The Open Rheumatology Journal
DOI: 10.2174/0118743129499823260727114010, 2026, 20, e18743129499823

ISSN: 1874-3129
1

SUPPLEMENTARY MATERIAL **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*

*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

DNA extraction

Following the manufacturer's instructions, a GeneJET genomic DNA purification kit (Thermo Fisher Scientific, USA) was used to extract the total genomic DNA from peripheral blood samples.

PCR amplification

CLC Main Workbench 5.5 software was used to designed primers to amplify *MUC5B* (rs35705950). The forward primer sequence was 5′-GGCTCAGACTCCGTCAGAG -3′, and the reverse primer sequence was 5′- ATTGAGACATCCCGGATGCG -3′. A 15-µL PCR reaction mixture was prepared with 1.5 µL of 10× PCR buffer, 1.5 µL of 2.5 mM dNTPs, 1.5 µL of each primer, 1.0 µL of template DNA, 9.4 µL of H_2O, and 0.1 µL of Taq DNA polymerase. The PCR conditions were as follows: 3 min at 98°C, denaturation for 10s at 98°C, annealing for 20s at 58°C, and extension for 60s at 72°C. The last three steps were repeated for 35 cycles, followed by a final extension for 2 min at 72°C. PCR amplification generated a fragment of 231 base pairs in length.

DNA sequencing

Direct sequencing of the purified PCR amplicons was performed using the ABI BigDye Terminator v3.1 Cycle Sequencing Kit, followed by analysis on an ABI 3500 Genetic Analyzer (Applied Biosystems, USA). The data were analyzed using CLC Main Workbench version 5.5 software to detect the MUC5B mutation.

Table S1. Characteristics of the RA cohorts.

Variables	Total (N = 306)
Age (years old)	59.00 [53.00; 66.00]
Female (N, %)	268 (87.6%)
Duration of illness (years)	4.00 [1.00; 7.00]
Onset age (years)	54.00 [46.00; 62.00]
Family history RA (N, %)	35 (11.4%)

Table S2. Characteristics of treatments between groups of RA patients with or without ILD presence.

Treatments		RA-ILD (n = 101)	RA-no-ILD (n = 205)	*p* value
Methotrexate	No	16 (15.84%)	74 (36.10%)	<0.001
	Yes	85 (84.16%)	131 (63.90%)	
HCQ	No	97 (96.04%)	193 (94.15%)	0.67[F]
	Yes	4 (3.96%)	12 (5.85%)	
Sulfasalazine	No	90 (89.11%)	187 (91.22%)	0.7
	Yes	11 (10.89%)	18 (8.78%)	
Biologics	No	90 (89.11%)	195 (95.12%)	0.09
	Yes	11 (10.89%)	10 (4.88%)	
Corticoids	No	21 (20.79%)	69 (33.66%)	0.029
	Yes	80 (79.21%)	136 (66.34%)	
NSAIDs	No	86 (85.15%)	147 (71.71%)	0.014
	Yes	15 (14.85%)	58 (28.29%)	

Note: Categorical data were presented as N (%). Differences between groups were calculated by Pearson's chi square or [F]Fisher's exact test.

