Published in final edited form as:

Arthritis Rheumatol. 2015 October ; 67(10): 2611–2623. doi:10.1002/art.39228.

Interactions between amino-acid-defined MHC class II variants and smoking for seropositive rheumatoid arthritis

Kwangwoo Kim, PhD^{1,*}, Xia Jiang, BSc^{2,*}, Jing Cui, MD, PhD^{3,*}, Bing Lu, MD, DrPH⁴, Karen H. Costenbader, MD, MPH⁴, Jeffrey A. Sparks, MD, MMSc⁴, So-Young Bang, MD, PhD⁵, Hye-soon Lee, MD, PhD⁵, Yukinori Okada, MD, PhD⁶, Soumya Raychaudhuri, MD, PhD⁷, Lars Alfredsson, PhD^{2,†}, Sang-Cheol Bae, MD, PhD, MPH^{5,†}, Lars Klareskog, MD, PhD^{8,†}, and Elizabeth W. Karlson, MD^{4,†}

¹Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of Korea, Brigham and Women's Hospital, Boston, Massachusetts, Harvard Medical School, Boston, Massachusetts, Broad Institute, Cambridge, Massachusetts

²Institute of Environmental Medicine, Karolinska Institute, Stockholm, Sweden

³Brigham and Women's Hospital, Boston, Massachusetts, Harvard Medical School, Boston, Massachusetts, Broad Institute, Cambridge, Massachusetts

⁴Brigham and Women's Hospital, Boston, Massachusetts, Harvard Medical School, Boston, Massachusetts

⁵Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of Korea

⁶Brigham and Women's Hospital, Boston, Massachusetts, Harvard Medical School, Boston, Massachusetts, Broad Institute, Cambridge, Massachusetts, Tokyo Medical and Dental University, Tokyo, Japan, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

⁷Brigham and Women's Hospital, Boston, Massachusetts, Harvard Medical School, Boston, Massachusetts, Broad Institute, Cambridge, Massachusetts, Manchester Academic Health Sciences Centre, Manchester, UK

⁸Rheumatology Unit, Department of Medicine, Karolinska Institutet/Karolinska Hospital, Stockholm, Sweden

Abstract

Address correspondence to: Dr. Elizabeth Karlson, Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115. Dr. Lars Klareskog, Rheumatology Unit, Karolinska Hospital, 171 16 Stockholm, Sweden. Dr. Sang-Cheol Bae, Hanyang University Hospital for Rheumatic Diseases, Seoul 133-792, Republic of Korea.

*Dr. Kwangwoo Kim, Xia Jiang and Dr. Jing Cui contributed equally to this work

†Dr. Elizabeth Karlson, Dr. Sang-Cheol Bae, Dr. Lars Klareskog and Dr. Lars Alfredsson contributed equally to this work

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Drs. Bae, Klareskog and Karlson had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Karlson, Klareskog, Bae, Alfredsson, Kim, Jiang

Acquisition of data. Kim, Jiang, Costenbader, Sparks, Bang, Lee, Okada, Raychaudhuri, Alfredsson, Bae, Klareskog, Karlson

Analysis and interpretation of data. Kim, Jiang, Cui, Lu, Alfredsson, Bae, Klareskog, Karlson

Objective—This study aimed to refine the interaction between cigarette smoking and human leukocyte antigen (HLA) polymorphisms in seropositive rheumatoid arthritis (RA), in the context of a recent amino-acid based HLA model for RA susceptibility.

Methods—We imputed HLA amino acids and classical alleles from case-control Immunochip array data of 3,588 Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA; case/control=1654/1934), 589 Nurses' Health Study (NHS; 229/360) and 2,125 Korean (1390/735) subjects. We examined interaction effects between heavy smoking (>10 pack-years) and genetic risk score (GRS) from RA-associated amino-acid positions (11, 13, 71 and 74 in HLA-DRβ1; 9 in HLA-B; and 9 in HLA-DPβ1) and from HLA-DRβ1 four-amino-acid haplotypes with an attributable proportion due to interaction (AP) using the additive interaction model.

Results—Heavy smoking and all investigated HLA amino-acid positions and haplotypes were associated with RA susceptibility in all three populations. In the interaction analysis, we found a significant deviation from the expected additive joint effect between heavy smoking and the HLA-DRβ1 amino-acid haplotype in all three studies (0.416 AP 0.796). We further identified the key interacting variants as being located at amino-acid positions 11 and 13 of HLA-DRβ1 but not the other RA-risk amino-acid positions in all populations. At the positions 11 and 13, there were similar patterns between RA-risk effects and interaction effects of residues.

Conclusion—Our findings of significant gene-environment interaction effects implicate that a physical interaction between citrullinated auto-antigens produced by smoking and HLA-DR molecules is characterized by the HLA-DRβ1 four-amino-acid haplotype, primarily by the positions 11 and 13.

INTRODUCTION

Rheumatoid arthritis (RA; OMIM180300) is a complex chronic inflammatory disease associated with multiple genetic and environmental factors. The strongest RA susceptibility gene is for alleles at the *HLA-DRB1* gene (1–3), the so-called “shared epitope”, and the influence of HLA genes are largely to the RA subset characterized by the presence of antibodies to citrullinated peptides (ACPA) (4–7). Cigarette smoking is the most established RA environmental risk factor, with accumulating evidence demonstrating its risk effect as well as dose-response effect, also mainly for seropositive (ACPA and/or rheumatoid factor (RF)-positive) RA risk (8–13) (14–16). Since 2004, several studies conducted in multiple populations with distinct ethnic backgrounds have consistently identified a significant interaction between smoking and shared epitope alleles for seropositive RA susceptibility based on the additive interaction model (4,12,17–21). Analyses have further revealed a dose-dependency of this synergistic effect as the most pronounced association and interaction estimates have been repeatedly observed among heavy smokers (>10 pack-years) carrying two copies of shared epitope risk alleles, compared to non-carrier never smokers. This suggests the importance of considering the cumulative “dose” for both smoking and major histocompatibility complex (MHC) class II genes when testing for their interactions (12,14–16,19). These findings have led to the hypothesis that citrullinated peptides produced by smoking are accommodated by HLA-DR molecules in an allele-specific manner and that presenting the citrullinated peptides by

HLA-DR with shared epitope-associated alleles is integrally related to seropositive RA pathogenesis (4,22).

