

Research Article

Ekrem Akbulut*, Tülay Yıldırım and Onur Öztürk

Endoplasmic reticulum aminopeptidase-1 polymorphism increases the risk of rheumatoid arthritis

Endoplazmik retikulum aminopeptidaz-1 polimorfizmi romatoid artrit riskini artırır



<https://doi.org/10.1515/tjb-2022-0007>

Received January 6, 2022; accepted February 17, 2022;

published online March 24, 2022

Abstract

Objectives: Endoplasmic reticulum aminopeptidase-1 (*ERAP1*) polymorphic changes cause autoimmunity. To understand the contribution of *ERAP1* to the occurrence of rheumatoid arthritis (RA) disease, we investigated the relationship between *ERAP1* and RA.

Methods: This study was conducted with 201 patients and 171 healthy controls. The rs26653, rs27044, rs27582, rs28096, and rs30187 polymorphic regions of *ERAP1* were investigated. The comparison was done with Arlequin software and logistic regression. Haplotypes were analyzed with Phylogenetic Network software. *ERAP1* was modeled using Promod3. Topological changes in *ERAP1* were analyzed with TM-Score.

Results: The results showed that rs26653G>C ($p=0.002$, $OR=2.001$, $95\%CI=1.276-3.137$), rs27044C>G ($p=0.037$, $OR=1.583$, $95\%CI=1.028-2.440$), rs27582G>A ($p<0.05$, $OR=0.348$, $95\%CI=0.194-0.622$) and rs30187C>T ($p=0.006$, $OR=1.849$, $95\%CI=1.191-2.870$) polymorphisms are associated with RA disease risk. The relationship between rs28096 polymorphism and RA disease risk could not be

determined ($p=0.509$). The risk haplotype for rheumatoid arthritis was determined as [CGAAT]. It was determined that polymorphisms of *ERAP1* cause changes in the entry pocket of substrate and ligand.

Conclusions: We report a haplotype [CGAAT] that is associated with RA risk from Turkey that has not been described before. These data will make important contributions to elucidating the molecular mechanism of RA.

Keywords: autoimmunity; *ERAP1*; genetic predisposition; peptide presentation; rheumatoid arthritis.

Öz

Amaç: Endoplazmik retikulum aminopeptidaz-1 (*ERAP1*) polimorfik değişiklikleri otoimmüniteye neden olur. *ERAP1*'in, romatoid artrit (RA) hastalığının oluşumuna katkısını anlamak için *ERAP1* ve RA arasındaki ilişkiyi araştırdık.

Gereç ve Yöntemler: Bu çalışma 201 hasta ve 171 sağlıklı kontrol ile yapılmıştır. *ERAP1*'in rs26653, rs27044, rs27582, rs28096 ve rs30187 polimorfik bölgeleri araştırıldı. Karşılaştırma Arlequin yazılımı ve lojistik regresyon ile yapıldı. Haplotipler, Phylogenetic Network yazılımı ile analiz edildi. *ERAP1*, Promod3 kullanılarak modellenmiştir. *ERAP1*'deki topolojik değişiklikler TM-Score ile analiz edildi.

Bulgular: Sonuçlar, rs26653G>C ($p=0.002$, $OR=2.001$, $95\%CI=1.276-3.137$), rs27044C>G ($p=0.037$, $OR=1.583$, $95\%CI=1.028-2.440$), rs27582G>A ($p<0.05$, $OR=0.348$, $95\%CI=0.194-0.622$) and rs30187C>T ($p=0.006$, $OR=1.849$, $95\%CI=1.191-2.870$) polimorfizminin RA hastalık riski ile ilişkili olduğunu gösterdi. rs28096 polimorfizmi ile RA hastalığı riski arasındaki ilişki belirlenemedi ($p=0.509$). Romatoid artrit için risk haplotipi [CGAAT] olarak belirlendi. *ERAP1* polimorfizmlerinin substrat ve ligandin giriş cebinde değişikliklere neden olduğu belirlendi.

*Corresponding author: Ekrem Akbulut, Bioengineering, Faculty of Engineering and Natural Sciences, Malatya Turgut Özal University, Malatya, Turkey, E-mail: ekremakbulut@gmail.com. <https://orcid.org/0000-0002-7526-9835>

Tülay Yıldırım, Physical Medicine and Rehabilitation, İnönü University, Malatya, Turkey. <https://orcid.org/0000-0002-3443-8102>

Onur Öztürk, Biophysics, Faculty of Medicine, Malatya Turgut Özal University, Malatya, Turkey. <https://orcid.org/0000-0003-3404-4135>

Sonuç: Türkiye’den RA riski ile ilişkili daha önce tanımlanmamış bir haplotip [CGAAT] bildiriyoruz. Bu veriler, RA’nın moleküler mekanizmasının aydınlatılmasına önemli katkılar sağlayacaktır.

Anahtar Kelimeler: otoimmünite; *ERAP1*; genetik yatkınlık; peptid sunumu; romatoid artrit.

Introduction

Rheumatoid arthritis (RA) is a systemic inflammatory disorder that affects between 0.5 and 1% of the population and ultimately induces persistent synovial inflammation, resulting in joint destruction and disablement [1]. The disease affects women more frequently than men 2 to 3 times and appears at every age. The highest incidence emerged during the sixth decade [2]. A highly specific RA biomarker linked to disease seriousness, the plurality of infected RA cases (about 70 percent) are seropositive for anti-citrullinated protein antibodies (ACPA) [3]. An increased risk for RA has been associated with multiple genetic and environmental factors. The highest connections have been seen with female gender, a familial history of RA, genetic factor, and tobacco smoke exposure [4]. RA inheritance is estimated to be approximately 60% [5]. Genetic factors including class II major histocompatibility antigens and non-human leukocyte antigen (HLA) genes have been implicated in the pathogenesis of RA [6, 7]. While the pathogenesis of RA is complicated and remains uncertain, it is necessary to consider and acknowledge its pathogenesis in disclosing the effective therapeutic strategy that could contribute to meaningful clinical benefits [8]. Early detection is critical to proper clinical progress. Especially in patients with proven risk factors, including high disease activity and antibodies, or early joint injury [9].