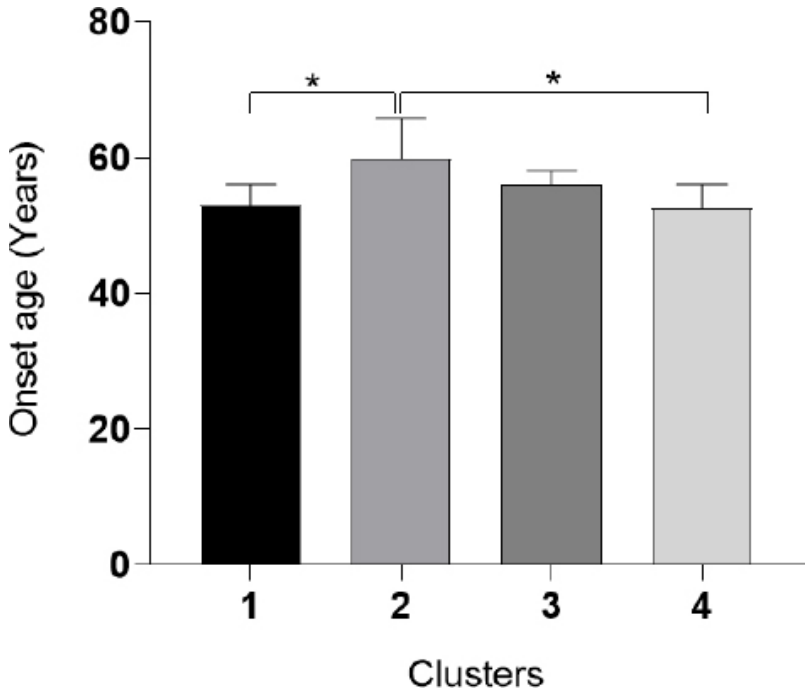

Fig. (S1). Characteristics of onset age (years old) of RA between clusters. Data was shown as median ± SEM. The Kruskal–Wallis test was used to assess differences between groups, followed by Dunn's post hoc test for multiple comparisons. *, $P<0.05$

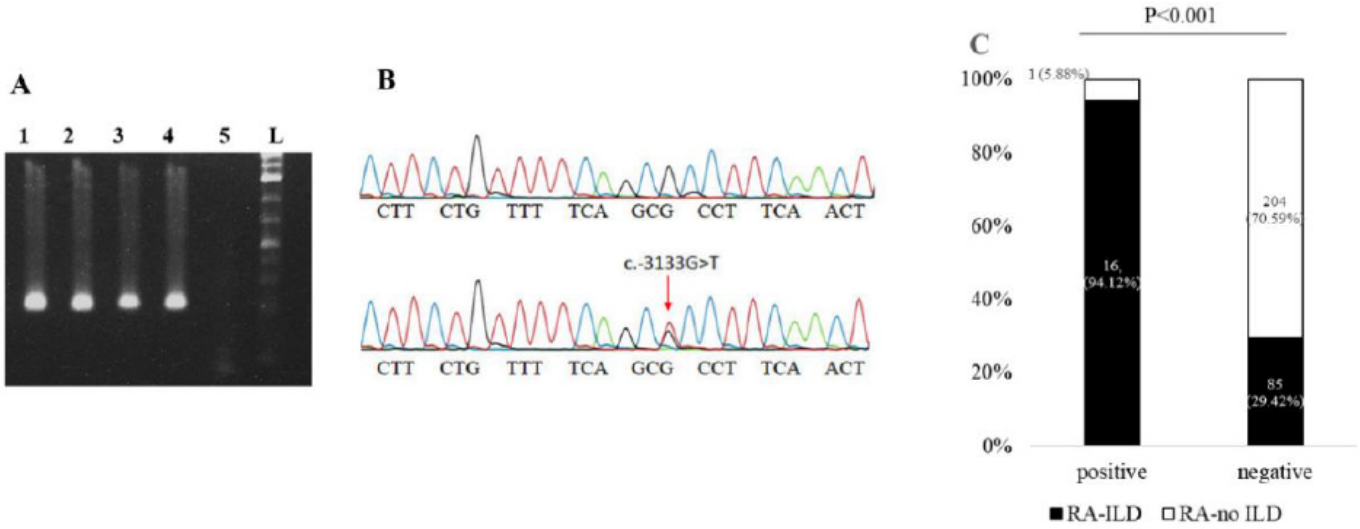

Fig. (S2). DNA sequencing results showing the mutation in MUC5B.**(A)**PCR results from MUC5B amplification: lanes 1–4, amplified MUC5B products from patient DNA samples; lane 5, the negative control obtained with H2O; and lane L, DNA marker. The expected product size was approximately 231 base pairs (bp).**(B)**Chromatograms obtained from Sanger sequencing reveal the presence of the rs35705950 variant in MUC5B. The red arrow indicates the specific location of the mutation within the gene. (C) The association of the MUC5B rs35705950 variant with ILD among RA patients; p values were determined using Pearson's chi-squared test.