



OPEN Polygenic risk scores of rheumatoid arthritis associated with seropositivity and bone erosions in a Taiwanese population

Ting-Shuan Wu^{1,2,3}, Yen-Ju Chen^{3,4,5}, Chia-Ni Hsiung⁶, Chien-Lin Mao³, Chia-Yi Wei³, I-Chieh Chen³, Chung-Mao Kao^{4,7}, Tzu-Hung Hsiao³, Wen-Nan Huang^{4,5,8}, Yi-Hsing Chen^{4,5,8} & Yi-Ming Chen^{3,4,5,8,9,10}✉

Polygenic risk scores (PRS) are widely used to estimate disease risks and predict clinical outcomes. However, ethnic differences exist worldwide. We aimed to use PRS for predicting the development of rheumatoid arthritis (RA) in a Taiwanese population and investigate whether PRS of RA may be associated with structural damage. PRS was calculated via GWAS data from 2,042 RA patients and 7,950 controls from the Taiwan Precision Medicine Initiative. LDpred2, PLINK, and PRSice-2 models were used to evaluate RA susceptibility via AUC. Clinical factors, including rheumatoid factor (RF), anti-citrullinated protein antibody (ACPA), medication, and bone erosion identified by X-ray or ultrasound, were compared across PRS quartiles. PRS derived from 97,396 SNPs via LDpred2 had the highest AUC. Participants in the highest quartile of RA-PRS had the highest proportions of rheumatoid factor (RF) and anti-citrullinated protein antibody (ACPA) positivity (74.8% and 65.0%, respectively), bone erosion (86.4%), and use of bDMARDs or tsDMARDs (42.3%) compared to the other quartiles. PRS remained a significant predictor of bone erosion, particularly in patients under 60 years of age, even after adjusting for RF, ACPA, and treatment. PRS is associated with seropositivity, erosive bone disease, and need for advanced therapy among Taiwanese patients with RA.

Keywords PRS, RA, Seropositivity, Rheumatoid factor, Anti-citrullinated protein antibody, Bone erosion

Abbreviations

PRS	Polygenic risk scores
RA	Rheumatoid arthritis
AUC	Area under curve
ACPA	Anti-citrullinated protein antibody
bDMARDs	Biologic disease modifying anti-rheumatic drugs
tsDMARDs	Targeted synthetic disease modifying anti-rheumatic drugs
SNPs	Single nucleotide polymorphisms
GWAS	Genome-wide association studies
HLA	Human leukocyte antigen
TPMI	Taiwan precision medicine initiative
VGHTC	Taichung veterans general hospital

¹Department of Biomedical Sciences, Chung Shan Medical University, Taichung 40201, Taiwan. ²Department of Medical Research, Chung Shan Medical University Hospital, Taichung 40201, Taiwan. ³Department of Medical Research, Taichung Veterans General Hospital, Taichung 40705, Taiwan. ⁴Division of Allergy, Immunology and Rheumatology, Department of Internal Medicine, Taichung Veterans General Hospital, 1650, Section 4, Taiwan Boulevard, Xitun Dist, Taichung City 40705, Taiwan. ⁵School of Medicine, National Yang Ming Chiao Tung University, Taipei 11221, Taiwan. ⁶Data Science Statistical Cooperation Center, Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan. ⁷Division of Translational Medicine, Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan. ⁸Department of Post-Baccalaureate Medicine, College of Medicine, National Chung Hsing University, Taichung, Taiwan. ⁹Graduate Institute of Clinical Medicine, College of Medicine, National Chung Hsing University, Taichung 402202, Taiwan. ¹⁰Precision Medicine Research Center, College of Medicine, National Chung Hsing University, Taichung 402202, Taiwan. ✉email: ymchen1@vghtc.gov.tw

GCR	Genomic control rate
MAF	Minor allele frequency
LD	Linkage disequilibrium
RF	Rheumatoid factor
SD	Standard deviation
ANOVA	Analysis of variance
OR	Odds ratios
CI	Confidence intervals
PC	Principal component
CEU	Utah residents with northern and western european ancestry
ASW	African ancestry from southwest USA
MKK	Maasai from kinyawa, kenya
MEK	Mende from sierra leone
CHD	Chinese from metropolitan denver
CHB	Han chinese from beijing, china
JPT	Japanese from tokyo
LWK	Luhya from webuye, kenya
TSI	Toscani from italia
GIH	Gujarati indians from houston
YRI	Yoruba from ibadan, nigeria
YORA	Younger age at onset under 60 years
EORA	Elderly age at onset over or equal to 60 years
SE	Shared epitope

Rheumatoid arthritis (RA) is an autoimmune rheumatic disease associated with chronic inflammation that requires early diagnosis and timely treatment to prevent structural deformities and bone erosions¹. During the last decade, the advancement of biologic and targeted synthetic disease modifying anti-rheumatic drugs (bDMARDs and tsDMARDs) has provided an opportunity to control synovial inflammation and greatly reduced the potential threats of permanent joint damage^{2,3}. However, a substantial number of patients with RA still fail to respond to bDMARDs or tsDMARDs and develop bone erosions and functional disabilities^{2,3}. Rheumatoid factor (RF) and anti-citrullinated protein antibody (ACPA) seropositivity is associated with bone destruction and appear to be clinical predictors for structural damages⁴. However, RF and ACPA alone are not sufficient to identify all radiographic progressions and disease remission in patients with RA^{5,6}. Therefore, it is imperative to optimize the prediction model for joint erosion to guide therapeutic strategies in patients with RA.

The inheritance of RA is often in a familial aggregation pattern, indicating a genetic contribution in its pathogenesis⁷. Previous genome-wide association studies (GWAS) identified genetic variations in the human leukocyte antigen (HLA) and non-HLA regions associated with the development of RA^{8,9}. Moreover, gene alleles at HLA-DRB1 and non-HLA genes were also linked to joint destructions¹⁰ indicating a potential role of genomic variation in the pathogenesis and prediction of structural damages in RA.

Polygenic risk score (PRS) is a novel technique to calculate and integrate the summative association of multiple genetic loci with clinical phenotypes¹¹. PRS for RA has recently been widely investigated across ethnicity groups^{12–15}. Genetic predisposition in terms of PRS predicted the propensity among patients with pre-RA condition to develop RA¹². Moreover, RA patients within the highest quartiles of PRS exhibited the greatest risk of radiographic progression in a Japanese cohort¹⁵. However, PRS derived from different ethnicity groups may have different prediction powers¹³. To date, no RA-PRS generated from the Taiwanese population has been reported. Furthermore, whether RA-PRS is associated with the treatment patterns of bDMARDs and tsDMARDs remains unknown.