The shared epitope hypothesis, originally described by Gregersen et al (1) defines causal alleles among classical 4-digit *HLA-DRB1* alleles that encode a common conserved amino-acid sequences (QKRAA, QRRRA or RRRRA) at position 70 to 74 in the third hypervariable region of the HLA-DR β -chain. For many years, it was difficult, however, to separate effects from the structure defined by the amino acids constituting the shared epitope and other structures within MHC class II molecules, due to the strong linkage disequilibrium (LD) within the MHC region (23). Several studies have indicated that different MHC structures are involved in defining the susceptibility to ACPA-positive RA. In addition to effects outside of *HLA-DRB1* (6,24), opposing (protective) effects have been described from some *HLA-DRB1* non-SE alleles, such as *03, *07, *08, *11, *12, and *13, in the development of seropositive RA (25–30). Likewise, an increasing number of studies have demonstrated significant interactions between certain *HLA-DRB1* non-SE alleles (*09:01 in Asians and *13:01 in Caucasians) and smoking (21,28,30). Taken together these studies demonstrate that the effects of MHC class II genetic variants as well as the interactions between smoking and MHC class II genes in conferring risk for RA are complex.

The redefinition of the structures of MHC class II molecules involved in the development of ACPA-positive RA in 2012, by Raychaudhuri *et al* (2) identified the strongest association signal for seropositive RA susceptibility at the HLA-DR β 1 amino-acid position 13 (and its tightly correlated position 11) whereas the traditional shared epitope (defined as positions from 70 to 74) as well as in HLA- DR β 1, position 9 in HLA-B and position 9 in HLA- DP β 1 all conferred independent but weaker effects. This amino-acid based HLA model was initially identified in Caucasian ancestries was later shown to be applicable also to Asian populations since the consistent association was observed in a large Asian study on ACPA-positive RA (3).

The use of the structure-based amino acid model (2) for analysis of how environmental risk factors interact with different variants of MHC class II molecules may allow better understanding of possible molecular mechanisms responsible for the development of different subsets of RA. In particular, these analyses may provide us with clues to which MHC class II structures are involved in presentation of neoantigens, for example citrullinated peptides, which may be generated from environmental exposure to smoking in the lungs and elsewhere (31). In the present study, we aimed to explore the interaction of smoking with the MHC class II structures defined by the variable amino-acid positions in three HLA genes: *HLA-DRB1*, *-B* and *-DPB1* by using three independent populations from Sweden, US and Korea.

PATIENTS AND METHODS

Subjects

We utilized genetic and epidemiologic data of 3,588 Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA), 589 Nurses' Health Study (NHS), and 2,125 Korean Observational Study Network for Arthritis subjects. The study subjects were surveyed for

smoking behavior prior to RA onset and genotyped by Immunochip, the high-density genotyping array for immune disease loci including the HLA region.

EIRA is a population-based case-control study initiated in 1996 and is still actively enrolling subjects. In the current study, we included all individuals recruited until 2009, consisting of 1,654 incident ACPA-positive RA cases and 1,934 controls available for both genetic and environmental information. Cases were recruited from all rheumatology providers in defined areas in Sweden with more than 85% recruited within one year after first symptoms onset and at first visit with a rheumatologist. Controls, matched by age, gender and residential area, were randomly selected from the Swedish National Register. At baseline, participants were asked to fill out a self-administrated questionnaire, as well as to provide blood samples for serological and genetic examinations. For the current analysis, the cumulative quantity of cigarettes smoked (pack-years) before the RA onset (or up to the index year among controls) were used.

Nurses' Health Study and Nurses' Health Study II are prospective cohort studies in the United States in which data on lifestyle and behavioral factors and disease outcomes are collected every two years by questionnaire. In a nested case-control study, 229 Caucasian female, seropositive (positive for RF or ACPA) incident RA cases and 360 controls were selected for genotyping on Immunochip. For this study, we used smoking information about quantity (pack-year) surveyed just prior to RA onset for the cases and matched time for the controls. The Korean subjects consist of 1,390 RA cases and 735 controls recruited from Hanyang University Hospital for Rheumatic Diseases in Seoul, Korea (21) and genotyped by Immunochip. All cases were positive for ACPA (disease duration at enrollment: mean 8.43 years; standard deviation: 8.44). Controls were recruited from healthy volunteers by a campaign at public places and the hospital, enriched in females aged >30 years to get the similar age range and gender ratio observed among cases. Clinical information was collected directly by trained interviewers using questionnaires. All control volunteers were examined for past medical history and those with a personal or family history of autoimmune diseases including rheumatic diseases were excluded. Smoking information was retrospectively recorded at RA onset for cases or at enrollment for controls.

The institutional review boards of all participating sites approved these studies. The baseline characteristics including smoking information of each cohort are shown in Table 1. The Immunochip data have been reported by Eyre S *et al* (32) for EIRA, by Karlson EW *et al* (33) for NHS, and by Kim K *et al* (34) for Korean.

HLA imputation

HLA imputation was performed using SNP2HLA software according to published instructions (35). For Caucasian datasets (EIRA and NHS), the dosage of HLA classical alleles and amino acids were imputed from the Immunochip data with a publicly released reference panel generated from 5,225 Caucasians by the Type 1 Diabetes Genetics Consortium (35). For the Korean dataset, imputation was performed with a recently developed reference panel constructed from 413 Koreans (36). Long-range haplotypes across the extended MHC were obtained, where we extracted the haplotypes defined from

HLA-DRβ1 amino-acid positions 11, 13, 71 and 74, as well as RA-associated amino-acid positions in HLA-DRβ1, -B and -DPβ1 for subsequent analyses.

Calculation of genetic risk score for HLA-DRβ1 amino-acid haplotypes and RA-associated amino-acid positions in HLA-DRβ1, -B and -DPβ1

To summarize the RA-risk effect of the previously characterized HLA-DRβ1 amino-acid haplotypes constructed by the amino-acid positions 11, 13, 71 and 74, we calculated a genetic risk score (GRS) for individual haplotypes by weighing each haplotype on its reported RA-risk effect size obtained from Raychaudhuri *et al* (2) for Caucasians and from Okada *et al* (3) for Koreans using the following equation:

$$GRS_i = \sum_{h=1}^n (\ln OR_h) \cdot haplotype_{hi}$$

where GRS_i is a haplotype GRS of individual i , OR_h is the reported odds ratio (OR) of the h^{th} haplotype and $haplotype_{hi}$ is the number of the h^{th} haplotype in individual i .

We also calculated GRS for each of the amino-acid positions 11, 13, 71 and 74 in HLA-DRβ1, position 9 in HLA-B and position 9 in HLA-DPβ1 by the equation:

$$GRS_{ki} = \sum_{a=1}^n (\ln OR_{ak}) \cdot dosage_{aki}$$

where GRS_{ki} is a GRS of amino-acid position k in individual i , OR_{ak} is the reported OR of the a^{th} residue at amino-acid position k and $dosage_{aki}$ is an imputed dosage value of the a^{th} residue at amino-acid position k in individual i .