Endoplasmic reticulum aminopeptidase-1 (*ERAP1*) is a zinc metallopeptidases that are active in trimming peptides in the endoplasmic reticulum (ER) to load them into the HLA class I antigen-binding groove [10]. *ERAP1* cleavages proinflammatory cytokine receptors, such as interleukin 1 receptor type 2 (*IL1R2*), interleukin 6 receptor subunit alpha (*IL6RA*), and tumor necrosis factor receptor 1 (*TNFR1*), leading to a decrease in cell transmission of inflammatory signaling [11]. *ERAP1* activity affects cell surface immunopeptidome and epitope immunodominance models, although its molecular mechanism is not fully elucidated. Studies have shown that *ERAP1* is an essential determinant of the immune response by producing mature antigenic epitopes from peptides that are extracted from degraded proteins [12]. Resolving the particulars of how *ERAP1* generates antigenic peptides would have significant

implications about how medications may be used to control the immunopeptidome and how *ERAP1* biology causes a predisposition to autoimmune disease [13].

Single nucleotide polymorphism (SNP) of *ERAP1* has been shown to be associated with cancer, rheumatic and viral diseases [14]. A genome-wide association study showed that polymorphic changes in *ERAP1* can result in genetic predisposition to the disease and changes in the immune response. Antigen presentation is influenced by SNPs which trigger changes to enzyme activity [14, 15].

In this study, the relationship between *ERAP1* polymorphic changes NM_016442.5:p.Arg127Pro (rs26653), NM_016442.5:p.Gln730Glu (rs27044), NC_000005.10:g.96762509G>A (rs27582), NC_000005.10:g.96773539G>A (rs28096) and NM_016442.5:p.Lys527Arg (rs30187) and susceptibility to RA were investigated. In this context; the genetic relationships between individuals clinically and serologically diagnosed with RA and healthy controls were determined at allele, genotype, haplotype, and protein level.

Materials and methods

Patients and ethical compliance

This case-control study was conducted with 201 patients who were clinically and serologically diagnosed with RA and 171 healthy controls. European League Against Rheumatism (EULAR)/American College of Rheumatology (ACR) classification criteria were used in the diagnosis of RA. The individuals without any autoimmune and endocrine diseases were included in the control group. Table 1 depicts the clinical and demographic specifications of study subjects and their values. Ethical rules were followed in this study and written consent of volunteers was obtained. The study protocol was approved by the ethics committee of Malatya Clinical Investigations (2016/12). The study followed the ethical standards of the Helsinki Declaration.

SNP selection and genotyping

We first retrieved the reference SNPs (rs) defined for *ERAP1* from the NCBI database and selected potential functional SNPs based on the following criteria: (i) located in the exon and intron regions; (ii) not studied in the published genome-wide association studies; (iii) potential functional SNPs identified using SNP info software; (iv) SNPs reported to be associated with different autoimmune diseases were selected. Linkage disequilibrium (LD) was checked. Finally, NM_016442.5:p.Arg127Pro (rs26653), NM_016442.5:p.Gln730Glu (rs27044), NC_000005.10:g.96762509G>A (rs27582), NC_000005.10:g.96773539G>A (rs28096) and NM_016442.5:p.Lys527Arg (rs30187) polymorphic regions were included in the study. DNA isolation from blood samples was performed using Invisorb Spin Blood Mini Kit™ (Catalog Number: 1031100300). Polymorphic regions were amplified by multiplex Polymerase Chain Reaction (PCR) using designed

Table 1: Clinical and demographic data.

Demographic data	Case (n=201)	Control (n=171)	p-Value Odds ratio (95%CI)
Anti-CCP (+) (Mean ± SD, range)	58.71% 140.78 ± 135.48 3–300	–	–
RF (+) (Mean ± SD, range)	64.18% 103.84 ± 152.16 0.994–1.040	–	–
ESR (+) (Mean ± SD, range)	50.75% 22.23 ± 13.96 2–71	–	–
CRP (+) (Mean ± SD, range)	49.75% 0.84 ± 1.05 0.303–4.84	–	–
Age (mean ± SD) (Age range)	54.99 ± 12.64 20–80	47.14 ± 11.26 23–71	–
Gender (female)	69.7% 140	49.12% 84	0.000 2.61 (1.595–4.253)
Disease duration (Year ± SD)	11.45 ± 10.07	–	–
Diagnosis delay (Year ± SD)	2.4 ± 5.4	–	–
Familial RA disease story	22.39%	–	0.030 2.26 (1.082–4.725)
Surgical intervention before symptoms occur	58.71%	–	0.000 6.22 (3.730–10.367)
Suffered severe infectious disease	13.43%	–	0.009 3.337 (1.346–8.275)
Presence of diabetes	16.42%	–	0.020 2.62 (1.165–5.887)
Presence of high blood pressure	20.9%	–	0.885 1.048 (0.552–1.991)
Mechanical trauma exposure (traffic accident, falling etc.)	12.44%	–	0.070 0.51 (0.250–1.055)
Smoking	19.9%	–	0.236 1.44 (0.788–2.630)

Significant $p < 0.05$; SD-standard deviation.

primer pairs. The genotype analysis was performed according to Gabriel et al. [16]. Primer pairs and probe information used in the single-base elongation reaction and multiplex PCR are presented in Supplemental Table 1.

