prevalence of RA-associated interstitial lung disease (RA-ILD) of 33%, consistent with previously reported estimates (10–50%). Cluster analysis revealed reciprocal phenotypes: two clusters were characterized by more severe articular involvement with limited pulmonary fibrosis, whereas two others showed greater pulmonary manifestations. Notably, patients with RA-ILD exhibited fewer swollen joints and lower pain scores than those without ILD, suggesting that lung involvement does not necessarily parallel joint disease severity at a single time point. Instead, systemic inflammatory burden may be more closely associated with pulmonary pathology, as clusters with greater lung involvement demonstrated higher ESR levels. RA-ILD predominantly occurred in older individuals, supporting the role of aging-related mechanisms in fibrotic susceptibility. A distinct cluster with male predominance, later disease onset, and minimal pulmonary lesions was also observed.

Conclusion: Vietnamese RA exhibits substantial clinical heterogeneity with divergent articular- and pulmonary-dominant phenotypes.

Keywords: Interstitial Lung Disease (ILD), MUC5B mutation, Rheumatoid arthritis (RA), Phenotype, Vietnam, Articular lesions.

*Address correspondence to this author at the Internal Medicine, Hue University of Medicine and Pharmacy, Hue, Vietnam and Department of Rheumatology, Nhan dan Gia Dinh Hospital, Ho Chi Minh City, Vietnam;
Tel: +84-986 972 707; E-mail: bsmtri2000@gmail.com

Cite as: Duong M, Hoang B, Bui M, Truong D, Trinh H. Cluster Analysis Identifies Rheumatoid Arthritis Phenotypes with Distinct Features of Pulmonary Involvement. Open Rheumatol J, 2026; 20: e18743129499823.
http://dx.doi.org/10.2174/0118743129499823260727114010

Received: March 09, 2026
Revised: June 04, 2026
Accepted: June 11, 2026
Published: July 31, 2026

1. INTRODUCTION

The chronic inflammatory disease referred to as Rheumatoid Arthritis (RA) is characterized by inflammation in the synovial fluid that damages the joints, ultimately decreasing patients' quality of life [1]. While minor joints are usually affected symmetrically, there may also be organ involvement, including the lungs, heart, and eyes [2]. The estimated frequency of RA worldwide is 0.5–1%, with higher rates in women and older populations [3]. It is unclear exactly what causes RA, although environmental stresses including infections and smoking, and genetic predisposition have been implicated [4]. Serologic indicators of disease severity, such as Rheumatoid Factor (RF) and Anti-Cyclic Citrullinated Peptide (ACPA) antibodies, are frequently utilized for the diagnosis of RA [5].

Interstitial Lung Disease (ILD) is a severe, common extra-articular symptom of RA, accounting for a significant proportion of the mortality and morbidity in patients. The prevalence of ILD among patients with RA is estimated to be 10–19%, although substantial racial and geographic variations have been reported [6]. Nonspecific interstitial pneumonia and Usual Interstitial Pneumonia (UIP) represent the two predominant histopathological characteristics of RA–associated Interstitial Lung Disease (RA-ILD), with the UIP subtype generally associated with a poor prognosis [7]. Risk factors for Rheumatoid Arthritis–Associated Interstitial Lung Disease (RA-ILD) include older age, male sex, high RA disease activity, smoking, and seropositivity for RF and ACPA antibodies [8]. Despite improvements in our understanding of the pathophysiology of RA-ILD, the mechanisms linking joint and lung involvement are unknown. Genetic variants may be associated with an increased likelihood of RA-ILD, for example, *MUC5B* rs35705950 [9]. Individuals with RA who carry the MUC5B risk allele have a 3.1-fold higher risk of developing ILD and a 6.1-fold greater likelihood of exhibiting radiographic pulmonary abnormalities on High-Resolution Computed Tomography (HRCT) compared with non-carriers [10]. However, published reports about the *MUC5B* rs 35705950 in Vietnamese patients with RA-ILD are lacking.

Another major obstacle for clinicians is the diverse clinical presentation of RA. Traditional classification systems emphasize clinical criteria and serologic markers. Individuals who test positive for the autoantibodies, including Rheumatoid Factor (RF) and Anti–Cyclic Citrullinated Peptide (ACPA), are classified as having "seropositive" RA, whereas those who fulfill clinical criteria for RA but lack these antibodies are categorized as having "seronegative" RA [11]. Seronegative RA is often perceived as having a milder disease course and being more manageable, but they had the highest disease burden of all phenotypes [12]. Recently, many phenotyping strategies have been proposed for RA, which facilitate understanding the heterogeneity of RA.

A population-based WHO-ILAR COPCORD study reported a prevalence of rheumatoid arthritis of 0.28% among adults in urban Vietnam [13]. There are studies describing the characteristics and prevalence of RA and RA-ILD (~31,6%) [14]. However, these investigations have primarily emphasized general clinical and laboratory characteristics rather than in-depth phenotypic stratification. Therefore, the present study aimed to determine phenotypes of RA using demographic, clinical, radiologic, and genetic factors associated with pulmonary involvement. Subsequently, we evaluated the prevalence of ILD and the associated characteristics of RA-ILD.

2. MATERIALS and METHODS

2.1. Study Enrollment and Procedures

This prospective study included 306 patients with RA who were admitted to the Department of Rheumatology at Gia Dinh People's Hospital between January 2023 and April 2024.

The inclusion criteria were: (1) age ≥18 years; (2) fulfillment of the 2010 American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria for RA; and (3) provision of written informed consent. In addition, the exclusion criteria included: (1) concomitant autoimmune diseases other than RA; (2) active pulmonary infection confirmed by a positive sputum culture or ongoing treatment for pneumonia; (3) incomplete clinical, laboratory, or imaging data; (4) contraindications to HRCT examination, including pregnancy or refusal to undergo HRCT; and (5) withdrawal from the study before completion.

Patients were categorized into two distinct groups: those with ILD (n = 101) and those without ILD (RA-no-ILD, n = 205). RA was diagnosed based on the history, clinical symptoms, serological markers, and acute phase

reactants, RF, and erythrocyte sedimentation rate (ESR based on the 2010 ACR/EULAR RA classification [11]. Written informed consent was obtained from all patients before participation. HRCT was conducted initially and, in certain cases, again a month later if structural abnormalities were found on one side of the lungs. EDTA tubes were used to collect 2 mL of peripheral venous blood for DNA sequencing to detect the *MUC5B* mutation, as described below.