The calculated GRS for haplotypes and amino-acid positions (which were continuous) were dichotomized at the median value of the controls in each population for a subsequent interaction analysis.

RA association tests

Pack-years of smoking was dichotomized at 10 pack-years: heavy smoking > 10 pack-years vs. non-heavy or never smoking ≤ 10 pack-years (11,12,14–16,19). Heavy smokers were compared with non-heavy/never smokers. The main effect of heavy smoking was calculated by logistic regression models with adjustment for age, gender and residential area in EIRA, for age in NHS, and for age and gender in Koreans.

The RA associations of overall residue difference at each of the reported RA-risk amino-acid positions were evaluated by a log-likelihood ratio test that estimates log-likelihood difference between null and full models. The null logistic regression model included the top five principal components (PCs) as predictors for RA risk. The full logistic regression model included all residues, excluding the most frequent residue, and the same five PCs. To calculate the omnibus P values for the combined dataset from the three datasets, a dummy variable indicating the three cohorts was additionally included in both null and full logistic

regression models. The association of dichotomized GRS was examined by logistic regression adjusting for the top five PCs.

Interaction tests

The interaction effects between heavy smoking and haplotype GRS as well as between heavy smoking and amino-acid position GRS were assessed using the attributable proportion due to interaction (AP) to determine if the joint effect between the two factors is significantly departed from additivity (19,37). For testing the interaction between heavy smoking and the dichotomized GRS for HLA-DRβ1 amino-acid haplotypes, we calculated disease odds ratios in three groups (with high GRS/heavy smoking, low GRS/heavy smoking, and high GRS/non-heavy smoking) with a reference group (with low GRS/non-heavy smoking) by using the following logistic regression model adjusting five PCs and continuous GRS at position 9 in HLA-B and -DPβ1 as covariates:

$$\log(odds) = \beta_0 + \sum_{i=1}^3 \beta_i G_i + \sum_{j=1}^5 \pi_j PC_j + \pi_6 GRS_{B_9} + \pi_7 GRS_{DP\beta1_9} + \varepsilon$$

Where G_1 , G_2 , and G_3 are indicate variables for high GRS/non-heavy smoking, low GRS/heavy smoking and high GRS/heavy smoking groups. PC_j is the j^{th} PC, and GRS_{B_9} and $GRS_{DP\beta1_9}$ are continuous GRS of HLA-B and -DPβ1 amino-acid position 9, respectively. β_0 is the logistic regression intercept. β_i and π are the effects of G_i and covariates (PC_j , GRS_{B_9} and $GRS_{DP\beta1_9}$). Similarly, to test the interactions between heavy smoking and the dichotomized GRS for the RA-associated amino-acid positions in HLA-DRβ1, -B and -DPβ1, we conditioned on five PCs and other amino-acid position GRS to remove GRS correlation effects generated by linkage disequilibrium (LD) among the amino-acid positions. For the two tightly correlated amino-acid positions at 11 and 13 of HLA-DRβ1, we did not adjust each position for the other position in the AP model. For example, when testing for HLA-DRβ1 amino-acid position 11, we conditioned on GRS from position HLA-DRβ1 amino-acid positions 71 and 74, HLA-B amino-acid position 9 and HLA-DPβ1 amino-acid position 9.

As an alternative method to remove the possible LD bias between the amino-acid positions, we residualized each continuous GRS of the tested amino-acid position by the other continuous GRS of the other amino-acid positions and the top five PCs using linear regression by the following equation:

$$R_k = GRS_k - \hat{GRS}_k$$

where GRS residual of amino-acid position k (R_k) is calculated by subtracting the predicted GRS ($\hat{GRS}_k$; e.g. for HLA-DRβ1 amino acid position 11,

$\hat{GRS}_{DR\beta1_11} = \beta_0 + \sum_{j=1}^5 \beta_j PC_j + \beta_6 GRS_{DR\beta1_71} + \beta_7 GRS_{DR\beta1_74} + \beta_8 GRS_{B_9} + \beta_9 GRS_{DP\beta1_9}$) from the actual GRS (GRS_k), considered as a proportion not explained by the RA-risk

effect (GRS) at the other positions in a regression model. We performed the same interaction analysis without covariates by using the dichotomized GRS residuals as a complement

Since each amino acid position harbors different residues, we finally investigated the residue effects of HLA-DR β 1 amino-acid positions 11 and 13 in the interaction with heavy smoking in RA. Each of the two amino-acid positions can be occupied by one of 6 possible residues. We coded each residue for each individual as a binary variable according to its presence (one residue or two residues) or absence (no residue). Since some residues have protective effect on RA susceptibility, we alternatively recoded these RA-protective residues to set up a correct reference combination of both genetic and epidemiologic factors in AP calculation, as previously suggested (38), and their APs were converted to the opposite direction to interpret them appropriately. All analyses were performed by PLINK, R and SAS.

RESULTS

Association of smoking with RA in the study subjects

Smoking status and intensity (pack-years of exposure) until the time of RA onset were obtained from the study populations. The smoking behavior varied considerably by population, especially, comparing ever-smoking rates among the Korean population where overall rates (7.6% in Korean controls) were much lower than the Caucasian populations (59.1% in EIRA controls and 58.3% in NHS controls) (Table 1). The overall low smoking rates in the Korean study reflects the high percentage of females who tend to be non-smokers for cultural reasons (Supplementary Table 1). For pack years of smoking, we saw similar population differences. In EIRA, we identified 754 heavy smoker and 893 non-heavy/never smokers among cases and 556 heavy smokers and 1,359 non-heavy/never smokers among the controls. In NHS, we identified 107 heavy smokers and 122 non-heavy/never smokers among RA cases and 129 heavy smokers and 231 non-heavy smokers among the controls. In Korean, we identified 164 heavy smokers and 1,226 non-heavy/never smokers among RA cases and 11 heavy smokers and 724 non-heavy smokers among controls.

Heavy smoking, a strong epidemiologic risk factor, was consistently associated with RA susceptibility in EIRA, NHS and Korean subjects (1.62 OR 4.73; 6.48×10^{-3} P 1.60×10^{-34}) (Table 2). Although rates of heavy smoking were much lower in the Korean subjects than the Caucasian subjects, the heavy smoking effect on RA was larger in the Koreans (Table 2). We note three important observations about the association of smoking with RA susceptibility on our study, in order to address potential concerns. First, to investigate whether there was a selection bias resulting in low ever-smoking rate among the Korean controls compared to Korean cases, we obtained the smoking status in Health2 Study (39), a community-based cohort study where the 2,987 participants aged 40 – 69 were recruited by Korea National Institute of Health (KNIH). We found very similar ever-smoking rates (3.2% in females and 76.4% in males vs. (3.0% in females and 76.1% in males) in this independent cohort of general subjects in Korea (Supplementary Table 1). Second, heavy smoking effects upon RA risk were similar in males and females in EIRA and Koreans (NHS includes females only). Third, when smoking was categorized as never

and ever smokers, the association became weaker in EIRA and Koreans, and disappeared in NHS (data not shown).