In this study, we aimed to establish a Taiwanese-ancestry specific PRS for RA in a hospital-based cohort. The primary objective was to determine the association of RA-PRS established from a Taiwanese population with RF/ACPA seropositivity, bone erosions, and RA treatment.

Methods

Study design, participants, and ethics statement

The study population was enrolled in the Taiwan Precision Medicine Initiative (TPMI) from **January 2010 to August 2022**. The TPMI study cohort ($N=43,035$) consists of individuals who participated in TPMI **before September 2021**, with a follow-up period of **at least one year**. The outcomes were last updated in **August 2022**, ensuring that our data reflects the most recent clinical information. The genetic profiles of TPMI participants were recorded and linked to their electronic health records, which included physical measurements, blood biochemical analyses, diagnostic information, and genotyping. In this study, we included patients diagnosed with RA, identified using International Classification of Diseases (ICD)–9 and ICD-10 codes (714, M05, M06), recorded more than once during inpatient admissions and twice in outpatient clinics. We excluded participants with ICD-10 code M06.1 and those younger than 20 years old. Our cohort's medical records were available longitudinally, allowing for a thorough evaluation of outcomes. A total of 2,183 arthritis (RA) patients were selected based on the 2010 American College of Rheumatology and the European League Against Rheumatism criteria for the classification of RA¹⁶. A control group of 8,732 individuals was selected by matching age, sex, and periodontitis symptoms with cases at a 1:4 ratio¹⁷. After removing the sex check error, heterozygosity, ancestry differences, and related samples from the 10,915 patients after matching, the final number of participants in the study was 9,992.

All eligible participants from all cohorts were required to complete a written informed consent form by following the protocols approved by their institutional ethical committees. The institutional review boards of Taichung Veterans General Hospital approved this study (No. SF19153A) and ensured compliance with the Declaration of Helsinki. We have complied with all relevant ethical regulations.

Genotyping and quality controls

The blood samples of patients were collected for DNA extraction and genotyped at Taichung Veterans General Hospital (VGHTC). The Affymetrix TWB 2.0 SNP chip contains 714,431 SNPs and was designed specifically for Taiwan's Han Chinese population. A quality control procedure was performed to exclude markers that were present on the sex chromosomes, had a genotype missingness rate $< 5\%$, had a minor allele frequency (MAF) $< 1\%$, failed the Hardy-Weinberg equilibrium tests with $P\text{-value} < 1.1 \times 10^{-7}$, and different genotype missing rates in case and control group. A total of 435,605 SNPs were retained after the quality control. Post-QC, we conducted a principal component analysis (PCA) to adjust for potential population stratification.

Genetic construction of polygenic risk scores

The class method of polygenic risk score (PRS) calculation is by computing the sum of effect alleles from the individual's genotype, weighted by the risk allele effect sizes as estimated by a GWAS on the phenotype. The PRS were calculated using summary statistics from a rheumatoid arthritis (RA) training cohort of 4,996 individuals. We chose to handle the complexity of the HLA region by incorporating only selected TagSNPs from this region. To handle the issue of linkage disequilibrium (LD), in this study, we evaluated the performance of polygenic risk scores (PRS) using three different methods: Plink, LDpred2, and PRSice¹¹. We controlled the LD issue by clumping and threshold (C+T) in plink^{18–20}. Plink used clumping $R^2 < 0.1$ and thresholding (C+T) with $P\text{-value} < 10^{-5}$ to control LD. LDpred2, which used the SNPs from the TPMI datasets and HapMap3, included 97,396 SNPs. LDpred2 did not use the C+T (clumping and threshold) method to select PRS SNPs, but rather a Bayesian approach and auto-regression model for constructing polygenic risk scoring. For PRSice-2, the threshold $p\text{-value}$ was set at 5×10^{-8} , and R^2 was 0.2, following the C+T method for scoring PRS.

The capacity of the polygenic risk score to predict RA development was evaluated using the area under the curve (AUC) for PRS calculated by the LDpred2, PLINK and PRSice-2 packages in R. In addition, using logistic regression analysis, we constructed quartile plots to examine the risk of RA development according to the quartile of polygenic risk score. The quartile was based on significance levels, with the top quartile indicating the highest and the bottom quartile displaying the lowest polygenic risk scores.

Age stratification

Studies on autoimmune diseases have shown that PRS are associated with age at onset^{21,22}. Moreover, disease onset age significantly affects the treatment and prognosis of RA^{15,23}. Therefore, patients were classified into two groups according to their age: the younger-onset group (YORA) for those with onset under 60 years, and the elderly-onset group (EORA) for those aged 60 years or older.

Outcome measurement

The primary outcome, bone erosions were assessed using both X-ray and ultrasonography. Plain films of both the hands and feet were obtained in the postero-anterior view for all participants, in line with the criteria outlined in the original Sharp score²⁴. For ultrasonography, we utilized the dorsal longitudinal view to examine all metacarpophalangeal (MCP) and metatarsophalangeal (MTP) joints. This technique aligns with the standards set forth by the Outcome Measures in Rheumatology Clinical Trials (OMERACT) group²⁵. The images obtained from both X-ray and ultrasonography were reviewed by two experts in the field: a rheumatologist and a radiologist. The presence of bone erosion was determined based on the identification of erosive changes in either modality—X-ray or ultrasound.

Covariates

Demographic data, comorbidities, rheumatoid factor (RF), anti-citrullinated protein antibody (ACPA), RA disease activity and treatment with biologic and targeted synthetic disease-modifying anti-rheumatic drugs (bDMARDs and tsDMARDs) were extracted from the electronic health record.

Periodontitis symptoms in our study were identified using ICD-9 codes 523.3-5 and ICD-10 codes K05.2-6, along with relevant procedure codes for periodontitis management. This identification was based on the presence of these codes in patients' medical records more than once during inpatient admissions or more than twice in outpatient clinic visits.

Statistical analyses

Quantitative variables are expressed as the mean \pm standard deviation (SD). A Student's *t*-test or analysis of variance (ANOVA) was used to compare the means of continuous variables. The statistical differences in the relationships between the categorical variables between the cases and controls were analyzed using the chi-square test. A logistic regression model was applied to analyze the association between quartiles of PRS and seropositivity of RF, ACPA, targeted therapy, and bone erosions with adjustment of age, sex and smoking status. Moreover, we evaluated factors associated with bone erosion by the enter mode of multivariate logistic regression with covariates of RA-PRS, seropositivity and targeted therapy in both EORA and YORA groups. The odds ratios (OR) and 95% confidence intervals (CI) were calculated. All statistical analyses were performed in SAS version 9.4 (SAS Institute Inc. Cary NC) and significance was set at $p < 0.05$.