This study was conducted in accordance with the principles of the Declaration of Helsinki, and the study protocol was approved by the Ethics Committee Review Board of Hue University of Medicine and Pharmacy (Approval No. H2022/519, approved by Prof. Vo Tam, Chair of the Ethics Committee). Part of the data has been published previously; the publication is cited herein to ensure scientific transparency [15, 16].

2.2. Evaluation of ILD

The thoracic HRCT results were used to diagnose ILD in accordance with the 2018 ATS/ERS/JRS/ALAT diagnostic criteria [17]: the simultaneous existence of fine reticulation and ground-glass opacification, centrilobular lung nodules, honeycombing, traction bronchiectasis, and traction bronchiolectasis affecting > 5% of any lung zone. The criteria recommend a second HRCT, performed one month later, if structural lung abnormalities are documented only on one side. Other diagnostic criteria consisted of a negative sputum culture and excluding several possible causes of ILD (medication toxicity, connective tissue illness, and exposures to environmental factors at work and at home) based on a detailed history of medication use, environmental exposure, and serological test results. HRCT patterns were independently read and interpreted by radiologists and pulmonologists. The diagnosis of ILD was made by consensus. For unclear readings, the images were discussed with another pulmonologist before a final decision was reached.

2.3. Demographic Factors and Assessment of Disease Activity

The following data were obtained: age, gender, disease duration, stage of illness (based on the Steinbroker classification), disease activity according to Disease Activity Score-C-Reactive Protein (DAS28-CRP), and the ILD imaging patterns on HRCT.

2.4. Amplification and Sequencing of the MUC5B Promoter Variant rs35705950

Genomic DNA was extracted from peripheral blood using the GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific, USA). The MUC5B promoter variant rs35705950 was amplified by Polymerase Chain Reaction (PCR) using specific primers designed with CLC Main Workbench software. Purified PCR products were subjected to direct sequencing using the ABI BigDye Terminator v3.1 Cycle Sequencing Kit and analyzed on an ABI 3500 Genetic Analyzer (Applied Biosystems, USA). Sequence data were processed with CLC Main Workbench to identify the rs35705950 variant. Detailed primer sequences, reaction components, and cycling conditions are provided in the **Supplementary Materials**.

2.5. Cluster Analysis

The Partitioning Around Medoids (PAM) algorithm was applied to classify ILD patients into groups sharing similar clinical phenotypes based on 10 baseline variables. Cluster analysis aims to group individuals according to predefined characteristics so that members within each cluster are as similar as possible, while differences between clusters are maximized [18]. Because the analysis included both continuous and categorical variables, Gower's distance was used to standardize the data prior to clustering. This approach rescales variables to a common range from 0 to 1, allowing mixed data types to be analyzed simultaneously [19]. The optimal number of clusters was determined using the silhouette width, which quantifies how well each subject matches its assigned cluster compared with neighboring clusters [20]. All analyses were conducted using the cluster package in R version 4.5.0 (The R Foundation for Statistical Computing, Vienna, Austria).

2.6. Statistical Analysis

Statistical analyses were performed using SPSS version 20.0 (SPSS Inc., USA) and R version 4.5.0 (The R Foundation for Statistical Computing, Vienna, Austria). Data normality was assessed with the Shapiro–Wilk test. Categorical variables were summarized as n (%), normally distributed data as mean ± SD, and non-normally distributed data as median [Q1; Q3]. Continuous variables were compared using Student's t-test or Wilcoxon rank-sum test, and categorical variables were assessed with Pearson's chi-squared or Fisher's exact test. A P value of less than 0.05 is considered statistically significant.

3. RESULTS

3.1. Demographic, Seropositivity and Inflammatory Characteristics of Four Clusters

A total of 306 patients were recruited into the study. The median age was 59 years, with female predominance (87.6%). The onset age was 54 years, and the duration of illness was 4 years. The prevalence of family history of RA was 11.4%. Data are shown in Table **S1**.

These patients were stratified into four clusters: Cluster 1 (n = 179), Cluster 2 (n = 23), Cluster 3 (n = 37), and Cluster 4 (n = 67). In Table **1**, there were no statistical differences in terms of age between the four clusters, although Clusters 3 and 4 consisted predominantly of patients aged ≥50 years. Disease duration differed significantly ($p < 0.001$), being shortest in Cluster 2 and longest in Clusters 3 and 4. Especially, patients in Cluster 2 were male-predominant (100%), and had significant later onset of disease (~ 59.8 years old) than patients in Cluster 1 (53.00 years old) and Cluster 4 (52.60 years old) (P=0.044, P=0.027, respectively) (Fig. **S1**). Inflammatory markers (CRP, ESR), Rheumatoid Factor (RF), ACPA levels, and BMI did not significantly differ between clusters.

Table 1. Unbiased phenotyping of RA patients.