Associations of the reported HLA amino-acid positions with RA in the study subjects

We imputed HLA classical alleles, amino-acid residues and SNPs within the extended MHC region from the Immunochip data of 3,588 EIRA (case/control=1,654/1,934), 589 NHS (229/360) and 2,125 Korean (1,390/735) subjects, and achieved high concordance rates between the imputed and previously genotyped HLA-DRB1 alleles in all three populations (EIRA, concordance rate for 2-digit *04 alleles: 97.3%, 4-digit *04 alleles: 95.0%; NHS, 2-digit: 93.8%, 4 digit: 91.8%; Koreans, 2-digit: 97.4%, 4-digit: 91.4%), indicating a high quality of the imputed data.

From the imputed results, the most significant association for RA risk was found within *HLA-DRB1* in the extended MHC region in all three populations (Supplementary Figure 1). We assessed the association of amino-acid positions 11, 13, 71 and 74 in HLA-DRβ1, position 9 in HLA-B and position 9 in HLA-DPβ1 that were reported to have independent RA-risk effect in Caucasian and Asian populations (2,3). Consistent with previous reports, the most significant associations were found at the amino-acid positions 11 and 13 before and after conditioning on other amino-acid positions in each of the three populations and a single analysis ($P_{\text{omnibus}} = 7.17 \times 10^{-169}$ at the position 11 and $P_{\text{omnibus}} = 2.67 \times 10^{-166}$ at the position 13 in an unconditional analysis) (Table 3). Relatively weak or no associations at the other positions in HLA-DRβ1, -B and -DPβ1 were found in each population. However, independent associations at these positions were observed after conditioning on other position's effects in the combined analysis ($P_{\text{omnibus}} = 2.74 \times 10^{-3}$; Table 3).

Interactions between HLA-DRβ1 amino-acid haplotypes and heavy smoking

The haplotype defined by the amino-acid positions 11, 13, 71 and 74 is the best model to explain the *HLA-DRB1* association with RA susceptibility (2). We found that RA susceptibility was strongly associated with the dichotomized haplotype GRS in all three populations (Supplementary Table 2).

We evaluated interaction effect between the dichotomized haplotype GRS and heavy smoking and found significant interaction effects between the two factors in all three populations ($AP = 0.416$, 95%CI = 0.306 to 0.526, $P_{AP} = 1.16 \times 10^{-13}$ in EIRA; $AP = 0.467$, 95%CI = 0.187 to 0.748, $P_{AP} = 1.07 \times 10^{-3}$ in NHS; and $AP = 0.796$, CI = 0.535 to 1.058, $P_{AP} = 9.73 \times 10^{-14}$ in Koreans; Table 4). We note that the number of heavy smokers in the Korean control group subjects was too small to get a stable AP value and might cause an overestimated AP in the Korean population.

Interactions between RA-risk HLA amino-acid positions and heavy smoking

We further evaluated interaction effects between heavy smoking at the HLA amino-acid position level using the dichotomized GRS for each RA-risk amino-acid positions in HLA-DRβ1, -B and -DPβ1. To minimize potential correlation between the amino-acid positions due to their LD, we added continuous GRS of non-tested amino-acid positions as covariates. Significant and consistent synergy was mapped to HLA-DRβ1 amino-acid positions 11 and

13 in all three populations (0.324 AP 0.721, 1.64×10^{-4} P_{AP} 1.45×10^{-5} ; Table 5). In contrast, inconsistent interaction effects were found at the other positions, including HLA-DR β 1 amino-acid positions 71 and 74 within the traditional shared epitope, in all three study populations. When we performed the same interaction analysis using the dichotomized GRS residual, we also observed significant AP at either HLA-DR β 1 position 11 or 13 in both EIRA and NHS (Supplementary Table 3).

Correlation between HLA amino-acid residue/heavy smoking effects and RA-risk effects

We observed a positive correlation between the interaction effect and RA-risk effect at most residues at the positions 11 and 13 (Figure 1) in all three populations. For example, the most significant interaction effect with positive direction was found at Val11 and His13, which are the same residues associated with the strongest risk of RA in previous studies (2,3). Similarly, the most RA-protective residues (Ser11 and Ser13) showed the strongest AP with negative direction.

DISCUSSION

In this study, we applied novel HLA models for RA risk in evaluating gene-environment interactions for RA risk. We found consistent and strong additive interactions between heavy smoking and the HLA-DR β 1 amino-acid haplotype model in Swedish, US, and Korean datasets, and further identified the key interacting polymorphisms by testing the interaction at the HLA amino-acid position level and amino-acid residue level.

In addition to the synergistically additive interaction between heavy smoking and HLA-DR β 1 amino-acid haplotype, we report two other novel observations in this study relevant to RA pathogenesis. First, our case-control study mapped the strongest interaction effect to the positions 11 and 13 in HLA-DR β 1 among the four amino acids comprising the haplotype interacting with smoking, whereas weaker or inconsistent effects were seen when interaction analysis was performed for HLA-DR β 1 positions 70 and 74, HLA-B amino-acid position 9 and HLA-DP β 1 amino-acid position 9 in our study populations. Second, we found a similar pattern between interaction effects of amino-acid residues at position 11 and 13 with heavy smoking and main effects on RA risk of those residues that are known to explain a majority of RA risk effect in the HLA-DR β 1 amino-acid haplotype (2,3).

We took advantage of a weighted GRS method (40,41) in the interaction analysis for the highly polymorphic HLA variants (e.g. 15 possible HLA-DR β 1 amino-acid haplotypes, 3 to 6 possible residues at each RA-risk amino-acid position). Because diverse amino-acid haplotypes or residues with different effect sizes could be aggregated into a single GRS value, all subsequent analyses became simplified, powerful and easily interpretable. Furthermore, in contrast to the traditional shared epitope model, our approach could accommodate effects of non-SE RA-risk alleles (e.g., HLA-DRB1*09:01) in the interaction tests.

The variety of the study designs included in our analysis adds to the strength of the consistency of the associations we described. Strengths of all three cohort studies include the enrollment of incident or recently diagnosed RA cases in NHS and EIRA, the collection of

based model for detailed analysis of the contribution of variants of MHC to the emergence of potential pathogenic MHC class II restricted and environmentally triggered immune reactions, offers novel opportunities to define and eventually modify specific immune reactions that drive defined subsets of RA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding sources:

The EIRA study was supported by grants from the Swedish Medical Research Council, from the Swedish Council for Working life and Social Research, from King Gustaf V's 80-year foundation, from the Swedish Rheumatism Foundation, from Stockholm County Council, from the insurance company AFA, the Innovative Medicines Initiative (IMI) supported BTCure project.