Results

Patient characteristics

The patients were randomly divided into training and validation groups, each of which consisted of RA patients and controls matched on age, gender, and proportion of periodontal disease. The training set was used for the selection of the best model, and the testing set was used for validation of the data (Supplementary Fig. 1A). We performed a GWAS to identify SNPs associated with RA. As shown in Supplementary Fig. 1B, the main SNPs at chromosome 6 were the HLA regions in all RA participants, and similar results were shown in the training group (Supplementary Fig. 1C). To establish the genetic distinctiveness of each ethnic group, we projected the genotypes of different ethnic groups on the principal component (PC) (Supplementary Fig. 2). We included the VGHTC cohort and 11 other cohorts, with the final study population comprising 2,042 patients with RA, 1,115 control individuals of CEU (Utah residents with Northern and Western European ancestry), ASW (African ancestry from southwest USA), MKK (Maasai from Kinyawa, Kenya), MEK (Mende from Sierra Leone), CHD (Chinese from metropolitan Denver), CHB (Han Chinese from Beijing, China), JPT (Japanese from Tokyo), LWK (Luhya from Webuye, Kenya), TSI (Toscani from Italia), GIH (Gujarati Indians from Houston), and YRI (Yoruba from Ibadan, Nigeria) ethnicities, and 7,950 TPMI controls. The results showed that while the genotypes of our study population, VGHTC, were similar to those of the JPT, CHB, and CHD ethnic groups in Asia, there were still distinct differences among them. However, the genetic variations between VGHTC and other ethnic populations were more substantial. Thus, it is crucial to establish a Taiwanese-specific PRS for the prediction of RA development.

PRS algorithms establishment

To establish the optimal PRS for RA in a hospital-based cohort, the training group was used to select the best algorithms. The RA susceptibility was evaluated by the AUC. In the training group, the AUC values of LDpred2, PLINK, and PRSice-2 were 0.751, 0.637, and 0.611, respectively (Fig. 1A). Although there were no statistically significant differences in the testing group among the AUCs using these three methods (Fig. 1B), the AUC of LDpred2 consisting of 97,396 SNPs was the numerically highest value (Fig. 1C). Therefore, LDpred2 was selected for the subsequent analyses.

Clinical characteristics of enrolled patients with RA by PRS quartiles

The RA patients were divided into quartiles by PRS score (RA-PRS), the lowest score group was the first quartile, and as the PRS score increased, it was divided into the second, third and the top groups. The patients' characteristics, including demographic data, seropositivity of rheumatoid factor (RF), anti-citrullinated protein antibody (ACPA), and treatment with targeted therapy including biologic and targeted synthetic disease modifying anti-rheumatic drugs (bDMARDs and tsDMARDs) are recorded in Table 1. Age of onset is known to be associated with differences in clinical features and prognosis^{15,23}. However, whether the age of onset is associated with RA-PRS remains unknown. Analyzing by RA-PRS grouping, the proportion of YORA was 70.84% among patients in the first quartile group and increased to 78.28% in the highest quartile (Fig. 2). In contrast, the proportion of EORA decreased from 29.16% to 21.72%. Because PRS appeared to be associated with RA onset age, we focused on the YORA group.

Autoantibody profiles by PRS quartiles

In the autoantibody analysis, RF positivity significantly increased from 57.37% to 74.84% in the first quartile to the top quartile of the RA-PRS score, $p < 0.0001$ (Table 1). Similarly, the proportion of ACPA positivity increased from 38.83% to 64.98%, $p < 0.0001$ (Table 1). The odds ratio was used to evaluate the risk factor, while the first quartile was used as the reference. The odds ratios of RF positivity significantly escalated with RA-PRS quartiles in the YORA group from 1.19 (95% CI: 0.87–1.62, $p = 0.27$) to 2.39 (95% CI: 1.72–3.22, $p < 0.001$) (Fig. 3A and Supplementary Table 1). Similarly, the odds ratios of ACPC positivity were upregulated from 1.35 (95% CI: 0.98–1.87, $p = 0.08$) to 2.85 (95% CI: 2.06–3.96, $p < 0.001$) (Fig. 3B and Supplementary Table 1). Additionally, the odds ratios of APCA positivity underwent a significant increase in the EORA group from 1.99 (95% CI: 1.11–3.56, $p = 0.02$) to 3.69 (95% CI: 1.98–6.88, $p < 0.001$), but significant changes were only observed in top quartile of RF positivity, 1.91 (95% CI: 1.05–3.48, $p = 0.03$) (Supplementary Table 1). The results indicated that in the YORA group, the positive rates and odds ratios of RA and ACPA served as significant identification markers above the 3rd quantile of RA-PRS.

RA treatment pattern by PRS quartiles

We then investigated whether RA-PRS was associated with RA treatment. As shown in Table 1, the proportion of glucocorticoid and conventional synthetic disease-modifying anti-rheumatic drugs (csDMARDs) treatments (i.e. Methotrexate, Hydroxychloroquine, Sulfasalazine, Cyclosporin, and Leflunomide) increased with PRS quartiles. In addition, we also found that the proportion of experienced bDMARDs or tsDMARDs treatments also increased with PRS scores, with 26.42% in the first quartile and 42.27% in the top quartile. In addition, the proportion of RA patients who had underwent two or more bDMARDs or tsDMARDs ranged from 9% in the first quartile to 16.63% in the top quartile (Table 1). Figure 3C and Supplementary Table 1 depicted the age and sex-adjusted risk for the necessity of targeted therapies in patients with RA by RA-PRS categories. We found that the need for targeted therapies was associated with RA-PRS, with the OR for the top quartile vs. the first quartile in YORA and EORA groups, 1.71 (95%CI: 1.27–2.31, $p = 0.0004$) and 3.31 (95% CI: 1.86–5.89, $p < 0.0001$), respectively. The results indicated that higher PRS scores significantly increased the likelihood of using targeted therapy to control inflammation in patients with RA.