-	Cluster 1 (N=179)	Cluster 2 (N=23)	Cluster 3 (N=37)	Cluster 4 (N=67)	*P*-values
Age (years old)	58.00 [50.00 – 66.50]	60.00 [53.00 – 68.50]	62.00 [57.00 – 67.00]	60.00 [53.50 – 65.50]	0.108[δ]
Age (years old)	-	-	-	-	0.01
<50	39 (21.8%)	3 (13.0%)	0 (0.00%)	10 (14.9%)	-
>=50	140 (78.2%)	20 (87.0%)	37 (100%)	57 (85.1%)	-
Sex	-	-	-	-	< 0.001[F]
Female	179 (100%)	0	22 (59.46%)	67 (100%)	-
Male	0 (0%)	23 (100%)	15 (40.54%)	0 (0%)	-
Duration of illness (years)	3.00 [1.00;7.00]	0.60 [0.30;3.50]	5.00 [2.00;10.0]	5.00 [3.00;10.0]	<0.001[δ]
BMI	22.0 [20.6;23.4]	22.0 [21.5;23.9]	22.7 [20.2;25.0]	22.1 [20.4;24.3]	0.683[δ]
Onset age of RA (years old)	53.00 [45.00 – 62.25]	59.80 [52.20 – 65.90]	56.00 [51.00 – 59.60]	52.60 [46.50 – 59.25]	0.092[δ]
RF (U/mL)	36.9 [17.0;100]	39.2 [18.5;116]	48.2 [12.6;120]	38.9 [18.0;93.4]	0.819[δ]
ACPA (U/mL)_	79.4 [10.5;196]	73.0 [15.8;189]	35.4 [17.7;106]	72.2 [14.9;196]	0.620[δ]
ESR (mm)	34.0 [15.0;80.0]	18.0 [14.0;27.0]	30.0 [20.0;46.0]	35.0 [18.0;61.0]	0.350[δ]
CRP (mg/L)	12.1 [4.00;40.0]	20.0 [7.04;44.9]	14.6 [5.50;53.0]	12.3 [4.00;40.0]	0.387[δ]
	-	-	-	-	-
RA-ILD	-	-	-	-	<0.001
No	143 (79.9%)	22 (95.7%)	15 (40.5%)	25 (37.3%)	-
Yes	36 (20.1%)	1 (4.35%)	22 (59.5%)	42 (62.7%)	-
MUC5B rs35705950 variant	-	-	-	-	0.449[F]
GG	169 (94.4%)	23 (100%)	36 (97.3%)	61 (91.0%)	-
GT	10 (5.59%)	0 (0.00%)	1 (2.70%)	6 (8.96%)	-
ILD_lesions:	-	-	-	-	0.187[F]
<=10%	30 (85.7%)	1 (100%)	15 (68.2%)	38 (90.5%)	-
10.1-25	3 (8.57%)	0 (0.00%)	5 (22.7%)	4 (9.52%)	-
>25	2 (5.71%)	0 (0.00%)	2 (9.09%)	0 (0.00%)	-
Pulmonary fibrosis:	-	-	-	-	0.040[F]
<=10%	18 (100%)	0 (%)	9 (75.0%)	27 (96.4%)	-
10.1-25	0 (0.00%)	0 (%)	1 (8.33%)	1 (3.57%)	-
>25	0 (0.00%)	0 (%)	2 (16.7%)	0 (0.00%)	-
ILD pattern on HRCT	-	-	-	-	0.512[F]
UIP	27 (75.0%)	0 (0.00%)	14 (63.6%)	29 (69.0%)	-
NSIP	6 (16.7%)	1 (100%)	6 (27.3%)	11 (26.2%)	-
OP/CVP	3 (8.33%)	0 (0.00%)	2 (9.09%)	2 (4.76%)	-
FVC (%)	-	-	-	-	0.537[F]
<50	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (2.56%)	-
50-69	5 (19.2%)	0 (0.00%)	3 (14.3%)	4 (10.3%)	-
70-79	4 (15.4%)	0 (0.00%)	4 (19.0%)	2 (5.13%)	-
>80	17 (65.4%)	1 (100%)	14 (66.7%)	32 (82.1%)	-
DLCO (%):	-	-	-	-	0.088[F]
<40	3 (11.5%)	0 (0.00%)	1 (4.76%)	1 (2.56%)	-
40-59	3 (11.5%)	0 (0.00%)	9 (42.9%)	7 (17.9%)	-
60-79	13 (50.0%)	0 (0.00%)	4 (19.0%)	17 (43.6%)	-
>80	7 (26.9%)	1 (100%)	7 (33.3%)	14 (35.9%)	-
Morning stiffness	30.0 [15.0;60.0]	30.0 [20.0;60.0]	30.0 [14.0;60.0]	15.0 [10.0;30.0]	0.064[δ]
Number of painful joints:	-	-	-	-	0.172[F]
0	11 (6.15%)	2 (8.70%)	6 (16.2%)	7 (10.4%)	-
1-10	134 (74.9%)	14 (60.9%)	24 (64.9%)	52 (77.6%)	-
>10	34 (19.0%)	7 (30.4%)	7 (18.9%)	8 (11.9%)	-
Number of swollen joints:	-	-	-	-	0.048[F]
0	73 (40.8%)	7 (30.4%)	13 (35.1%)	33 (49.3%)	-
1-10	104 (58.1%)	16 (69.6%)	20 (54.1%)	32 (47.8%)	-
>10	2 (1.12%)	0 (0.00%)	4 (10.8%)	2 (2.99%)	-
mMRC scale	-	-	-	-	.

-	Cluster 1 (N=179)	Cluster 2 (N=23)	Cluster 3 (N=37)	Cluster 4 (N=67)	*P*-values
Stage 0	179 (100%)	23 (100%)	1 (2.70%)	1 (1.49%)	-
Stage 1	0 (0.00%)	0 (0.00%)	23 (62.2%)	59 (88.1%)	-
Stage 2	0 (0.00%)	0 (0.00%)	6 (16.2%)	6 (8.96%)	-
Stage 3	0 (0.00%)	0 (0.00%)	3 (8.11%)	1 (1.49%)	-
Stage 4	0 (0.00%)	0 (0.00%)	4 (10.8%)	0 (0.00%)	-

Note: [§]Numerical data were presented as median [Q1; Q3]. Differences between groups were calculated by the Kruskal-Wallis test.
Categorical data were presented as N (%). Differences between groups were calculated by Pearson's chi-square or [F]Fisher's exact test.