Sequence data and modeling of polymorphic protein

Reference nucleotide and protein sequence information of *ERAP1* were obtained from the NCBI (NM_016442.5) and Uniprot portal (Q9NZ08), respectively. The polymorphic residues information of *ERAP1* was processed with MEGAX software. Models were built based on the target-template alignment using ProMod3 (ver3.1.1). 6rxq (RCSB protein data bank code) was selected as template. ProSA and MolProbity tools were used for structural validation of *ERAP1* polymorphic model [17, 18]. Conformational analyzes of polymorphic *ERAP1* were performed with PyMOL (ver2.4.1). Topological differences of wild and polymorphic *ERAP1* were calculated with the i-Tasser TM-Score and root mean square deviation (RMSD) algorithm [19].

Statistical analysis

The association of SNP alleles and genotypes with disease predisposition was analyzed by a logistic regression test (SPSS 24.0, Chicago). Binary Logistic Regression model was established in which polymorphic changes, smoking, family history, diabetes, blood pressure, surgery, mechanical trauma, infection history, and gender variables represent covariates if the group represents the patient and control variable is dependent. Associations of RA with polymorphisms are shown as odds ratios (ORs), and risks were estimated with 95% confidence intervals (CIs). Power analysis was performed by using G*Power 3.1. Demographic data were given as mean ± standard deviation and min/max values. “p” value smaller than 0.05 was regarded as statistically significant. The associated haplotype varieties for healthy control samples and RA population were calculated with Arlequin 3.5 software [20]. LD analysis was performed using Haploview v. 4.2 program [21]. Hardy–Weinberg equilibrium (HWE) test was performed at each locus for HWE analysis in populations. Arlequin software was used to calculate p values in these analyzes. Arlequin

software uses Fisher's exact test to calculate the significance of the deviation from the HWE. The probably historical movement model of RA was estimated by neutrality tests and mismatch distribution analysis method. For this purpose, tau (τ) and theta initial (θ_0) were estimated assuming theta final (θ_1) to calculate the possible mode of mismatch distribution using Arlequin software. The parameters of the infinity and the raggedness index and the effects of the Tajima and Fu neutrality tests were determined using Arlequin.

Results

The clinical and demographic data for the 201 patients and 171 controls evaluated in this study are shown in Table 1. The percentages of the patients for anti-CCP (+), rheumatoid factor-RF (+), erythrocyte sedimentation rate-ESR (+), C-reactive protein-CRP (+) were 58.71%, 64.18%, 50.75%, 49.75%, respectively. It was found that the diagnosis of the disease was delayed for 2.4 years. Gender was associated with the risk of disease, and the incidence of the disease in women was 2.61 times higher than men ($p < 0.05$). There was a familial RA disease story of 22.39% of the patients ($p = 0.030$). Before the onset of disease complaints, surgical intervention ($p < 0.05$), severe infectious disease ($p = 0.009$) and presence of diabetes ($p = 0.020$) may be associated with the risk of disease (Table 1). Hosmer–Lemeshow statistics was used to test the goodness of fit in the model established to predict the group variable of polymorphic, clinical and demographic variables, and it was found to be statistically sufficient ($\chi^2 = 6.362$, $SD = 8$, $p = 0.607 > 0.05$).

Genotyping results showed that rs26653 ($p = 0.002$), rs27044 ($p = 0.037$), rs27582 ($p < 0.05$) and rs30187 ($p = 0.006$) polymorphisms are associated with RA disease risk (Table 2). The individuals carrying the C allele in the rs26653 polymorphic region had a 2-fold higher risk of RA disease (OR: 2.001). It was determined that the G allele

for rs27044 increased the disease risk 1.6 fold. The C>G polymorphic change at this position results in Gln730 Glu amino acid change in the protein structure. Similarly, T allele for rs30187 increased the disease risk 1.9 fold. Similarly, it was shown that rs30187C>T polymorphic change causing Lys528Arg amino acid change increased predisposition to RA 1.9 fold. The relationship between rs28096 polymorphism and RA disease risk could not be determined ($p = 0.509$).

Table 3 summarizes the frequencies for rs26653, rs27044, rs27582, rs28096, and rs30187 haplotypes determined for five unrelated heterozygous loci in RA and healthy control populations respectively. The RA population observed in the Malatya region is linked with the haplotype [GCGAT] (15.9% and No.1). In the healthy control population the haplotype [GCGAT] is in the high frequency (14.3% and No.1). There is a risk haplotype [CGAAT] in the RA population. However, in healthy control population the haplotype [CGAAT] does not have any frequency value.

Populations diversity parameters (h , π , k) are shown at Supplemental Table 2. The h parameter is highly calculated and seems to be compatible with the control population. (0.91 ± 0.005) and the RA population (0.92 ± 0.004). Nucleotide diversity (π) has low and similar values; 2.15 ± 1.32 (healthy control population) and 2.37 ± 1.43 (RA population). Fu's F_s test results show that both populations have common growth rates (Supplemental Table 2). Tajima's D results ($p > 0.05$) indicate that there is neutral equilibrium among the populations. These compatible data show that unlikely haplotype diversity is not possible in future generations. On the other hand, Tajima's D values indicate that heterozygotes have a selective advantage in these populations. Estimations of time τ and the current and historical population parameters θ imply a similar historical episode of expansion for the two populations ($\tau > 0$ and $\theta_1 > \theta_0$; Supplemental Table 2). The

Table 2: *ERAP1* genotyping results in RA.