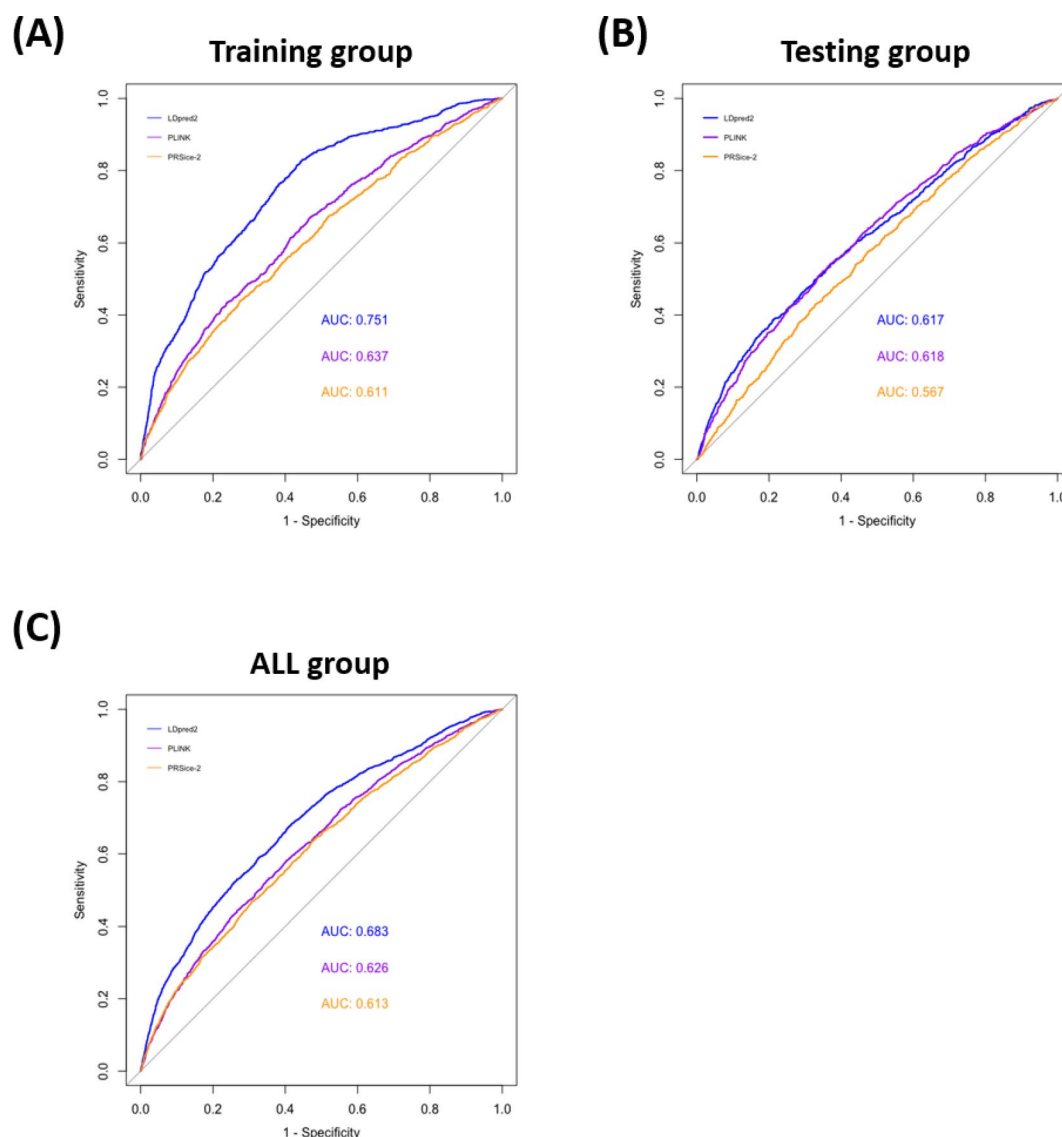


Fig. 1. Algorithm establishment of polygenic risk scores in predicting rheumatoid arthritis (RA) development. LDpred2 (blue), PLINK (purple), and PRSice-2 (yellow) methods were used to estimate the RA susceptibility. Area under the curve (AUC) analyses for assessment of the accuracy of the model in predicting RA development in the training set (A), the testing set (B), and in all patients (C).

Association of bone erosions and RA-PRS

Bone erosion, detected by either plain X-ray or ultrasonography, is one of the most crucial complications of RA that warrants intensive treatment and monitoring to prevent further structural damages and functional disability². We discovered that RA-PRS was associated with the prevalence of bone erosion, ranging from 77.07% in the first quartile to 86.44% in the top quartile ($p=0.01$) (Table 1). It is worthnoting that patients in the third and top quartiles of RA-PRS exhibited a significant risk of bone erosion compared to those in the first quartile (OR: 2.21 [95% CI 1.25–3.91, $p=0.01$], and OR: 1.84 [95% CI 1.08–3.11, $p=0.02$]) in the YORA group (Fig. 3D). However, the tendency of higher RA-PRS associated with bone erosion was not observed in the EORA group. (Supplementary Table 1).

Multiple logistic regression was used to analyze variables associated with bone erosions, with adjustments for RA-PRS, ACPA/RF positivity, and target therapy. We demonstrated that RA-PRS (OR: 1.86, 95% CI: 1.12–3.1, $p=0.02$), ACPA positivity (OR: 3.11, 95% CI: 1.82–5.32, $p<0.001$), and treatment with targeted therapies (OR: 2.73, 95% CI: 1.75–4.27, $p<0.001$) were independent predictors for bone erosions in the YORA group (Fig. 4 and Supplementary Table 2). However, only ACPA seropositivity (OR: 3.32, 95% CI: 1.04–10.58, $p=0.04$) was associated with bone erosion in patients with EORA. Taken together, our results suggested that, in addition to autoantibody positivity, RA-PRS could be a valuable predictor for structural damage.