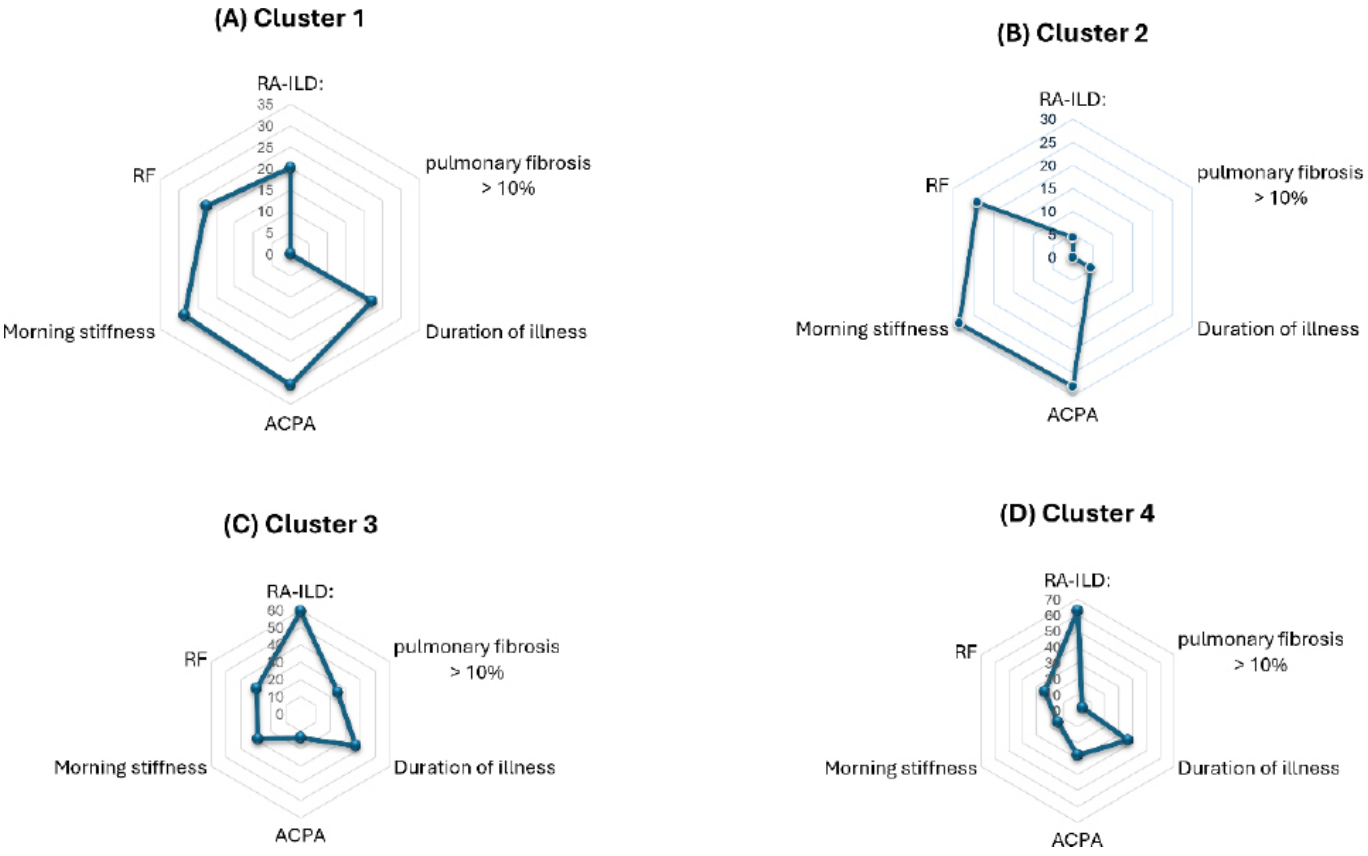

Fig. (1). Distinct characteristics of the four clusters of RA. A radar plot was used to illustrate the frequency of RA-ILD presence, pulmonary fibrosis > 10%, morning stiffness, autoantibody titers (RF and ACPA), and duration of illness. ACPA, anti-cyclic citrullinated peptide; ILD, interstitial lung disease; RA, rheumatoid arthritis; RA-ILD, RA patients with ILD; RF, rheumatoid factor.

3.2. Features of ILD and Joints Lesions

Distinct pulmonary and joint lesions were seen between four clusters (Table **1**, Fig. **1**). Firstly, the distribution of RA-ILD differed significantly among clusters ($P< 0.001$). RA-ILD prevalence was highest in Cluster 4 (62.7%) and Cluster 3 (59.5%), whereas Clusters 1 and 2 showed substantially lower proportions (20.1% and 4.35%, respectively).

The overall prevalence of the *MUC5B* rs35705950 variant was low, 5.6% (17/306) (Fig. **S2**) and did not differ significantly among clusters ($P= 0.449$). Among the 101 RA patients with ILD, 16 (15.8%) tested positive for the MUC5B variant, whereas among the 205 RA patients without ILD, only 1 (0.5%) had the variant (Fig. **S2**).

Regarding radiological characteristics, the extent of ILD lesions did not significantly differ among clusters ($P=0.187$). However, the degree of pulmonary fibrosis showed significant variation ($P=0.040$), with more extensive fibrosis observed in Cluster 3. The Usual Interstitial Pneumonia (UIP) pattern predominated across ILD cases, although no significant difference in HRCT pattern distribution was found among clusters ($P=0.512$). Pulmonary function parameters, including Forced Vital Capacity (FVC) and diffusing Capacity For Carbon monoxide (DLCO), were comparable across clusters ($p = 0.537$ and $p = 0.088$, respectively).

Articular manifestations showed partial variation. The number of swollen joints differed significantly among

clusters (*P*= 0.048), while painful joint count did not (*P*=0.172). Dyspnea severity assessed by the mMRC scale clearly distinguished clusters, with Clusters 3 and 4 demonstrating higher grades of breathlessness, whereas Clusters 1 and 2 were almost exclusively asymptomatic. These findings corresponded to the pulmonary fibrosis degrees (data were shown above).

Generally, Clusters 1 and 2 exhibited a higher prevalence of articular involvement compared with pulmonary manifestations, whereas Clusters 3 and 4 were characterized by predominant pulmonary involvement (Fig. **1**).

3.3. Comparison between RA Patients with and without ILD

Among the 306 patients, 101 (33.0%) were diagnosed with ILD. Sex distribution and age were comparable between the RA-ILD and RA-no-ILD groups. The median age was approximately 59–60 years in both groups (*P*= 0.948), and no significant difference in sex distribution was observed (*P*=0.192). Disease duration was significantly longer in patients with ILD (median 5.0 years) compared with those without ILD (median 3.0 years; p = 0.001). Stratified analysis further confirmed that longer disease duration was associated with ILD presence (*P*= 0.048).

Inflammatory markers showed mixed findings. CRP levels were significantly higher in the RA-ILD group (p = 0.014), and RF titers were also significantly elevated ($p < 0.001$). ACPA levels did not significantly differ between groups.

Respiratory symptoms were significantly more common in the RA-ILD group. Cough was reported in 16.8% versus 5.9% (p = 0.002), and dyspnea was markedly more frequent (63.4% *vs*. 19.0%, $p < 0.001$). mMRC grading demonstrated significantly greater breathlessness severity in the ILD group ($p < 0.001$).

In contrast, articular manifestations appeared less pronounced in patients with ILD (Table **2**, Fig. **2**). A higher proportion of ILD patients had no swollen joints (55.5% *vs*. 34.2%, p = 0.002), and fewer had more than 10 painful joints (12.9% *vs*. 21.0%, p = 0.008). Except for the number of deformed joints, the numbers of painful joints, swollen joints, and level of morning stiffness were significantly lower RA-ILD (3.00 [1.00;6.00], 0.00 [0.00; 2.00], 15.00 [5.00; 30.00], respectively) compared to RA-no-ILD (6.00 [2.00;10.00]; 2.00 [0.00; 4.00]; and 30.00 [20.00-60.00], respectively) ($p < 0.001$). The DAS28-CRP was calculated based on tender and swollen joints, showing that RA patients with ILD were more frequently in remission or low disease activity compared with patients without ILD (58.4% *vs*. 36.6%, $p < 0.001$). Treatments between groups were displayed in Table **S2**.

Table 2. Characteristics of RA patients according to ILD presence.