SNP	Alleles		Case (n=201)				Control (n=171)				p-Value	Or (95%CI)
	1	2	Genotypes				Genotypes					
			11	12	22	MAF	11	12	22	MAF		
rs26653 R127P	C	G	37	105	59	0.45	12	73	86	0.28	0.002*	2.001 (1.276–3.137)
rs27044 Q730E	C	G	83	99	19	0.34	96	67	8	0.24	0.037*	1.583 (1.028–2.440)
rs27582	G	A	81	101	19	0.39	59	94	18	0.40	0.000*	0.348 (0.194–0.622)
rs28096	A	G	19	119	63	0.39	16	97	58	0.38	0.509	1.167 (0.738–1.847)
rs30187 K528R	T	C	49	88	64	0.46	23	65	83	0.33	0.006*	1.849 (1.191–2.870)

Significant $p < 0.05$; OR-odds ratio; bold alleles-minor alleles; bold p-value-associated with disease risk.

Table 3: rs26653, rs27044, rs27582, rs28096 and rs30187 haplotypes for the five loci in association with RA and healthy control population.

No	Haplotype		Frequency		SD	
	Case	Control	Case	Control	Case	Control
1	GCGAT	GCGAT	0.159204	0.14327	0.018272	0.01897
2	CGAGC	GCGAC	0.111944	0.13158	0.015745	0.01831
3	GCGAC	GCGGC	0.084577	0.11988	0.013895	0.01759
4	GCGGT	CCAGC	0.082099	0.09064	0.013708	0.01555
5	CCAGC	CGAGC	0.077114	0.08772	0.013322	0.01532
6	GCGGC	GCGGT	0.069652	0.07895	0.012712	0.0146
7	CGAGT	GCAGC	0.044776	0.06725	0.010328	0.01356
8	CCGGC	GGGGC	0.039801	0.06725	0.009762	0.01356
9	GGGGC	GCAAC	0.037313	0.02632	0.009465	0.00867
10	CCGGC	CCAGT	0.034826	0.02047	0.009155	0.00767
11	CCGAT	CGAGT	0.023338	0.01754	0.008834	0.00711
12	GGGGT	GGAGC	0.029851	0.01754	0.008498	0.00711
13	CCAGT	GCAGT	0.024876	0.01462	0.007778	0.0065
14	CCGGT	CCGGC	0.019999	0.01462	0.006974	0.0065
15	CCGAC	GCAAT	0.019999	0.01462	0.006974	0.0065
16	GCAAT	CCGAC	0.019999	0.01462	0.006974	0.0065
17	GGGAC	GGGAC	0.017413	0.01175	0.006532	0.00582
18	GGAGT	GGGGT	0.014925	0.00877	0.006055	0.00505
19	CGAAC	CCGAT	0.014925	0.00877	0.006055	0.00505
20	GGGAT	CGGGC	0.014925	0.00585	0.006055	0.00413
21	GCAGC	CCGGT	0.007463	0.00585	0.004298	0.00413
22	CGGGT	GGAGT	0.007463	0.00585	0.004298	0.00413
23	GCAAC	GGAAC	0.007463	0.00585	0.004298	0.00413
24	CCAAC	CCAAC	0.007463	0.00585	0.004298	0.00413
25	GCAGT	CGAAC	0.004975	0.00585	0.003514	0.00413
26	CCAAT	CCGAT	0.002488	0.00292	0.002488	0.00292
27	GGAGC	GGGAT	0.002488	0.00292	0.002488	0.00292
28	GGAAC	CGGAC	0.002488	0.00292	0.002488	0.00292
29	CGGAC		0.002488		0.002488	
30	CGAAT		0.002488		0.002488	
31	CGAGT		0.002488		0.002488	

SD-Standard deviation.

demographic histories of the two populations were examined using a mismatch distribution analysis (Supplemental Table 3). The null hypothesis of the Neutrality test expresses the population balance. The opposite hypothesis describes the population growth according to the results of SSD and rg tests. The distribution of pairwise differences in control and RA is significantly different ($p < 0.05$) from the expected distribution based on both SSD and rg analysis results performed with the sudden population expansion model (Supplemental Table 3). According to the SSD and rg tests, the null hypothesis is not accepted and the distribution is defined as unimodal. In our results, distribution parameters, the control and RA population is departure from the unimodal distribution (Supplemental Table 3). The reason for this difference in distribution is that the control and RA

populations presented in Supplemental Table 4 are departure from HWE of the 3., 4. and 3. locus, respectively [22, 23].

We performed the phylogenetic network analysis of the haplotypes linked with the RA population by calculating the frequency and possible link estimations with NETWORK software (Supplemental Figure 1). We tested the LD using Haploview software. LD analysis showed weak allelic combinations in the rs27582 and rs26653 polymorphic regions in the RA and control populations (Figure 3).

Missense changes caused by polymorphic substitutions in tertiary structure were modeled. The molprobit score of polymorphic *ERAP1* was 1.25. The Z-score of the polymorphic model was -13.17 and the model was found in the native proteins range that were identified by X-ray crystallography. It was determined that polymorphic *ERAP1* has topological and conformational differences compared to reference *ERAP1* (RMSD=0.13 Å and TM-score=0.97). The data obtained from wild and polymorphic model comparisons indicate topological and conformational changes.

Discussion

ERAP1 modulates immune responses by regulating the peptide repertory eligible for presenting by MHC molecules [24]. In doing so, *ERAP1* clips peptides by a unique molecular-scale mechanism that recognizes both the N- and C-ends of the peptides to read the length of the substrate [25]. The polymorphic changes of *ERAP1* cause HLA-dependent inflammatory autoimmunity [15]. Some polymorphic modifications, including the Lys528Arg and Gln730Glu, are located close substrate binding and regulatory sites and may influence the processing and clipping specificity of peptides [26].