The NHS was supported by National Institutes of Health (NIH) grants AR049880, AR052403, AR059073, CA87969, CA176726 CA186107, CA49449, CA67262.

The Korean cohort was supported by the Korea Healthcare technology R&D project of Ministry for Health & Welfare (HI13C2124). Health2 Study data was provided from Korean Biobank Project supported by the Korea Center for Disease Control and Prevention at the Korea National Institute of Health.

Dr. Kim was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2013R1A6A3A03023016).

Dr. Raychaudhuri was supported by R01AR063759-01A1, U01GM092691-04, and R01AR065183-01.

We are grateful to all participants for this study in Sweden, the US and Korea for contributing samples and data.

References

1. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis and rheumatism*. 1987; 30(11):1205–1213. [PubMed: 2446635]
2. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature genetics*. 2012; 44(3):291–296. [PubMed: 22286218]
3. Okada Y, Kim K, Han B, Pillai NE, Ong RT, Saw WY, et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. *Human molecular genetics*. 2014; 23(25):6916–6926. [PubMed: 25070946]
4. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, et al. A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis and rheumatism*. 2006; 54(1): 38–46. [PubMed: 16385494]
5. Verpoort KN, Cheung K, Ioan-Facsinay A, van der Helm-van Mil AH, de Vries-Bouwstra JK, Allaart CF, et al. Fine specificity of the anti-citrullinated protein antibody response is influenced by the shared epitope alleles. *Arthritis and rheumatism*. 2007; 56(12):3949–3952. [PubMed: 18050209]
6. Ding B, Padyukov L, Lundstrom E, Seielstad M, Plenge RM, Oksenberg JR, et al. Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. *Arthritis and rheumatism*. 2009; 60(1):30–38. [PubMed: 19116921]

7. Han B, Diogo D, Eyre S, Kallberg H, Zhernakova A, Bowes J, et al. Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. *American journal of human genetics*. 2014; 94(4):522–532. [PubMed: 24656864]
8. Criswell LA, Merlino LA, Cerhan JR, Mikuls TR, Mudano AS, Burma M, et al. Cigarette smoking and the risk of rheumatoid arthritis among postmenopausal women: results from the Iowa Women's Health Study. *The American journal of medicine*. 2002; 112(6):465–471. [PubMed: 11959057]
9. Stolt P, Bengtsson C, Nordmark B, Lindblad S, Lundberg I, Klareskog L, et al. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Annals of the rheumatic diseases*. 2003; 62(9):835–841. [PubMed: 12922955]
10. Pedersen M, Jacobsen S, Klarlund M, Pedersen BV, Wiik A, Wohlfahrt J, et al. Environmental risk factors differ between rheumatoid arthritis with and without auto-antibodies against cyclic citrullinated peptides. *Arthritis research & therapy*. 2006; 8(4):R133. [PubMed: 16872514]
11. Costenbader KH, Feskanich D, Mandl LA, Karlson EW. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. *The American journal of medicine*. 2006; 119(6):503 e501–509. [PubMed: 16750964]
12. Mikuls TR, Sayles H, Yu F, Levan T, Gould KA, Thiele GM, et al. Associations of cigarette smoking with rheumatoid arthritis in African Americans. *Arthritis and rheumatism*. 2010; 62(12):3560–3568. [PubMed: 20722010]
13. Sugiyama D, Nishimura K, Tamaki K, Tsuji G, Nakazawa T, Morinobu A, et al. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. *Annals of the rheumatic diseases*. 2010; 69(1):70–81. [PubMed: 19174392]
14. Kallberg H, Ding B, Padyukov L, Bengtsson C, Ronnelid J, Klareskog L, et al. Smoking is a major preventable risk factor for rheumatoid arthritis: estimations of risks after various exposures to cigarette smoke. *Annals of the rheumatic diseases*. 2011; 70(3):508–511. [PubMed: 21149499]
15. Di Giuseppe D, Orsini N, Alfredsson L, Askling J, Wolk A. Cigarette smoking and smoking cessation in relation to risk of rheumatoid arthritis in women. *Arthritis research & therapy*. 2013; 15(2):R56. [PubMed: 23607815]
16. Di Giuseppe D, Discacciati A, Orsini N, Wolk A. Cigarette smoking and risk of rheumatoid arthritis: a dose-response meta-analysis. *Arthritis research & therapy*. 2014; 16(2):R61. [PubMed: 24594022]
17. Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis and rheumatism*. 2004; 50(10):3085–3092. [PubMed: 15476204]
18. Lundstrom E, Kallberg H, Alfredsson L, Klareskog L, Padyukov L. Gene-environment interaction between the DRB1 shared epitope and smoking in the risk of anti-citrullinated protein antibody-positive rheumatoid arthritis: all alleles are important. *Arthritis and rheumatism*. 2009; 60(6):1597–1603. [PubMed: 19479873]
19. Karlson EW, Chang SC, Cui J, Chibnik LB, Fraser PA, De Vivo I, et al. Gene-environment interaction between HLA-DRB1 shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis. *Annals of the rheumatic diseases*. 2010; 69(1):54–60. [PubMed: 19151010]
20. Bang SY, Lee KH, Cho SK, Lee HS, Lee KW, Bae SC. Smoking increases rheumatoid arthritis susceptibility in individuals carrying the HLA-DRB1 shared epitope, regardless of rheumatoid factor or anti-cyclic citrullinated peptide antibody status. *Arthritis and rheumatism*. 2010; 62(2):369–377. [PubMed: 20112396]
21. Bang SY, Lee HS, Lee KW, Bae SC. Interaction of HLA-DRB1*09:01 and *04:05 with smoking suggests distinctive mechanisms of rheumatoid arthritis susceptibility beyond the shared epitope. *The Journal of rheumatology*. 2013; 40(7):1054–1062. [PubMed: 23637323]
22. Klareskog L, Ronnelid J, Lundberg K, Padyukov L, Alfredsson L. Immunity to citrullinated proteins in rheumatoid arthritis. *Annual review of immunology*. 2008; 26:651–675.
23. de Bakker PI, McVean G, Sabeti PC, Miretti MM, Green T, Marchini J, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nature genetics*. 2006; 38(10):1166–1172. [PubMed: 16998491]