-		RA-ILD (n = 101)	RA-no-ILD (n = 205)	*p*-value
Sex	Male	9 (8.91%)	29 (14.15%)	0.192
	Female	92 (91.09%)	176 (85.85%)	
Age (years)	-	59.0 [53.0;66.0]	60.0 [52.0;67.0]	0.948[δ]
Age group	< 40 years	4 (3.96%)	18 (8.78%)	0.333
	40–49 years	10 (9.90%)	20 (9.76%)	
	50–59 years	40 (39.60%)	62 (30.24%)	
	60–69 years	30 (29.70%)	72 (35.12%)	
	≥ 70 years	17 (16.83%)	33 (16.10%)	
Medical history	Bronchiectasis	1 (0.99%)	6 (2.93%)	0.333[F]
	COVID-19 infection	28 (27.72%)	55 (26.83%)	0.869
	Family history of RA	9 (8.91%)	26 (12.68%)	0.330
Duration of illness (years)	-	5.00 [2.00;9.00]	3.00 [1.00;7.00]	0.001[δ]
Duration of illness (years)	< 3 years	38 (37.62%)	108 (52.68%)	0.048
	3–5 years	21 (20.79%)	39 (19.02%)	
	5–10 years	26 (25.74%)	41 (20.00%)	
	> 10 years	16 (15.84%)	17 (8.29%)	
CRP (mg/L)	-	30.2 [5.10;40.0]	11.2 [4.00;36.8]	0.014[δ]
RF (UI/mL)	-	73.1 [25.2;119]	30.2 [14.4;77.0]	< 0.001[δ]
DAS28CRP score	Remission to low disease activity	59 (58.42%)	75 (36.59%)	0.003
	Moderate to high disease activity	42 (41.58%)	130 (63.41%)	
Respiratory symptoms	Cough	17 (16.83%)	12 (5.85%)	0.002
	Dyspnea	64 (63.37%)	39 (19.02%)	< 0.001

-		RA-ILD (n = 101)	RA-no-ILD (n = 205)	*p*-value
mMRC scale	Stage 0	37 (36.63%)	167 (81.46%)	< 0.001[F]
	Stage 1	50 (49.50%)	32 (15.61%)	
	Stage 2	10 (9.90%)	2 (0.98%)	
	Stage 3	1 (0.99%)	3 (1.46%)	
	Stage 4	3 (2.97%)	1 (0.49%)	
Morning stiffness	-	15.0 [5.00;30.0]	30.0 [20.0; 60.0]	< 0.001[δ]
Number of painful joints	0	15 (14.85%)	11 (5.37%)	0.008
	1–10	73 (72.28%)	151 (73.66%)	
	> 10	13 (12.87%)	43 (20.98%)	
Number of swollen joints	0	56 (55.45%)	70 (34.15%)	0.002
	1–10	43 (42.57%)	129 (62.93%)	
	> 10	2 (1.98%)	6 (2.93%)	
ILD pattern on HRCT	UIP	70 (69.31%)	NA	NA
	NSIP	24 (23.76%)	NA	-
	OP	3 (2.97%)	NA	-
	Acute ILD	4 (3.96%)	NA	-

Note:[δ] Numerical data were presented as median [Q1; Q3]. Differences between groups were calculated by the Kruskal-Wallis test. Categorical data were presented as N (%). Differences between groups were calculated by Pearson's chi-square or [F]Fisher's exact test.

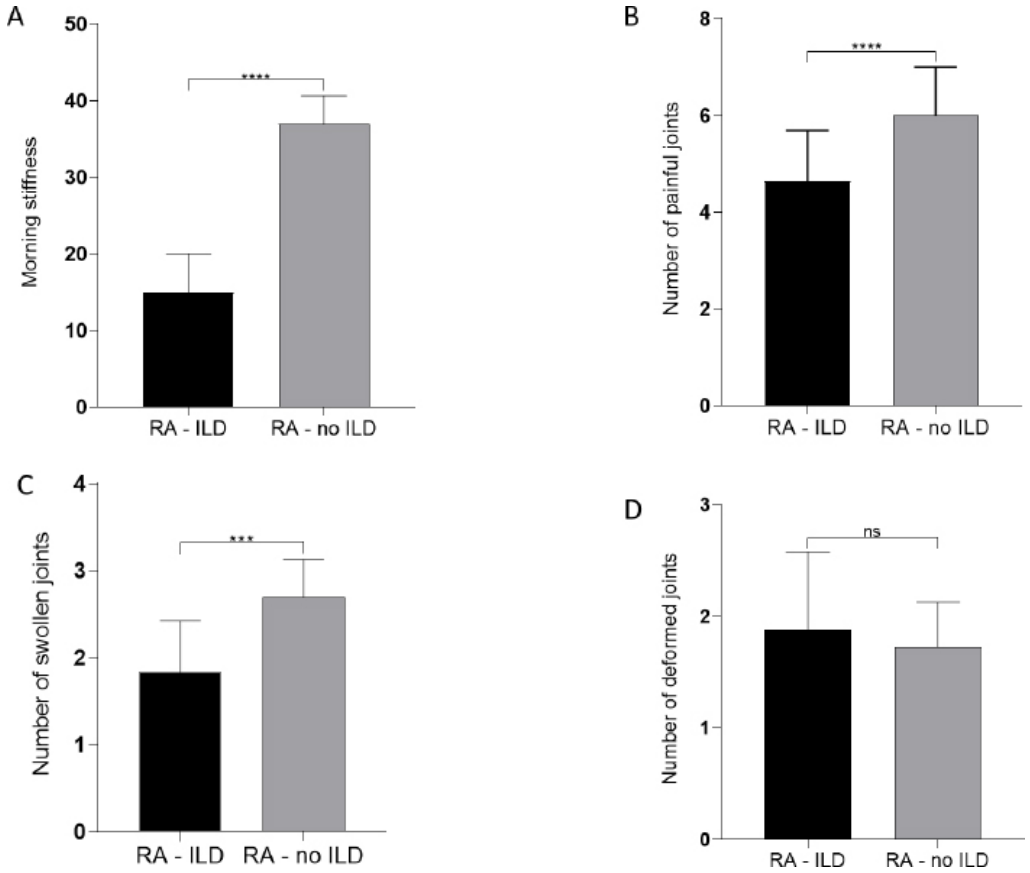

Fig. (2). Association of articular manifestations in RA patients according to ILD presence. The (**A**) degrees of morning stiffness, (**B**) number of painful joints, (**C**) number of swollen joints, and (**D**) number of deformed joints were calculated between the two groups RA-ILD and RA–noILD. Mann-Whitney U test was used to compare between groups.
*, *P*<0.05; **, *P*<0.01; ***, *P*<0.001. *ILD, interstitial lung disease; RA, rheumatoid arthritis; RA-ILD, RA patients with ILD; RA-noILD, RA patients without ILD.*

3.4. HRCT Patterns in RA-ILD

Among patients with ILD, the UIP pattern was the most frequent radiological subtype, observed in 69.3% of cases. NSIP accounted for 23.8%, while organizing pneumonia and acute ILD patterns were uncommon. These findings indicate a predominance of fibrotic phenotypes within the RA-ILD population.