RA PRS (N= 2042)									
	1 st quartile (n = 511)		2nd quartile (n = 510)		3rd quartile (n = 510)		top quartile (n = 511)		p value
	N	%	N	%	N	%	N	%	
Demography									
Age	57.71	14.13	56.84	12.89	58.36	12.87	56.64	13.91	0.15
RA Age									0.04
< 60	362	70.84	389	76.27	386	75.69	400	78.28	
≥ 60	149	29.16	121	23.73	124	24.31	111	21.72	
Gender									0.34
Man	107	20.94	96	18.82	108	21.18	120	23.48	
Female	404	79.06	414	81.18	402	78.82	391	76.52	
BMI (mean)	23.7	4.10	24.3	4.41	23.9	4.14	23.6	4.06	0.04
< 24	320	62.62	292	57.25	304	59.61	322	63.01	0.17
≥ 24	191	37.38	218	42.75	206	40.39	189	36.99	
Smoking	40	7.83	40	7.84	42	8.24	42	8.22	0.99
Immunology									
RF (positive)	249	57.37	274	60.62	303	65.16	345	74.84	< 0.0001
ACPA (positive)	153	38.83	201	48.32	256	60.52	282	64.98	< 0.0001
ESR (mm/hr)	31.38	26.18	32.56	29.24	33.61	27.82	35.93	29.05	0.07
CRP (mg/L)	0.77	1.80	0.89	1.90	0.79	2.21	0.97	2.21	0.39
DAS28 (first)	3.47	1.49	3.58	1.39	3.41	1.33	3.60	1.52	0.38
DAS28 (highest)	4.29	1.44	4.27	1.44	4.12	1.35	4.42	1.47	0.10
Comorbidities/outcome									
Periodontitis	36	7.05	32	6.27	37	7.25	30	5.87	0.79
Bone erosion	158	77.07	193	78.78	227	85.34	255	86.44	0.01
Drug									
Methotrexate	280	54.79	320	62.75	353	69.22	390	76.32	< 0.0001
Hydroxychloroquine	412	80.63	432	84.71	444	87.06	452	88.45	0.003
Sulfasalazine	151	29.55	195	38.24	188	36.86	223	43.64	< 0.0001
Cyclosporin	98	19.18	96	18.82	118	23.14	157	30.72	< 0.0001
Leflunomide	107	20.94	116	22.75	131	25.69	154	30.14	0.004
Glucocorticoids									
No	109	21.33	89	17.45	78	15.29	65	12.72	0.002
Yes	402	78.67	421	82.55	432	84.71	446	87.28	
Target drug									
Treatment with bDMARDs or tsDMARDs									< 0.0001
No	376	73.58	348	68.24	316	61.96	295	57.73	
Yes	135	26.42	162	31.76	194	38.04	216	42.27	
Number of bDMARDs or tsDMARDs treatment									< 0.0001
0	376	73.58	348	68.24	316	61.96	295	57.73	
1	89	17.42	98	19.22	119	23.33	131	25.64	
≥ 2	46	9	64	12.55	75	14.71	85	16.63	
TNFi	88	17.22	102	20	130	25.49	153	29.94	< 0.0001
Etanercept	35	6.85	39	7.65	59	11.57	67	13.11	0.001
Adalimumab	40	7.83	37	7.25	53	10.39	63	12.33	0.02
Golimumab	16	3.13	21	4.12	16	3.14	29	5.68	0.13
Certolizumab	8	1.57	13	2.55	9	1.76	13	2.54	0.58
Infliximab	0	0	0	0	0	0	0	0	-
Non-TNFi	56	10.96	81	15.88	92	18.04	105	20.55	0.0003
Tocilizumab	23	4.5	43	8.43	41	8.04	41	8.02	0.05
Abatacept	20	3.91	23	4.51	24	4.71	38	7.44	0.05
Rituximab	21	4.11	28	5.49	40	7.84	39	7.63	0.04
JAKi	31	6.07	41	8.04	48	9.41	41	8.02	0.26
Tofacitinib	23	4.5	34	6.67	35	6.86	31	6.07	0.38
Baricitinib	11	2.15	8	1.57	18	3.53	11	2.15	0.20
Continued									

RA PRS (N=2042)									
1 st quartile (n= 511)		2nd quartile (n= 510)		3rd quartile (n= 510)		top quartile (n= 511)		p value	
N	%	N	%	N	%	N	%		
TNFi: tumor necrosis factor inhibitors									
JAKi: Janus kinase inhibitors									
Note: “Age” and the values for “BMI,” “ESR,” “CRP,” and “DAS28” are presented as the mean with the corresponding standard deviation (SD). All other variables are represented as Number (N) and Percentage (%), which reflect the count and the proportion of patients within each categorical variable.									

Table 1. Basic characteristics of study populations.

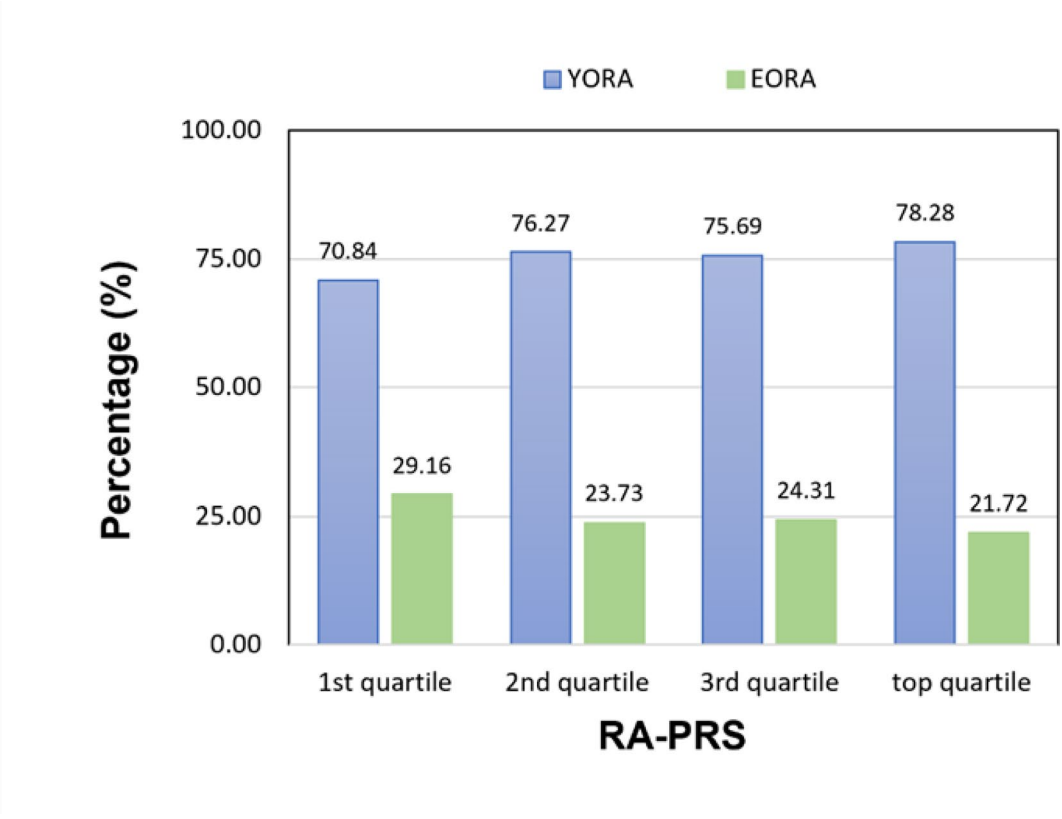


Fig. 2. Stratified Analysis by RA-onset age. Proportion of incidence in RA-PRS quartiles. The blue bar indicates the RA onset age of under 60 years (YORA); the green bar indicates the RA onset age of equal to or more than 60 years (EORA).