The rs30187 polymorphism was associated with RA disease risk ($p=0.006$, OR=1.849, 95%CI=1.191–2.870). Until today, the relationship of rs30187 polymorphism with susceptibility to different diseases other than RA has been shown in many studies. In the *ERAP1*-ankylosing spondylitis (AS) association study by Harvey et al. with 730 patients and 1,021 controls, the rs30187 polymorphism was shown to be strongly associated with disease risk ($p=0.0034$) [27]. While Gureini et al. showed the genetic predisposition of rs30187 to multiple sclerosis, Giancchetti et al. showed the contribution of rs30187 to the risk of type 1 diabetes mellitus [28, 29].

One of the important processes in *ERAP1* catalytic activity is to increase the efficiency of the catalytic activity by closing domain 4, which consists of alpha helices, on the active site. In the transition between the closed-open formation, Gly529 and its surrounding residues act as a hinge, moving it 28° away from the protease area of domain 4 and performing 12° of lateral rotation [30, 31]. The transformation

of the open-closed formation transition, which has an important role in the immune response, after the mutation in the Lys528 position from Lys containing 2 nitrogen atoms to Arg containing 4 nitrogen atoms results in an increase in protonation capacity. In this case, two alternative scenarios will result in a change in catalytic activity. Firstly, a decrease in catalytic activity can be seen as the full transition from open to closed formation cannot be achieved and the activity distance (29 Å) between the N-terminal catalytic region and the C-terminal regulatory region will change. Secondly, when the complete transition from closed to open formation cannot be achieved, sustained catalytic activity and antigenic peptide presentation may result in unwanted changes and products in the immune response.

It has been shown that *ERAP1* Lys528Arg polymorphism will cause a significant change in peptide processing properties as a result of disruption in interdomain interactions. *ERAP1* is a multifunctional enzyme involved in the regulation of immune and inflammatory responses [32].

The Gln730 is one of the polymorphic residues of *ERAP1*, which has been shown to be associated with disease susceptibility and antigen presentation [26]. Gln730 is located in the interior of the catalytic cavity of *ERAP1* and its interaction with catalytic residues makes it important in *ERAP1* efficacy. The polymorphic change of Gln730Glu will change the surface electrostatic potential of the inner cavity, affecting substrate specificity and/or the conformational dynamics of the enzyme. Sui et al. conducted a study with the -IINFEKL-substrate to understand the roles of the C-terminal regulator region of *ERAP1* in catalytic activity. *ERAP1* acts by binding to the substrate placed in the catalytic cavity, the N terminal catalytic region and the C terminal regulator region. In this mechanism, Gln730 located at the midpoints of the substrate interacts with the asparagine (N) of the peptide by hydrogen bonding [25]. On the other hand, the Gln730Glu polymorphic residue is located positionally next to the C-terminal specificity pocket (Figure 1). Electrostatic charge change beside the C-terminal specificity pocket, which has an important role in peptide specificity, may affect the binding affinity of the C-terminal region. It is thought that the rs27044 polymorphic modification may change the hydrogen bonding pattern and this may affect the substrate affinity and peptide presentation mechanism. The data obtained in this study showed that the rs27044 polymorphism was associated with a genetic predisposition to RA disease ($p=0.037$, $OR=1.583$, $95\%CI=1.028-2.440$).

The *ERAP1* rs26653 polymorphism has been associated with the risk of many autoimmune and infectious diseases [33, 34]. Macieja et al. conducted a study with 148 patients and 146 healthy controls showed that the rs26653 polymorphism

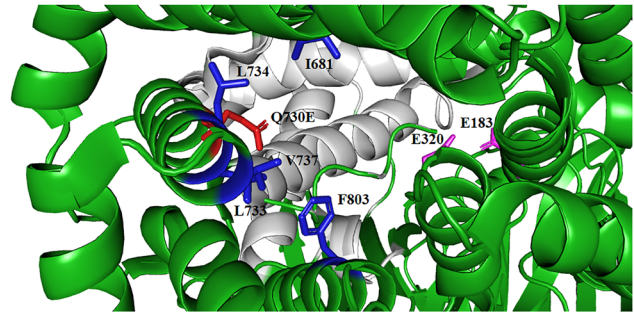


Figure 1: Cartoon illustration of the positional state of the Gln730Glu polymorphism of *ERAP1* in the C-terminal specificity pocket of *ERAP1*.

Blue color: C-terminal specificity residues; red color: polymorphic residue; magenta color: N-terminal anchor.

was associated with the risk of psoriasis ($p=3.1 \times 10^{-5}$) [34]. In another study, Li et al. showed the association of *ERAP1* rs26653 ($p=0.001$) and rs27044 ($p=0.003$) polymorphisms with cervical cancer risk [35]. In this study, the rs26653 Arg127Pro polymorphism was shown to be associated with the risk of rheumatoid arthritis disease ($p=0.002$, $OR=2.001$, $95\%CI=1.276-3.137$). It was determined that the Arg127Pro polymorphic change caused a change in the *ERAP1* substrate entry pocket, and the distance between 127th residues and 900th residues in the entrance region expanded from 7.5 to 10.3 Å (Figure 2). It was observed that Arg→Pro change at residue 127 transformed the loop secondary formation into a semi-helical form. The association of the rs26653 polymorphism with the RA disease risk indicates that this formal change in the entry pocket of ligand and substrate exerts effects on the activity of the enzyme that will affect the peptide processing mechanism.

Intronic variant rs27582 was found to be associated with RA disease risk in the Turkish population ($p<0.05$, $OR=0.348$, $95\%CI=0.194-0.622$). The rs27582 polymorphic change has been associated with an autoimmune disease, AS, in different populations. The *ERAP1* rs27582 polymorphism was associated with the risk of AS disease in a study conducted with 602 patients and 619 controls in the Beijing Han community [36]. Despite the link established in East Asia, the association between *ERAP1* and rs27582 could not be shown in terms of susceptibility to AS in a study conducted with 730 patients and 1,021 controls in white European cohorts [27].