4. DISCUSSION

Recent evidence indicates that Rheumatoid Arthritis (RA) is not a homogeneous entity but rather encompasses multiple distinct biological and clinical phenotypes. This study is the first to comprehensively characterize RA phenotypes in Vietnamese cohorts. By analyzing 306 patients with RA, we identified divergent articular and pulmonary phenotypes of RA with or without ILD. In addition, we identified 33% of cases with RA-ILD. This prevalence falls within the broad range reported in previous studies (10–50%) [21, 22], which reinforces that RA-ILD is a frequent extra-articular manifestation and should not be considered a rare complication.

Efforts have been made to well characterize RA phenotypes for personalizing treatment. Traditionally, RA can be divided into seropositive or seronegative phenotypes based on the presence of RF or ACPA [23, 24]. However, ACPA status may not fully capture the biological heterogeneity between ACPA– and ACPA+ RA subgroups [25]. Integrated multi-omic analyses (proteomic and metabolomic) of plasma samples have identified distinct immune and metabolic signatures between ACPA-positive and ACPA-negative RA [25]. At the tissue level, synovial phenotyping into lympho-myeloid, diffuse-myeloid, and pauci-immune/fibroid subtypes demonstrates differences in inflammatory burden, clinical manifestations, and therapeutic response. The pauci-immune phenotype highlights the contribution of non-inflammatory mechanisms in disease pathogenesis [26]. Furthermore, an increased vascular synovial phenotype has been associated with poorer response to JAK inhibitors [26]. At the immune-cell level, an imbalance in T helper cell subsets, characterized by reduced Th1 and Th17.1 and increased Th2, correlates with disease activity [27]. Comorbidity-based cluster analyses have identified distinct clinical phenotypes with significant differences in disease activity, treatment patterns, and prognosis [28]. Each phenotyping strategy contributes to elucidate the multi-layered heterogeneity in RA, spanning molecular, histopathological, immunological, and clinical domains, and highlight the potential of phenotype-driven stratification to support personalized management and prognostic assessment.

In this study, our analytical approach identified reciprocal articular and pulmonary phenotypes. Clusters 1 and 2 were characterized by more severe articular involvement but relatively limited pulmonary fibrosis, in contrast to Clusters 3 and 4, which exhibited greater pulmonary involvement. Consistently, patients With RA-Associated Interstitial Lung Disease (RA-ILD) presented with fewer swollen joints and lower pain scores compared with those without ILD. These findings suggest that pulmonary manifestations do not necessarily parallel the severity of peripheral joint disease at a single time point. Rather, the overall systemic inflammatory burden may be more closely linked to lung pathology than joint counts alone. Supporting this hypothesis, Clusters 3 and 4 demonstrated higher ESR levels, whereas Cluster 2 showed lower ESR values, male predominance, and less pulmonary involvement. Although pro-inflammatory cytokines were not measured in the present study, previous evidence has shown that mediators such as TNF-α, IL-6, and Platelet-Derived Growth Factor (PDGF) can promote fibroblast proliferation and extracellular matrix deposition [29–31], thereby providing a mechanistic link between synovial inflammation and pulmonary fibrosis. We also identified Cluster 2 as having a predominance of male patients, later RA onset, and the lowest burden of pulmonary lesions. Reported female-to-male ratios in RA range from 2:1 to 5:1 [32–36], although some studies have observed higher male proportions [37–39]. Such discrepancies likely reflect regional variation influenced by complex interactions among sex hormones, environmental exposures, and genetic background. However, longitudinal follow-up of this cluster is warranted to clarify its clinical trajectory.

Regarding extra-articular manifestations, RA-ILD accounted for approximately one-third of our RA cohort, with a mean patient age of around 60 years. This aligns with prior reports demonstrating that RA-ILD predominantly affects middle-aged and older individuals between the ages of 50 and 59, and age is an independent risk factor for ILD development [40]. Aging-related mechanisms, including epithelial senescence, telomere shortening, and impaired tissue repair, may contribute to increased fibrotic susceptibility in this population [41, 42]. Although RA itself is more prevalent in women, previous literature has suggested male sex as a predictor of the development of pulmonary disease, including ILD and bronchiectasis, in RA patients [43]. Our cohort included a lower proportion of males, and Cluster 2, despite male predominance, was associated with milder pulmonary disease. These differences compared with Western populations may reflect ethnic variability in disease susceptibility, but they may also indicate under-recognition or under-assessment of articular disease among Vietnamese men.

Several limitations should be considered. The cross-sectional design precludes assessment of causality or longitudinal progression. The study was conducted at a single center, which may limit generalizability. Additionally, pulmonary function trends and long-term survival outcomes were not evaluated. Although we examined the MUC5B rs35705950 variant in patients with RA, the relatively small number of variant carriers restricted meaningful subgroup analyses; therefore, these data were not included in the final analysis. Lastly, we did not evaluate the vimentin-specific antibodies between the two RA groups with and without ILD.

CONCLUSION

In conclusion, our cluster-based analysis highlights the substantial clinical heterogeneity of rheumatoid arthritis by identifying divergent articular- and pulmonary-dominant phenotypes. These findings suggest that joint and lung manifestations may not simply represent cumulative disease burden but rather reflect distinct expression patterns within RA. Further longitudinal and mechanistic studies are warranted to clarify the biological pathways underlying this divergence and to determine its prognostic and therapeutic implications.

AUTHORS' CONTRIBUTIONS

The authors confirm their contribution to the paper as follows: M.T.D.: Contributed to conceptualization, methodology, data curation, investigation, and writing (original draft and review and editing); H.B.B.: Involved in conceptualization, methodology, supervision, and writing (review and editing); M.T.B.: Contributed to methodology, data curation, and formal analysis; D.K.D.T.: Participated in methodology and writing (original draft and review and editing); H.K.T.T.: Contributed to data curation, formal analysis, and writing (original draft and review and editing). All authors reviewed the results and approved the final version of the manuscript.