24. Vignal C, Bansal AT, Balding DJ, Binks MH, Dickson MC, Montgomery DS, et al. Genetic association of the major histocompatibility complex with rheumatoid arthritis implicates two non-DRB1 loci. *Arthritis and rheumatism*. 2009; 60(1):53–62. [PubMed: 19116923]
25. del Rincon I, Escalante A. HLA-DRB1 alleles associated with susceptibility or resistance to rheumatoid arthritis, articular deformities, and disability in Mexican Americans. *Arthritis and rheumatism*. 1999; 42(7):1329–1338. [PubMed: 10403259]
26. Irigoyen P, Lee AT, Wener MH, Li W, Kern M, Batliwalla F, et al. Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles. *Arthritis and rheumatism*. 2005; 52(12):3813–3818. [PubMed: 16320316]
27. Cui J, Taylor KE, Destefano AL, Criswell LA, Izmailova ES, Parker A, et al. Genome-wide association study of determinants of anti-cyclic citrullinated peptide antibody titer in adults with rheumatoid arthritis. *Molecular medicine*. 2009; 15(5–6):136–143. [PubMed: 19287509]
28. Lundstrom E, Kallberg H, Smolnikova M, Ding B, Ronnelid J, Alfredsson L, et al. Opposing effects of HLA-DRB1*13 alleles on the risk of developing anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis and rheumatism*. 2009; 60(4):924–930. [PubMed: 19333936]
29. Chun-Lai T, Padyukov L, Dhaliwal JS, Lundstrom E, Yahya A, Muhamad NA, et al. Shared epitope alleles remain a risk factor for anti-citrullinated proteins antibody (ACPA)--positive rheumatoid arthritis in three Asian ethnic groups. *PloS one*. 2011; 6(6):e21069. [PubMed: 21698259]
30. van der Woude D, Lie BA, Lundstrom E, Balsa A, Feitsma AL, Houwing-Duistermaat JJ, et al. Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. *Arthritis and rheumatism*. 2010; 62(5):1236–1245. [PubMed: 20131291]
31. Catrina AI, Ytterberg AJ, Reynisdottir G, Malmstrom V, Klareskog L. Lungs, joints and immunity against citrullinated proteins in rheumatoid arthritis. *Nature reviews Rheumatology*. 2014; 10(11):645–653. [PubMed: 25072264]
32. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature genetics*. 2012; 44(12):1336–1340. [PubMed: 23143596]
33. Karlson EW, Ding B, Keenan BT, Liao K, Costenbader KH, Klareskog L, et al. Association of environmental and genetic factors and gene-environment interactions with risk of developing rheumatoid arthritis. *Arthritis care & research*. 2013; 65(7):1147–1156. [PubMed: 23495093]
34. Kim K, Bang SY, Lee HS, Cho SK, Choi CB, Sung YK, et al. High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. *Annals of the rheumatic diseases*. 2014
35. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PloS one*. 2013; 8(6):e64683. [PubMed: 23762245]
36. Kim K, Bang SY, Lee HS, Bae SC. Construction and application of a korean reference panel for imputing classical alleles and amino acids of human leukocyte antigen genes. *PloS one*. 2014; 9(11):e112546. [PubMed: 25398076]
37. Andersson T, Alfredsson L, Kallberg H, Zdravkovic S, Ahlbom A. Calculating measures of biological interaction. *European journal of epidemiology*. 2005; 20(7):575–579. [PubMed: 16119429]
38. Knol MJ, VanderWeele TJ, Groenwold RH, Klungel OH, Rovers MM, Grobbee DE. Estimating measures of interaction on an additive scale for preventive exposures. *European journal of epidemiology*. 2011; 26(6):433–438. [PubMed: 21344323]
39. Hwang JY, Sim X, Wu Y, Liang J, Tabara Y, Hu C, et al. Genome-Wide Association Meta-analysis Identifies Novel Variants Associated With Fasting Plasma Glucose in East Asians. *Diabetes*. 2015; 64(1):291–298. [PubMed: 25187374]

40. Karlson EW, Chibnik LB, Kraft P, Cui J, Keenan BT, Ding B, et al. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. *Annals of the rheumatic diseases*. 2010; 69(6):1077–1085. [PubMed: 20233754]
41. Yarwood A, Han B, Raychaudhuri S, Bowes J, Lunt M, Pappas DA, et al. A weighted genetic risk score using all known susceptibility variants to estimate rheumatoid arthritis risk. *Annals of the rheumatic diseases*. 2015; 74(1):170–176. [PubMed: 24092415]
42. Terao C, Ohmura K, Ikari K, Kawaguchi T, Takahashi M, Setoh K, et al. Effects of smoking and shared epitope on the production of anti-citrullinated Peptide antibody in a Japanese adult population. *Arthritis care & research*. 2014; 66(12):1818–1827. [PubMed: 24942650]
43. Scally SW, Petersen J, Law SC, Dudek NL, Nel HJ, Loh KL, et al. A molecular basis for the association of the HLA-DRB1 locus, citrullination, and rheumatoid arthritis. *The Journal of experimental medicine*. 2013; 210(12):2569–2582. [PubMed: 24190431]
44. James EA, Moustakas AK, Bui J, Papadopoulos GK, Bondinas G, Buckner JH, et al. HLA-DR1001 presents “altered-self” peptides derived from joint-associated proteins by accepting citrulline in three of its binding pockets. *Arthritis and rheumatism*. 2010; 62(10):2909–2918. [PubMed: 20533291]
45. James EA, Rieck M, Pieper J, Gebe JA, Yue BB, Tatum M, et al. Citrulline-specific Th1 cells are increased in rheumatoid arthritis and their frequency is influenced by disease duration and therapy. *Arthritis & rheumatology*. 2014; 66(7):1712–1722. [PubMed: 24665079]
46. Mahdi H, Fisher BA, Kallberg H, Plant D, Malmstrom V, Ronnelid J, et al. Specific interaction between genotype, smoking and autoimmunity to citrullinated alpha-enolase in the etiology of rheumatoid arthritis. *Nature genetics*. 2009; 41(12):1319–1324. [PubMed: 19898480]
47. van der Woude D, Alemayehu WG, Verduijn W, de Vries RR, Houwing-Duistermaat JJ, Huizinga TW, et al. Gene-environment interaction influences the reactivity of autoantibodies to citrullinated antigens in rheumatoid arthritis. *Nature genetics*. 2010; 42(10):814–816. author reply 816. [PubMed: 20877316]
48. Willemze A, van der Woude D, Ghiddey W, Levarht EW, Stoeken-Rijsbergen G, Verduyn W, et al. The interaction between HLA shared epitope alleles and smoking and its contribution to autoimmunity against several citrullinated antigens. *Arthritis and rheumatism*. 2011; 63(7):1823–1832. [PubMed: 21506097]
49. Lundberg K, Bengtsson C, Kharlamova N, Reed E, Jiang X, Kallberg H, et al. Genetic and environmental determinants for disease risk in subsets of rheumatoid arthritis defined by the anticitrullinated protein/peptide antibody fine specificity profile. *Annals of the rheumatic diseases*. 2013; 72(5):652–658. [PubMed: 22661643]
50. Fisher BA, Bang SY, Chowdhury M, Lee HS, Kim JH, Charles P, et al. Smoking, the HLA-DRB1 shared epitope and ACPA fine-specificity in Koreans with rheumatoid arthritis: evidence for more than one pathogenic pathway linking smoking to disease. *Annals of the rheumatic diseases*. 2014; 73(4):741–747. [PubMed: 23505239]

Figure 1.