Discussion

In this study, we established a Taiwanese-specific PRS for predicting the development of RA. We discovered that RA-PRS was associated with RF/ACPA seropositivity, bone erosion, and the use of bDMARDs/tsDMARDs. Taken together, polygenic genetic information and clinical phenotypes contribute to the prediction model for structural damages. Our result shed light on the integration of PRS in the clinical assessment strategy to identify RA patients at high risk of structural damage.

Several PRS with different approaches for the prediction of RA development have been reported^{13–15}. Rostami et al. calculated the best-fitting risk score of 27 SNPs selected from relevant GWAS studies using *P*-value selection criteria $\leq 5 \times 10^{-8}$ in a Norwegian cohort of 489 RA cases and 61,584 controls¹⁴. Another study by Honda et al. used Asian GWAS data to select 43,784 SNP and calculate PRS¹⁵. They further trained and tested the performance of PRS in a Japanese cohort from the Institute of Rheumatology, Rheumatoid Arthritis (IORRA) cohort of 1,240 patients with RA¹⁵. In this study, we selected SNPs from GWAS data of 686,534 SNPs from 10,915 samples with a *P*-value threshold $< 1.11 \times 10^{-7}$. We compared the PRS performance in predicting RA susceptibility and selected the LDpred2 method, which had the best AUC. We believed that using different approaches for the PRS calculation of the GWAS cohort from distinct ethnicities could lead to variation in prediction performance. A previous report indicated that PRS based on multi-ancestry GWAS may outperform PRS derived from a single ethnicity¹³. We explored the use of the Ishigaki 2022 GWAS multi-ancestry dataset

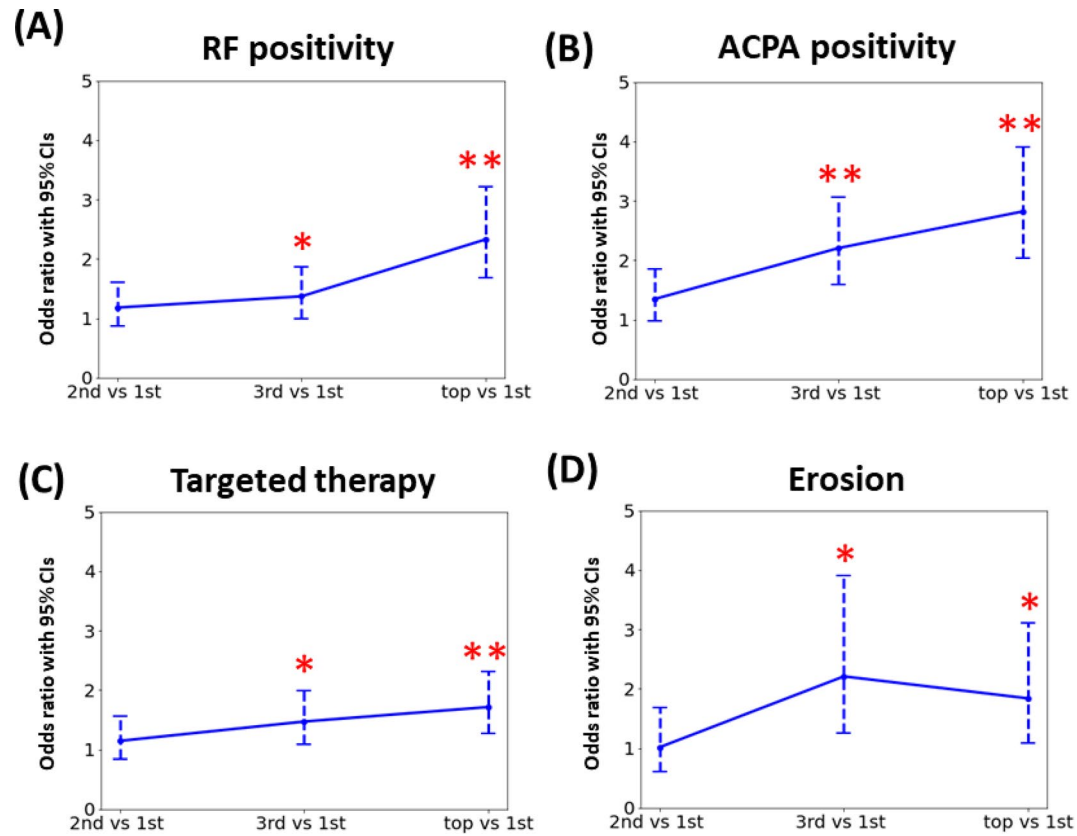


Fig. 3. PRS quartile plots of RF, ACPA, targeted therapy, and bone erosion positivity for YORA. According to the PRS and odds ratios, the association between the quartile of polygenic risk score and (A) RF positivity, (B) ACPA positivity, (C) Targeted therapy, and (D) erosion was generated using logistic regression analysis. The RA patients under 60 years RA onset were separated into quartiles. The top quartile in RA patients showed significantly higher risks for RF positivity, ACPA positivity, and targeted therapy compared with the first quartile. However, the second and third quartiles were significantly higher in the erosion group. Values are the odds ratios with 95% confidence intervals.