The rs28096 polymorphism in the intron region of the *ERAP1* gene was not associated with RA disease risk ($p=0.509$, $OR=1.167$, $95\%CI=0.738-1.847$). As in other polymorphic regions investigated in this study, although the rs28096 region has not been previously examined in terms of RA disease risk, its relationship with different

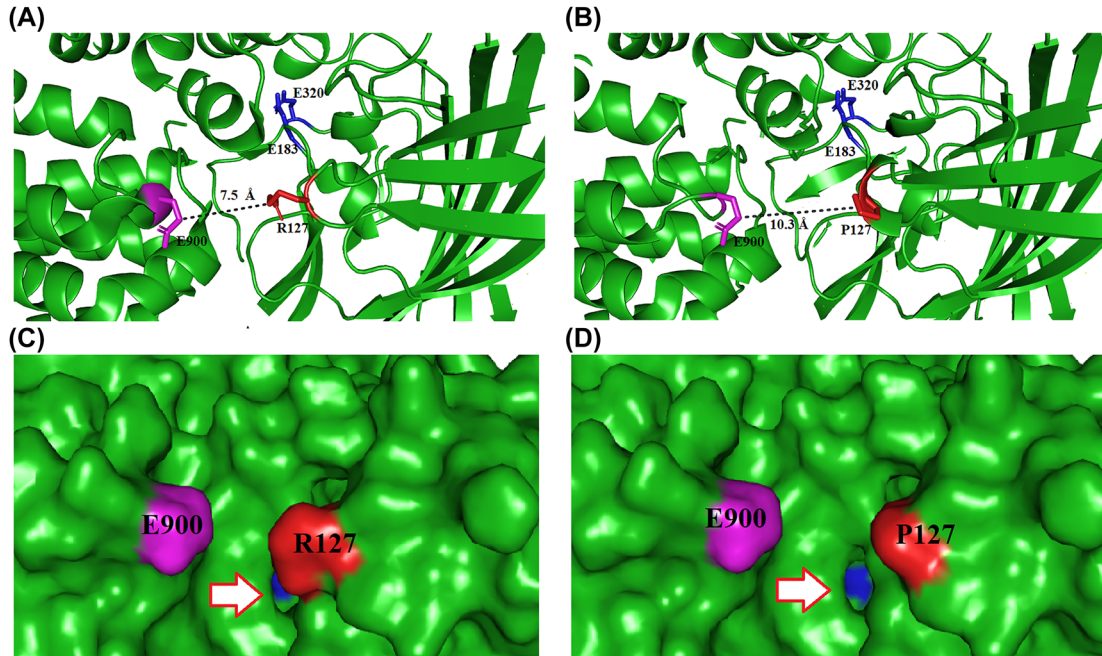


Figure 2: The representation of the changes caused by Arg127Pro polymorphism in protein structure of *ERAP1*.

Blue color: N-terminal anchor, red color: polymorphic residue, magenta color: substrate entry pocket residue; (A) Cartoon illustration of the distance between residues Glu900 and Arg127 in wild *ERAP1*; (B) Surface illustration of the positional states of substrate entry pocket residues Glu900 and Arg127 in wild *ERAP1*; (C) Cartoon illustration of the distance between residues Glu900 and Pro127 in polymorphic *ERAP1*; (D) Surface illustration of the positional states of substrate entry pocket residues Glu900 and Pro127 in polymorphic *ERAP1*.

diseases was examined. While Dargahi et al. could not correlate the risk of preeclampsia disease with rs28096 polymorphism, Harvey et al. stated that the rs28096 polymorphic change increased *ERAP1* gene expression excessively and was associated with AS disease risk ($p=1.6 \times 10^{-32}$) [27, 37]. It is known that the 3' untranslated regions of genes are important structures that affect the post-transcriptional regulation and translation rate [38]. This basic function lies behind the rs28096 polymorphic change within this region causing a change in the expression levels of the gene.

Evaluating these data contributes to approaches phylogeographical and genetic basis and risk factors for diseases. In addition, expected to provide valuable contributions to discuss the mechanisms of variations and mechanisms of modeling to the potential of gene therapy approaches. The data obtained for both healthy control and RA populations are in HWE ($p>0.05$) for except three loci of the five polymorphic loci as shown in Supplemental Table 4. However, utilizing a mismatch distribution method, the demographic histories of the two groups are examined. For these two populations, the mismatch distribution seemed to be multimodal (multimodal distribution; the exponential increase is not regular) (Supplemental Table 3, $p \leq 0.05$). These results show that there may be genetic drift-induced transitions such

as population motions, emigrations, and environmental impacts in the Malatya province gene pool that lead to the formation of the RA population. Historical population parameters (θ_0 , and θ_1) and τ values show a similar historical growth period for the populations (Supplemental Table 2). Supplemental Table 2 demonstrated high and peer haplotypic diversity (h), weak nucleotide diversity (π) and peer average number of pairwise nucleotide differences (k) between the two populations [39]. In addition, for the two groups, the Fu's F_s stats revealed a strong negative value, suggesting a comparable population expansion for both communities across history. For two populations, Tajima's D values were negligible ($p>0.05$), indicating that populations are at neutral equilibrium (Supplemental Table 2). These results demonstrate that in the historical period, the molecular diversity of both groups has genetically similar growth and expansion [40]. These data also indicate that in Malatya, Turkey, the RA population does not derive from the healthy control population. Despite similar genetic heritage, both populations have distinct haplotype variants. The reason for these differences is thought to be due to the departure from HWE at the third and fourth loci (Supplemental Table 2). The probable impact of the RA on polymorphic loci in the rs26653, rs27044, rs27582, rs28096, and rs30187 may reason for this haplotypic alteration between healthy control and RA