Correlation between RA-risk effects and interaction effects of each residue at HLA-DRβ1 amino-acid positions 11 and 13. For each residue at the HLA-DRβ1 amino-acid positions 11 and 13, attributable proportion due to interaction (AP) with smoking for RA was plotted against to the reported odds ratio (OR) for RA susceptibility in Caucasians (Raychaudhuri S et al. *Nature genetics* 2012, 44(3):291) and Asians (Okada Y et al. *Human Molecular Genetics* 2014, 23(25):6916). Error bars display their 95% confidence intervals. Unstable AP with large 95% confidence intervals spanning from less than -1 to more than 1 were excluded from the plots. There are 6 correlation plots for the HLA-DRβ1 amino-acid position 11 in (a) EIRA, (b) NHS and (c) Korean, and HLA-DRβ1 amino acid position 13 in (d) EIRA, (e) NHS and (f) Korean.

Table 1

Baseline characteristics of RA cases and controls from the Epidemiologic Investigation of Rheumatoid Arthritis (EIRA), Nurses' Health Study (NHS), and Korean ACPA-positive RA or seropositive RA case control studies

Characteristics	EIRA (n=3588)		NHS (n=589)		Korean (n=2125)	
	ACPA-positive RA cases	Controls	Sero-positive RA cases	Controls	ACPA-positive RA cases	Controls
Subjects (n)	1654	1934	229	360	1390	735
Age, mean (SD)	51.08 (12.36)	54.22 (11.17)	54.33 (8.23)	54.42 (8.13)	44.02 (12.52)	34.67 (9.98)
Female (%)	70.44	73.47	100.00	100.00	84.17	93.74
Smoking characteristics						
Ever cigarette smokers (%)	72.95	59.09	62.01	58.33	18.13	7.62
Pack-years among smokers, mean (SD)	17.37 (14.56)	14.59 (19.93)	23.56 (15.83)	22.10 (21.20)	21.82 (18.34)	8.78 (10.52)
Smokers with > 10 pack-years (%)	45.76	29.03	46.72	35.83	11.80	1.50

Table 2

Association between heavy smoking and ACPA-positive or sero-positive RA risk in EIRA, NHS and Korean cohorts

Population	Smoking intensity (pack-years)		OR* (95%CI)	p-value
	0–10	>10		
EIRA				
Controls	1359 (70.97)	556 (29.03)	1.0	ref.
ACPA-positive RA	893 (54.22)	754 (45.78)	2.52 (2.18–2.92)	1.60E-34
NHS				
Controls	231 (64.17)	129 (35.83)	1.0	ref.
Sero-positive** RA	122 (53.28)	107 (46.72)	1.62 (1.14–2.29)	6.48E-03
Korean				
Controls	724 (98.50)	11 (1.50)	1.0	ref.
ACPA-positive RA	1226 (88.20)	164 (11.80)	4.73 (2.27–9.85)	3.28E-05

* OR were adjusted for age, gender and residential area in EIRA; for age in NHS; and for age and gender in Koreans.

** NHS cases defined as seropositive based on positive rheumatoid factor in charts since 1976, positive anti-CCP2 in charts since 1990, or positive anti-CCPs in subset of subjects with plasma samples.

Table 3
Association of the known RA-risk amino-acid positions with ACPA-positive or sero-positive RA susceptibility in the EIRA, NHS and Korean cohorts

HLA genes	Amino acid Positions	Number of residues	EIRA		NHS		Korean		All**	
			P_{omnibus}	Conditional* P_{omnibus}	P_{omnibus}	Conditional* P_{omnibus}	P_{omnibus}	Conditional* P_{omnibus}	P_{omnibus}	Conditional* P_{omnibus}
HLA-DRB1	11	6	3,43E-97	5,47E-22	1,03E-11	5,59E-04	1,41E-43	4,53E-29	7,17E-169	5,32E-58
HLA-DRB1	13	6	6,45E-95	3,82E-19	1,74E-11	4,04E-04	4,35E-42	8,83E-26	2,67E-166	1,15E-49
HLA-DRB1	71	4	1,38E-44	3,27E-05	2,76E-04	0,035	5,98E-06	0,282	4,12E-54	3,35E-08
HLA-DRB1	74	5	5,63E-19	9,44E-04	3,84E-03	0,266	1,36E-10	1,23E-03	8,49E-34	7,88E-13
HLA-B	9	3	1,68E-08	4,47E-03	0,397	0,209	2,03E-03	0,040	0,507	2,74E-03
HLA-DPB1	9	3	2,96E-09	1,03E-05	0,409	0,844	0,394	0,938	4,62E-09	1,67E-04

* The effect of each amino acid position was conditioned on other amino acid positions. HLA-A-DRB1 positions 11 and 13 were not conditioned on each other due to their tight correlation. All the analyses were adjusted for top 5 PCs from each population.

** For the combined dataset, a dummy variable indicating the three cohorts was additionally included in the logistic regression.

interaction effect between HLA-DRβ1 haplotype genetic risk score (GRS) and heavy smoking in ACPA-positive or sero-positive RA susceptibility in EIRA, NHS and Korean cohorts

Table 4

Population	Pack-year	GRS level [*]	Cases	Controls	OR (95%CI)	AP(95%CI) ^{**} and P
EIRA	0-10	Low	161 (19.05)	684 (80.95)	1.0 ref.	
	>10	Low	128 (31.30)	281 (68.70)	1.96 (1.49-2.57)	
	0-10	High	732 (52.03)	675 (47.97)	4.38 (3.51-5.46)	0.416 (0.306-0.526) p=1.16E-13
	>10	High	626 (69.48)	275 (30.52)	8.49 (6.62-10.89)	
NHS	0-10	Low	38 (25.50)	111 (74.50)	1.0 ref.	
	>10	Low	26 (27.37)	69 (72.63)	1.09 (0.60-1.99)	
	0-10	High	84 (41.18)	120 (58.82)	1.80 (1.11-2.92)	0.467 (0.187-0.748) p=1.07E-3
	>10	High	81 (57.45)	60 (42.55)	4.42 (2.59-7.57)	
Korean	0-10	Low	232 (39.19)	360 (60.81)	1.0 ref.	
	>10	Low	43 (84.31)	8 (15.69)	8.09 (3.71-17.63)	
	0-10	High	994 (73.20)	364 (26.80)	3.66 (2.92-4.59)	0.796 (0.535-1.058) p=2.41E-9
	>10	High	121 (97.58)	3 (2.42)	61.49 (19.05-198.48)	

^{*} HLA-DRβ1 Haplotype was constructed from the amino-acid positions 11, 13, 71 and 74. Haplotype GRS was calculated by weighing each haplotype on its reported RA-risk effect size. The calculated GRS were dichotomized at the median in controls of each population.