LIST OF ABBREVIATIONS

ACR	=	American College of Rheumatology
ALAT	=	Latin American Thoracic Association
ACPA	=	Anti–Cyclic Citrullinated Peptide Antibody
ATS	=	American Thoracic Society
CI	=	Confidence Interval
COVID-19	=	Coronavirus Disease 2019
CRP	=	C-Reactive Protein
CT	=	Computed Tomography
DAS28-CRP	=	Disease Activity Score-28 (CRP-based)
DNA	=	Deoxyribonucleic Acid
EDTA	=	Ethylenediaminetetraacetic Acid
ERS	=	European Respiratory Society
ESR	=	Erythrocyte Sedimentation Rate
EULAR	=	European League Against Rheumatism
HRCT	=	High-Resolution Computed Tomography
ILD	=	Interstitial Lung Disease
IPF	=	Idiopathic Pulmonary Fibrosis
JRS	=	Japanese Respiratory Society
mMRC	=	Modified Medical Research Council Dyspnea Scale
MUC5B	=	Mucin 5B gene
OR	=	Odds Ratio
PCR	=	Polymerase Chain Reaction
RA	=	Rheumatoid Arthritis
RA-ILD	=	Rheumatoid Arthritis–Associated Interstitial Lung Disease
RF	=	Rheumatoid Factor
SD	=	Standard Deviation
SNP	=	Single Nucleotide Polymorphism
SPSS	=	Statistical Package for the Social Sciences
UIP	=	Usual Interstitial Pneumonia

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The study protocol was approved by the Ethics Committee Review Board of the Hue University of Medicine and Pharmacy (Approval No. H2022/519, approved by Prof. VT, Chair of the Ethics Committee).

HUMAN AND ANIMAL RIGHTS

All human research procedures followed were in accordance with the ethical standards of the committee responsible for human experimentation (institutional and national), and with the Helsinki Declaration of 1975, as revised in 2013.

CONSENT FOR PUBLICATION

Informed written consent was obtained from each patient prior to any study-related procedure.

AVAILABILITY OF DATA AND MATERIALS

All the data and supporting information is provided upon request.

STANDARDS FOR REPORTING

STROBE guidelines were followed.

FUNDING

None.

CONFLICT OF INTEREST

The authors declared no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

REFERENCES

[1] Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. Lancet 2016; 388(10055): 2023-38. http://dx.doi.org/10.1016/S0140-6736(16)30173-8 PMID: 27156434

[2] Firestein GS, McInnes IB. Immunopathogenesis of rheumatoid arthritis. Immunity 2017; 46(2): 183-96.

http://dx.doi.org/10.1016/j.immuni.2017.02.006 PMID: 28228278

[3] Cross M, Smith E, Hoy D, *et al*. The global burden of rheumatoid arthritis: Estimates from the Global Burden of Disease 2010 study. Ann Rheum Dis 2014; 73(7): 1316-22. http://dx.doi.org/10.1136/annrheumdis-2013-204627

[4] Deane KD, Holers VM. The natural history of rheumatoid arthritis. Clin Ther 2019; 41(7): 1256-69. http://dx.doi.org/10.1016/j.clinthera.2019.04.028 PMID: 31196652

[5] van der Helm-van Mil AH, Verpoort KN, Breedveld FC, Toes RE, Huizinga TW. Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis. Arthritis Res Ther 2005; 7(5): R949-58. http://dx.doi.org/10.1186/ar1767 PMID: 16207336

[6] Gabbay E, Tarala R, Will R, *et al*. Interstitial lung disease in recent onset rheumatoid arthritis. Am J Respir Crit Care Med 1997; 156(2 Pt 1): 528-35. http://dx.doi.org/10.1164/ajrccm.156.2.9609016 PMID: 9279235

[7] Zamora-Legoff JA, Krause ML, Crowson CS, Ryu JH, Matteson EL. Patterns of interstitial lung disease and mortality in rheumatoid arthritis. Rheumatology 2017; 56(3): 344-50. http://dx.doi.org/10.1093/rheumatology/kew391 PMID: 27940586

[8] 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(6): 1583-91. http://dx.doi.org/10.1002/art.27405

[9] Juge P, Solomon JJ, van Moorsel CHM, *et al*. MUC5B promoter variant rs35705950 and rheumatoid arthritis associated interstitial lung disease survival and progression. Semin Arthritis Rheum 2021; 51(5): 996-1004. http://dx.doi.org/10.1016/j.semarthrit.2021.07.002 PMID: 34411838

[10] 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): 2209-19. http://dx.doi.org/10.1056/NEJMoa1801562 PMID: 30345907

[11] Aletaha D, Neogi T, Silman AJ, *et al*. 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis 2010; 69(9): 1580-8. http://dx.doi.org/10.1136/ard.2010.138461 PMID: 20699241

[12] Perniola S, Chimenti M, Spinelli F, *et al*. Rheumatoid Arthritis from easy to complex disease: From the "2022 GISEA International Symposium". J Clin Med 2023; 12(8): 2781. http://dx.doi.org/10.3390/jcm12082781

[13] Minh Hoa TT, Darmawan J, Chen SL, Van Hung N, Thi Nhi C, Ngoc An T. Prevalence of the rheumatic diseases in urban Vietnam: A WHO-ILAR COPCORD study. J Rheumatol 2003; 30(10): 2252-6. PMID: 14528525

[14] Van Tuan N. Study of clinical, paraclinical characteristics and treatment situation of rheumatoid arthritis patients at Nghe An Friendship Multi-Specialty Hospital. Vietnam Med J 2021; 504(2) http://dx.doi.org/10.51298/vmj.v504i2.953

[15] Trí DM, Bảo HB, Tú THK. Survey of MUC5B gene mutation prevalence and risk factors of interstitial lung disease in rheumatoid arthritis. Vietnam Med J 2024; 167-72.