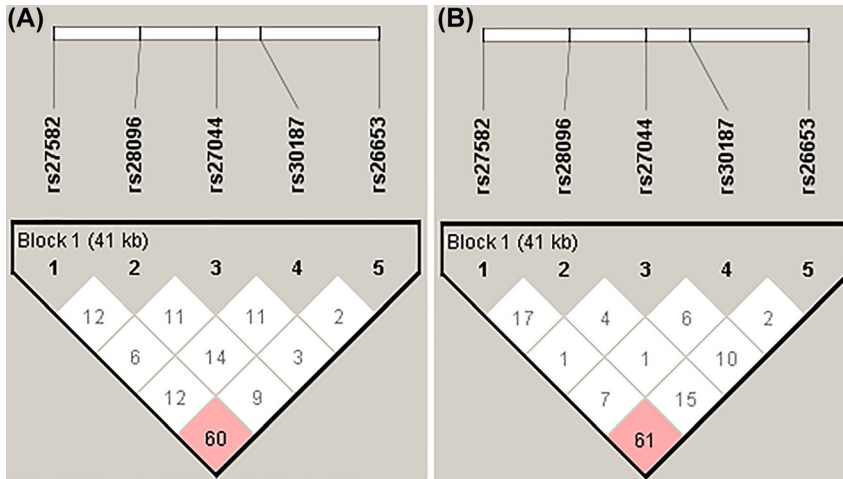


Figure 3: Haploview plot illustrating the LD of the RA and control populations.

At the top the SNPs are shown according to their succession from the start of translation of the gene. Numbers indicate the D' value expressed as a percentile. The plot of pair-wise LD analysis of (A) RA population standard D' values and (B) control population standard D' values (Diamonds are colored pink if the D' value is low [LD is weak] and the confidence in the value of D' is high as well, strong LD with LOD scores for $LD \geq 2$; white squares, $D' < 1.0$ and $LOD \leq 2$).

populations. LD analysis revealed weak allelic combinations at positions (rs27582 and rs26653) in the RA and control populations (Figure 3). This result supports the low frequency of the risk haplotype [CGAAT] in the RA population (Table 3).

In conclusion, it is known that polymorphic changes, which were examined for the first time in this study in terms of the RA relationship, are also associated with different diseases. It was determined that four exonic and intronic polymorphic regions (rs26653, rs27044, rs27582, and rs30187) were associated with disease risk. We report a haplotype [CGAAT] that is associated with RA risk from Turkey that has not been described before. It has been shown that polymorphic missense changes cause alterations in protein tertiary structure that may affect functional properties. The association of *ERAP1* with the risk of many autoimmune diseases of polymorphic regions involved in RA disease risk does not explain its specific role in RA development. It is thought that other biological structures (genes, ligands, etc.) that interact with peptide products resulting from *ERAP1* polymorphic changes may contribute to the development of the disease. It is thought that studies to be carried out with large sample groups including different polymorphic regions of *ERAP1* will contribute to the elucidation of the RA-*ERAP1* relationship.

Research funding: This study was supported by Munzur University Research Budget (2017-DT).

Author contribution: Conceptualization, EA, TY; methodology, EA, TY, OÖ; validation, EA, TY, OÖ; formal analysis, EA, TY, OÖ; investigation, EA, TY, OÖ; data

curation, EA, OÖ; writing-original draft preparation, EA, TY, and OÖ; writing—review and editing, EA, TY, and OÖ; visualization, EA; supervision, EA; project administration, EA; funding acquisition, EA. All authors have read and agreed to the published version of the manuscript.

Conflict of interest: The authors have no conflict of interest.

Informed consent: Informed consent was obtained from all individuals included in this study.

Ethical approval: Ethical rules were followed in this study and written consent of volunteers was obtained. The study protocol was approved by the ethics committee of Malatya Clinical Investigations (2016/12). The study followed the ethical standards of the Helsinki Declaration.

References

1. Alamanos Y, Drosos AA. Epidemiology of adult rheumatoid arthritis. *Autoimmun Rev* 2005;4:130–6.
2. Smolen J, Aletaha D, Barton A, Burmester G, Emery P, Firestein G, et al. Rheumatoid arthritis. *Nat Rev Dis Prim* 2018; 4:1–23.
3. Okada Y, Kim K, Han B, Pillai N, Ong R, Saw W, et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. *Hum Mol Genet* 2014;23:6916–26.
4. Deane KD, Demoruelle MK, Kelmenson LB, Kuhn K, Norris J, Holers V. Genetic and environmental risk factors for rheumatoid arthritis. *Best Pract Res Clin Rheumatol* 2017;31:3–18.
5. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000;43:30–7.