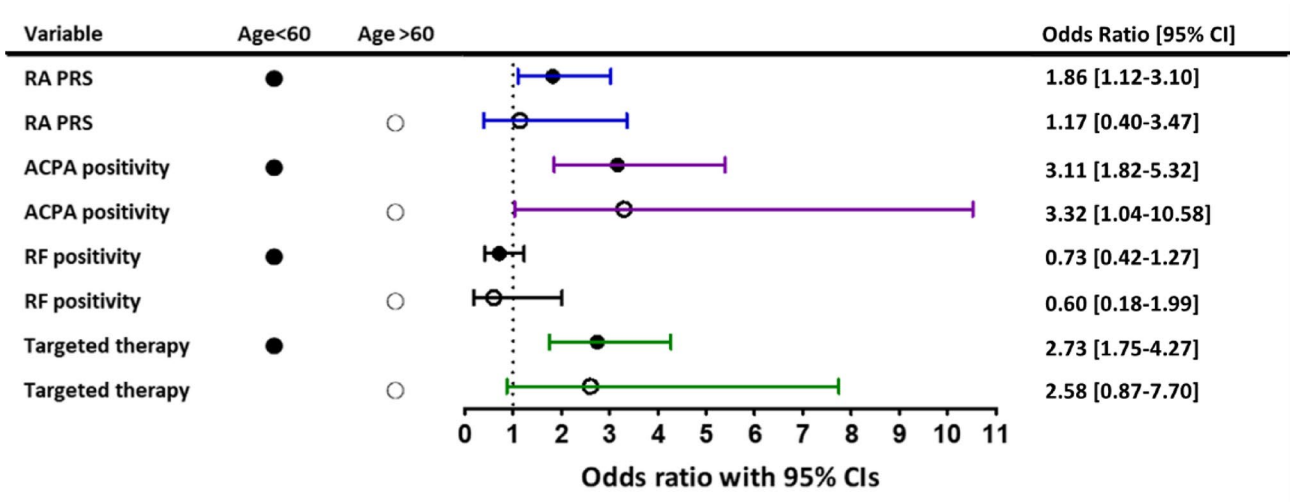


Fig. 4. Forest plot of RA-PRS, autoantibody positivity, and targeted therapy associated with bone erosion. Multiple logistic regression was used to analyze the odds ratios. The circles indicate odds ratio, and the line indicates the range of the 95% CI. ○ indicates the RA onset age of over or equal to 60 years; ● indicate the RA onset age under 60 years; blue line indicates RA-PRS; purple line indicates ACPA positivity; black line indicates RF positivity; green line indicates experienced targeted therapy. Targeted therapy refers to treatment with bDMARDs or tsDMARDs.

to construct our PRS. The results show that we need more SNPs based on the threshold of 0.00270005, which included 2388 SNPs. The PRS R^2 was 0.0437746. Additionally, we did not observe a significant enhancement in predictive ability when estimating the AUC. Protein heritability could vary substantially across a population with diverse genetic backgrounds²⁶. The present study is the first to use GWAS data and a training and testing cohort from a Taiwanese hospital-based population that is of Chinese Han ancestry. Future studies are needed to confirm whether training and testing cohort differences in ancestry might contribute to PRS performance.

In our comparison with previous GWAS studies, we meticulously analyzed our results alongside existing GWAS data, obtained from the GWAS catalog as of January 15, 2024. In this analysis, 201 SNPs in our GWAS were found to have been reported as significantly associated with RA in other studies. Notably, among these, 3 SNPs (rs1571878, rs2301888, and rs9268839) were also significant in our study ($P < 1 \times 10^{-4}$), indicating that these SNPs are significant markers related to RA across different studies and populations. Furthermore, we discovered 787 previously unreported SNPs in our GWAS, of which 73 SNPs shared LD with the reported rs9268839, with an $R^2 > 0.85$. Nonetheless, these findings still indicate that the SNPs discovered in this study suggest potential ancestry-specific genetic variations. These findings may offer a deeper understanding of the genetic predisposition to RA in this specific ancestry group.

GWAS is a hypothesis-free research approach that enables the exploration of genetic variants associated with specific phenotypes across the entire genome. The advantage of this method lies in its ability to uncover previously unknown pathogenic mechanisms without being limited by existing functional annotations, thereby providing novel insights into disease mechanisms. SNPs identified through GWAS are often located in non-coding regions, particularly within introns, which have crucial regulatory functions that can influence gene expression, splicing regulation, and even the production of non-coding RNAs. We further analyzed the SNPs used for PRS and found that these variants could potentially affect up to 822 genes. However, only 12 SNPs were found to result in coding sequence variants. A subsequent analysis of these affected genes using DAVID revealed that the primary functional clusters were closely related to MHC immune regulatory functions, highlighting the potentially pivotal role of immune mechanisms in these genetic variations.

Additionally, we considered the effect of the shared epitope (SE) and found that only one SNP, rs14391466, located in *DRB1*, was involved. Removing SE did not affect the predictive ability of the PRS.

Our study was the first to indicate that the patients in the top quartile of RA-PRS exhibited a higher probability of RF and ACPA seropositivity, particularly among RA patients aged under than 60 years (YORA group). This suggests that genetic risk scores may be more predictive in younger-onset RA cases. A Japanese study similarly reported that ACPA-positive RA patients in the top PRS quintile had significantly higher rates of radiographic progression¹⁵. A Dutch cohort of patients with clinically suspect arthralgia, a pre-RA stage, demonstrated that PRS was associated with early RA in the ACPA-positive population, but not in ACPA-negative group¹². Interestingly, a significant interaction between occupational inhalation, smoking, and genetic background in terms of PRS was observed only in ACPA-positive RA patients from a Swedish cohort²⁷. This suggests that genetic susceptibility may manifest differently depending on environmental exposures and disease onset timing. Overall, RA-PRS could provide an opportunity to evaluate the complex interactions between heritability, autoimmunity, and environmental exposure in the future.

We found that RA-PRS was associated with bone erosion, particularly in RA patients aged less than 60 years (YORA group). This was consistent with the Japanese study by Honda et al. that reported that quartiles of PRS were associated with severe radiographic progression in patients with a younger onset age of RA (<40 years)¹⁵. However, we showed that PRS quartiles were correlated with the use of targeted therapy in our cohort. According to Taiwan's National Health Insurance reimbursement criteria, patients with RA were only eligible for treatment with either bDMARDs or tsDMARDs if the disease activity was high despite a combination therapy of three csDMARDs. Although our analysis did not demonstrate the significant correlation between PRS and disease activity score by 28 joints (DAS28), we observed that the PRS was increased and speculated that RA patients with high PRS might exhibit a more severe disease activity. This also explains why our results indicated that treatment with bDMARDs or tsDMARDs was associated with bone erosion. Further study is needed to confirm our findings.

One of the strengths of our study was that it incorporated a large sample size of 2042 RA cases and 7950 non-RA controls. We also compared the prediction power of RA susceptibility using three PRS algorithms of LDpred2, plink, and PRSice2. However, some limitations did exist in this study. Firstly, our cohort was enrolled in a single medical institution and the PRS algorithm was not validated in an independent cohort. Furthermore, our study population was all Taiwanese. Our data might not be applicable for extrapolation to RA patients with other ancestry. Secondly, the study design was retrospective and cross-sectional. Therefore, clinical parameters were not prospectively collected, and missing data might exist in our analysis. Lastly, shared epitope (SE) HLA alleles were not included in the analysis, and to determine whether PRS might have an interaction with SE was beyond the scope of the current study.

Conclusions

PRS is associated with autoantibody seropositivity and bone erosion in Taiwanese patients with RA, especially in RA patients aged less than 60 years (YORA group). Future studies might incorporate PRS to identify the groups that are at high-risk of radiographic progression and guide the early initiation of targeted therapy in genetically vulnerable patients with RA.