[16] Duong Minh Tri DM, Hoang Bui Bao HB, Trinh Hoang Kim Tu THK. Investigation of clinical and imaging characteristics of lung lesions in rheumatoid arthritis patients with interstitial lung disease. Hue Med Pharm J 2024; 14(5) http://dx.doi.org/10.34071/jmp.2024.5.26

[17] Raghu G, Remy-Jardin M, Myers JL, *et al*. Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med 2018; 198(5): e44-68. http://dx.doi.org/10.1164/rccm.201807-1255ST PMID: 30168753

[18] La D, Livesay DR. Predicting functional sites with an automated algorithm suitable for heterogeneous datasets. BMC Bioinformatics 2005; 6: 116. http://dx.doi.org/10.1186/1471-2105-6-116

[19] Gower JC. A general coefficient of similarity and some of its properties. Biometrics 1971; 27(4): 857-71. http://dx.doi.org/10.2307/2528823

[20] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987; 20: 53-65. http://dx.doi.org/10.1016/0377-0427(87)90125-7

[21] Doyle TJ, Lee JS, Dellaripa PF, *et al*. A roadmap to promote clinical and translational research in rheumatoid arthritis-associated interstitial lung disease. Chest 2014; 145(3): 454-63. http://dx.doi.org/10.1378/chest.13-2408 PMID: 24590021

[22] 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(10): 1700-6. http://dx.doi.org/10.1136/annrheumdis-2017-211138 PMID: 28611082

[23] Perera J, Delrosso CA, Nerviani A, Pitzalis C. Clinical phenotypes, serological biomarkers, and synovial features defining seropositive and seronegative rheumatoid arthritis: A literature review. Cells 2024; 13(9): 743. http://dx.doi.org/10.3390/cells13090743 PMID: 38727279

[24] Kimbrough BA, George RJ, Crowson CS, *et al*. Long-term outcomes in seronegative rheumatoid arthritis. Arthritis care res 2025. http://dx.doi.org/10.1002/acr.70011 PMID: 41386743

[25] Hur B, Gupta VK, Oh M, *et al*. Integrative multi-omic profiling in blood reveals distinct immune and metabolic signatures between ACPA-negative and ACPA-positive rheumatoid arthritis. Front immunol 2025; 16: 1667662. http://dx.doi.org/10.3389/fimmu.2025.1667662 PMID: 41235232

[26] Giollo A, Salvato M, Frizzera F, *et al*. Clinical application of synovial biopsy in noninflammatory and persistent inflammatory refractory rheumatoid arthritis. Ann rheum dis 2026; 85(1): 91-102. http://dx.doi.org/10.1016/j.ard.2025.07.023 PMID: 40846588

[27] Xue M, Lin H, Lynch T, *et al*. Th cell phenotypes and their correlations with disease activity in patients with rheumatoid arthritis. J clin med 2025; 14(12) http://dx.doi.org/10.3390/jcm14124220 PMID: 40565964

[28] Crowson CS, Atkinson EJ, Kronzer VL, *et al*. Comorbidity clusters in patients with rheumatoid arthritis identify a patient phenotype with a favourable prognosis. Ann Rheum Dis 2024; 83(5): 556-63. http://dx.doi.org/10.1136/ard-2023-225093 PMID: 38331589

[29] Żurawek M, Ziółkowska-Suchanek I, Iżykowska K. Fibrosis in immune-mediated and autoimmune disorders. J Clin Med 2025; 14(18): 6636. http://dx.doi.org/10.3390/jcm14186636 PMID: 41010839

[30] Rosengren S, Corr M, Boyle DL. Platelet-derived growth factor and transforming growth factor beta synergistically potentiate inflammatory mediator synthesis by fibroblast-like synoviocytes. Arthritis Res Ther 2010; 12(2): R65. http://dx.doi.org/10.1186/ar2981

[31] Usher KM, Zhu S, Mavropalias G, Carrino JA, Zhao J, Xu J. Pathological mechanisms and therapeutic outlooks for arthrofibrosis. Bone Res 2019; 7: 9. http://dx.doi.org/10.1038/s41413-019-0047-x PMID: 30937213

[32] Trang TTH, Hung NV, Phuong PT. Study of some factors related to interstitial pneumonia in patients with rheumatoid arthritis. J Med Res 2021; 139(3): 29-36. http://dx.doi.org/10.52852/tcncyh.v139i3.120

[33] Iqbal Z, Khan JA, Hassan MM, *et al*. Frequency of interstitial lung disease in rheumatoid arthritis patients: A hospital-based study. Cureus 2022; 14(10): e30145. http://dx.doi.org/10.7759/cureus.30145 PMID: 36381717

[34] Fadda S, Khairy N, Fayed H, Mousa H, Taha R. Interstitial lung disease in Egyptian patients with rheumatoid arthritis: Frequency, pattern and correlation with clinical manifestations and anti-citrullinated peptide antibodies level. Egypt Rheumatol 2018; 40(3): 155-60. http://dx.doi.org/10.1016/j.ejr.2017.10.006

[35] Juge P, Borie R, Kannengiesser C, *et al*. Shared genetic predisposition in rheumatoid arthritis-interstitial lung disease and familial pulmonary fibrosis. Eur Respir J 2017; 49(5): 1602314. http://dx.doi.org/10.1183/13993003.02314-2016

[36] Koduri G, Norton S, Young A, *et al*. Interstitial lung disease has a poor prognosis in rheumatoid arthritis: Results from an inception cohort. Rheumatol 2010; 49(8): 1483-9. http://dx.doi.org/10.1093/rheumatology/keq035 PMID: 20223814

[37] Kim Y, Yang H, Kim K. Etiology and pathogenesis of rheumatoid arthritis-interstitial lung disease. Int J Mol Sci 2023; 24(19): 14509. http://dx.doi.org/10.3390/ijms241914509

[38] Klester E, Klester K, Shoykhet Y, *et al*. Risk factors of interstitial lung diseases in patients with rheumatoid arthritis. Eur Respir J 2019; 54(suppl 63) http://dx.doi.org/10.1183/13993003.congress-2019.PA1365

[39] Restrepo JF, del Rincón I, Battafarano DF, Haas RW, Doria M, Escalante A. Clinical and laboratory factors associated with interstitial lung disease in rheumatoid arthritis. Clin rheumatol 2015; 34(9): 1529-36. http://dx.doi.org/10.1007/s10067-015-3025-8 PMID: 26255186

[40] Assayag D, Lubin M, Lee JS, King TE, Collard HR, Ryerson CJ. Predictors of mortality in rheumatoid arthritis-related interstitial lung disease. Respirol 2014; 19(4): 493-500. http://dx.doi.org/10.1111/resp.12234 PMID: 24372981

[41] Liu R, Liu G. Cell senescence and fibrotic lung diseases. Exp gerontol 2020; 132: 110836. http://dx.doi.org/10.1016/j.exger.2020.110836 PMID: 31958492

[42] Ku JC, Raiten J, Li Y. Understanding fibrosis: Mechanisms, clinical implications, current therapies, and prospects for future interventions. Biomed Eng Adv 2024; 7: 100118. http://dx.doi.org/10.1016/j.bea.2024.100118

[43] Aubart F, Crestani B, Nicaise-Roland P, *et al*. High levels of anti-cyclic citrullinated peptide autoantibodies are associated with co-occurrence of pulmonary diseases with rheumatoid arthritis. J rheumatol 2011; 38(6): 979-82. http://dx.doi.org/10.3899/jrheum.101261 PMID: 21362759