6. Bogunia-Kubik K, Świerkot J, Malak A, Wyszczkańska B, Nowak B, Białowas K, et al. *IL-17A*, *IL-17F* and *IL-23R* gene polymorphisms in Polish patients with rheumatoid arthritis. *Arch Immunol Ther Exp* 2015;63:215–21.
7. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;506:376–81.
8. Kim JM, Kim HY. Pathogenesis of rheumatoid arthritis. *J Korean Med Assoc* 2010;53:853–61.
9. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. *Lancet* 2016;388:2023–38.
10. Saric T, Chang SC, Hattori A, York I, Markant S, Rock K, et al. An IFN- γ -induced aminopeptidase in the ER, *ERAP1*, trims precursors to MHC class I-presented peptides. *Nat Immunol* 2002;3:1169–76.
11. Haroon N, Inman RD. Endoplasmic reticulum aminopeptidases: biology and pathogenic potential. *Nat Rev Rheumatol* 2010;6:461–7.
12. Evnouchidou I, Papakyriakou A, Stratikos E. A new role for Zn(II) aminopeptidases: antigenic peptide generation and destruction. *Curr Pharmaceut Des* 2009;15:3656–70.
13. Mpakali A, Maben Z, Stern LJ, Stratikos E. Molecular pathways for antigenic peptide generation by ER aminopeptidase 1. *Mol Immunol* 2019;113:50–7.
14. Stratikos E, Stamogiannos A, Zervoudi E, Fruci D. A role for naturally occurring alleles of endoplasmic reticulum aminopeptidases in tumor immunity and cancer predisposition. *Front Oncol* 2014;4:1–10.
15. López de Castro JA, Alvarez-Navarro C, Brito A, Guasp P, Martín-Esteban A, Sanz-Bravo A. Molecular and pathogenic effects of endoplasmic reticulum aminopeptidases *ERAP1* and *ERAP2* in MHC-I-associated inflammatory disorders: towards a unifying view. *Mol Immunol* 2016;77:193–204.
16. Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the sequenom massarray iPLEX platform. *Curr Protoc Hum Genet* 2009;60:1–18.
17. Chen VB, Arendall WB, Headd JJ, Keedy D, Immormino R, Kapral G, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr Sect D Biol Crystallogr* 2010;66:12–21.
18. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 2007;35:407–10.
19. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct Funct Genet* 2004;57:702–10.
20. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinf Online* 2005;1:47–50.
21. Barrett JC, Fry B, Maller J, Daly M. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–65.
22. Schneider S, Excoffier L. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* 1999;152:1079–89.
23. Rogers AR. Genetic evidence for a pleistocene population explosion. *Evolution* 1995;49:608–15.
24. Giastas P, Neu M, Rowland P, Stratikos E. High-Resolution crystal structure of endoplasmic reticulum aminopeptidase 1 with bound phosphinic transition-state analogue inhibitor. *ACS Med Chem Lett* 2019;10:708–13.
25. Sui L, Gandhi A, Guo HC. Crystal structure of a polypeptide's C-terminus in complex with the regulatory domain of ER aminopeptidase 1. *Mol Immunol* 2016;80:41–9.
26. Stamogiannos A, Koumantou D, Papakyriakou A, Stratikos E. Effects of polymorphic variation on the mechanism of endoplasmic reticulum aminopeptidase 1. *Mol Immunol* 2015;67:426–35.
27. Harvey D, Pointon JJ, Evans DM, Karaderi T, Farrar C, Appleton L, et al. Investigating the genetic association between *ERAP1* and ankylosing spondylitis. *Hum Mol Genet* 2009;18:4204–12.
28. Guerini FR, Cagliani R, Forni D, Agliardi C, Caputo D, Cassinotti A, et al. A functional variant in *ERAP1* predisposes to multiple sclerosis. *PLoS One* 2012;7:1–4.
29. Giancchetti E, Crinò A, Palma A, Luciano R, Perri V, Fruci D, et al. Case-control analysis of the *ERAP1* polymorphism rs30187 in Italian type 1 diabetes mellitus patients. *Health* 2013;05:2150–5.
30. Evnouchidou I, Momburg F, Papakyriakou A, Chroni A, Leondiadis L, Chang S, et al. The internal sequence of the peptide-substrate determines its N-terminus trimming by *ERAP1*. *PLoS One* 2008;3:1–12.
31. Liddle J, Hutchinson JP, Kitchen S, Rowland P, Neu M, Ceconie T, et al. Targeting the regulatory site of ER aminopeptidase 1 leads to the discovery of a natural product modulator of antigen presentation. *ACS Appl Mater Interfaces* 2020;63:3348–58.
32. Kochan G, Krojer T, Harvey D, Fischer R, Chen L, Vollmar M, et al. Crystal structures of the endoplasmic reticulum aminopeptidase-1 (*ERAP1*) reveal the molecular basis for N-terminal peptide trimming. *Proc Natl Acad Sci U S A* 2011;108:7745–50.
33. Fu Y, Li X, Chen Y, Liu R, Wang R, Bai N. Association of *ERAP1* gene polymorphisms with the susceptibility to psoriasis vulgaris a case-control study. *Medicine* 2018;97:1–5.
34. Stawczyk-Macieja M, Szczerkowska-Dobosz A, Rebała K, Gabig-Ciminska M, Nowicki R, Haras A, et al. *ERAP1* and *HLA-C*06* are strongly associated with the risk of psoriasis in the population of northern Poland. *Postepy Dermatol Alergol* 2018;35:286–92.
35. Li C, Li Y, Yan Z, Dai S, Liu S, Wang X, et al. Polymorphisms in endoplasmic reticulum aminopeptidase genes are associated with cervical cancer risk in a Chinese Han population. *BMC Cancer* 2020;20:1–11.
36. Zhang Z, Dai D, Yu K, Yuan F, Jin J, Ding L, et al. Association of *HLA-B27* and *ERAP1* with ankylosing spondylitis susceptibility in Beijing Han Chinese. *Tissue Antigens* 2014;83:324–9.
37. Dargahi H, Nicknam MH, Mirahmadian M, Mahmoudi M, Aslani S, Sadrosadat M, et al. Association study of single nucleotide polymorphisms of endoplasmic reticulum aminopeptidase 1 and 2 genes in Iranian women with preeclampsia. *Iran J Public Health* 2019;48:531–40.
38. Schwerk J, Savan R. Translating the untranslated region. *J Immunol* 2015;195:2963–71.
39. Harpending HC. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 1994;66:591–600.
40. Tajima F. Measurement of DNA polymorphism. In: Clark A G, Takahata N, editors *Mechanisms of molecular evolution*. Tokyo: Sinauer Associates; 1993. pp. 37–59.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/tjb-2022-0007>).