Data availability

All data used in this study in this article are available through the corresponding author.

Received: 30 October 2024; Accepted: 19 June 2025

Published online: 16 July 2025

References

- Smolen, J. S. et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. *Ann. Rheum. Dis.* **82**, 3–18 (2023).
- Aletaha, D. & Smolen, J. S. Diagnosis and management of rheumatoid arthritis: A review. *JAMA* **320**, 1360–1372 (2018).
- Lin, C. T. et al. Predictors of drug survival for biologic and targeted synthetic DMARDs in rheumatoid arthritis: analysis from the TRA clinical electronic registry. *PLoS One*. **16**, e0250877 (2021).
- Syversen, S. W. et al. High anti-cyclic citrullinated peptide levels and an algorithm of four variables predict radiographic progression in patients with rheumatoid arthritis: results from a 10-year longitudinal study. *Ann. Rheum. Dis.* **67**, 212–217 (2008).
- Chen, Y. M. et al. Association between autophagy and inflammation in patients with rheumatoid arthritis receiving biologic therapy. *Arthritis Res. Ther.* **20**, 268 (2018).
- Traylor, M. et al. Genetic associations with radiological damage in rheumatoid arthritis: Meta-analysis of seven genome-wide association studies of 2,775 cases. *PLoS One*. **14**, e0223246 (2019).
- Frisell, T. et al. Familial risks and heritability of rheumatoid arthritis: role of rheumatoid factor/anti-citrullinated protein antibody status, number and type of affected relatives, sex, and age. *Arthritis Rheum.* **65**, 2773–2782 (2013).
- Okada, Y. et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat. Genet.* **44**, 511–516 (2012).
- Chung, C. W. et al. Machine learning approaches for the genomic prediction of rheumatoid arthritis and systemic lupus erythematosus. *BioData Min.* **14**, 52 (2021).
- Suzuki, T. et al. PADI4 and HLA-DRB1 are genetic risks for radiographic progression in RA patients, independent of ACPA status: results from the IORRA cohort study. *PLoS One*. **8**, e61045 (2013).
- Choi, S. W., Mak, T. S. & O'Reilly, P. F. Tutorial: a guide to performing polygenic risk score analyses. *Nat. Protoc.* **15**, 2759–2772 (2020).
- Maurits, M. P. et al. The role of genetics in clinically suspect arthralgia and rheumatoid arthritis development: A large Cross-Sectional study. *Arthritis Rheumatol.* **75**, 178–186 (2023).
- Ishigaki, K. et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. *Nat. Genet.* **54**, 1640–1651 (2022).
- Rostami, S., Hoff, M., Brown, M. A., Hveem, K. & Videm, V. Comparison of methods to construct a genetic risk score for prediction of rheumatoid arthritis in the population-based Nord-Trøndelag health study. *Nor. Rheumatol. (Oxford)*. **59**, 1743–1751 (2020).
- Honda, S. et al. Association of polygenic risk scores with radiographic progression in patients with rheumatoid arthritis. *Arthritis Rheumatol.* **74**, 791–800 (2022).
- Aletaha, D. et al. 2010 rheumatoid arthritis classification criteria: an American college of rheumatology/european league against rheumatism collaborative initiative. *Arthritis Rheum.* **62**, 2569–2581 (2010).
- Stahl, E. A. et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508–514 (2010).
- Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
- Lam, M. et al. RICOPIII: rapid imputation for consortia pipeline. *Bioinformatics* **36**, 930–933 (2020).
- Ni, G. et al. A comparison of ten polygenic score methods for psychiatric disorders applied across multiple cohorts. *Biol. Psychiatry*. **90**, 611–620 (2021).
- Roberts, G. H. L., Fain, P. R., Santorico, S. A. & Spritz, R. A. Inverse relationship between polygenic risk burden and age of onset of autoimmune vitiligo. *Am. J. Hum. Genet.* **111**, (2024).
- Aranda-Guillén, M. et al. A polygenic risk score to help discriminate primary adrenal insufficiency of different etiologies. *J. Intern. Med.* **294**, 96–109 (2023).
- Deal, C. L. et al. The clinical features of elderly-onset rheumatoid arthritis. A comparison with younger-onset disease of similar duration. *Arthritis Rheum.* **28**, 987–994 (1985).
- Sharp, J. T. et al. How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis? *Arthritis Rheum.* **28**, 1326–1335 (1985).
- Wakefield, R. J. et al. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. *J. Rheumatol.* **32**, 2485–2487 (2005).
- Thareja, G. et al. Differences and commonalities in the genetic architecture of protein quantitative trait loci in European and Arab populations. *Hum. Mol. Genet.* **00**, 1–10 (2022).
- Tang, B. et al. Occupational inhalable agents constitute major risk factors for rheumatoid arthritis, particularly in the context of genetic predisposition and smoking. *Ann. Rheum. Dis.* **0**, 1–8 (2022).

Acknowledgements

The authors would like to thank the Data Science Statistical Cooperation Center of Academia Sinica (AS-CFII-111-215) for statistical support. We thank all the participants and investigators from Taiwan Precision Medicine Initiative.

Author contributions

T.S. Wu and Y.M. Chen conceptualized the study. Y.J. Chen, C.N. Hsiung, C.Y. Wei, I.C. Chen, and T.H. Hsiao were responsible for data curation, while C.N. Hsiung and C.Y. Wei performed the formal analysis. Y.M. Chen secured funding for the research. W.N. Huang, Y.H. Chen, and Y.M. Chen were responsible for patient recruitment and clinical assessment. Supervision was provided by T.H. Hsiao, C.M. Kao, W.N. Huang, Y.H. Chen, and Y.M. Chen. The original draft was written by T.S. Wu and Y.M. Chen, with all other authors contributing to the review and editing of the manuscript.

Funding

This study was funded by Academia Sinica 40-05-GMM and AS-GC-110-MD02, National Science and Technology Council, Taiwan [NSTC – 111-2634-F-A49-014, NSTC-111-2218-E-039-001, and NSTC-111-2314-B-075 A-003-MY3], and Taichung Veterans General Hospital, Taiwan [TCVGH-1127301 C, TCVGH-1127302D, and TCVGH-YM1120110].

Declarations

Ethics approval and consent to participate

Participants from all groups gave written informed consent following approved protocols by their institutional ethics committees. This study was approved by the institutional review board of Taichung Veterans General Hospital (No. SF19153A).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Running title

RA-PRS, seropositivity, bone erosion.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-08254-1>.

Correspondence and requests for materials should be addressed to Y.-M.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025