^{**} Dichotomized two main factors, GRS and smoking pack-year, were used in the AP model adjusting for HLA-B amino-acid position 9 GRS, HLA-DPB1 amino-acid position 9 GRS and five principal components.

Table 5

Additive interaction between HLA-DRβ1 amino acid genetic risk score (GRS) and smoking in ACPA-positive or sero-positive RA susceptibility in EIRA, NHS and Korean cohorts

Amino acid GRS	Pack-year	GRS level*	Cases	Controls	OR (95%CI)	AP(95%CI)** and P
EIRA						
DRB1_P11	0-10	Low	123 (17.65)	574 (82.35)	1.0 ref.	0.333 (0.183-0.484) p=1.45E-05
	>10	Low	98 (28.99)	240 (71.01)	1.92 (1.41-2.62)	
	0-10	High	770 (49.52)	785 (50.48)	2.30 (1.76-3.01)	
	>10	High	656 (67.49)	316 (32.51)	4.26 (3.16-5.73)	
DRB1_P13	0-10	Low	115 (17.14)	556 (82.86)	1.0 ref.	0.324 (0.172-0.477) p=2.98E-05
	>10	Low	95 (28.79)	235 (71.21)	2.00 (1.45-2.75)	
	0-10	High	778 (49.21)	803 (50.79)	2.39 (1.82-3.12)	
	>10	High	659 (67.24)	321 (32.76)	4.30 (3.18-5.80)	
DRB1_P71	0-10	Low	237 (27.56)	623 (72.44)	1.0 ref.	0.158 (-0.046-0.361) p=0.129
	>10	Low	232 (45.49)	278 (54.51)	2.13 (1.67-2.71)	
	0-10	High	656 (47.13)	736 (52.87)	1.37 (1.07-1.75)	
	>10	High	522 (65.25)	278 (34.75)	2.84 (2.13-3.80)	
DRB1_P74	0-10	Low	315 (31.50)	685 (68.50)	1.0 ref.	0.26 (0.065-0.456) p=9.09E-03
	>10	Low	233 (46.05)	273 (53.95)	1.80 (1.42-2.27)	
	0-10	High	578 (46.17)	674 (53.83)	1.16 (0.90-1.49)	
	>10	High	521 (64.80)	283 (35.20)	2.32 (1.76-3.07)	
B_P9	0-10	Low	358 (39.51)	548 (60.49)	1.0 ref.	0.031 (-0.206-0.268) p=0.796
	>10	Low	535 (39.75)	811 (60.25)	2.22 (1.76-2.80)	
	0-10	High	349 (58.46)	248 (41.54)	1.19 (0.97-1.45)	
	>10	High	405 (56.80)	308 (43.20)	2.78 (2.18-3.55)	
DRB1_P9	0-10	Low	293 (33.76)	575 (66.24)	1.0 ref.	0.244 (0.054-0.433) p=0.012
	>10	Low	222 (49.12)	230 (50.88)	1.93 (1.49-2.50)	
	0-10	High	600 (43.35)	784 (56.65)	1.31 (1.08-1.59)	
	>10	High	532 (62.00)	326 (38.00)	2.97 (2.38-3.71)	

Amino acid	GRS	Pack-year	GRS level*	Cases	Controls	OR (95%CI)	AP (95%CI)** and P
NHS							
DRB1_P11	0-10	Low	Low	32 (23.02)	107 (76.98)	1.0 ref.	
	>10	Low	Low	20 (22.99)	67 (77.01)	0.89 (0.46-1.73)	
	0-10	High	High	90 (42.06)	124 (57.94)	1.52 (0.85-2.70)	0.534 (0.257-0.812) p=1.60E-04
	>10	High	High	87 (58.39)	62 (41.61)	3.31 (1.81-6.04)	
DRB1_P13	0-10	Low	Low	30 (22.39)	104 (77.61)	1.0 ref.	
	>10	Low	Low	19 (22.35)	66 (77.65)	0.89 (0.45-1.76)	
	0-10	High	High	92 (42.01)	127 (57.99)	1.60 (0.89-2.86)	0.529 (0.254-0.805) p=1.64E-04
	>10	High	High	88 (58.28)	63 (41.72)	3.48 (1.88-6.44)	
DRB1_P71	0-10	Low	Low	53 (31.93)	113 (68.07)	1.0 ref.	
	>10	Low	Low	38 (36.19)	67 (63.81)	1.39 (0.78-2.47)	
	0-10	High	High	69 (36.90)	118 (63.10)	0.99 (0.55-1.77)	0.449 (0.043-0.855) p=0.030
	>10	High	High	69 (52.67)	62 (47.33)	2.19 (1.17-4.11)	
DRB1_P74	0-10	Low	Low	26 (22.22)	91 (77.78)	1.0 ref.	
	>10	Low	Low	32 (41.56)	45 (58.44)	2.54 (1.29-4.98)	
	0-10	High	High	96 (40.68)	140 (59.32)	1.55 (0.87-2.74)	-0.442 (-1.217-0.333) p=0.264
	>10	High	High	75 (47.17)	84 (52.83)	2.32 (1.23-4.38)	
B_P9	0-10	Low	Low	61 (36.53)	106 (63.47)	1.0 ref.	
	>10	Low	Low	49 (47.12)	55 (52.88)	1.76 (1.02-3.03)	
	0-10	High	High	61 (32.80)	125 (67.20)	0.92 (0.55-1.52)	-0.084 (-0.776-0.609) p=0.812
	>10	High	High	58 (43.94)	74 (56.06)	1.55 (0.89-2.70)	
DRB1_P9	0-10	Low	Low	47 (28.48)	118 (71.52)	1.0 ref.	
	>10	Low	Low	46 (47.42)	51 (52.58)	2.54 (1.43-4.53)	
	0-10	High	High	75 (39.89)	113 (60.11)	1.51 (0.93-2.45)	-0.580 (-1.429-0.269) p=0.180
	>10	High	High	61 (43.88)	78 (56.12)	1.85 (1.11-3.10)	
Korean							
DRB1_P11	0-10	Low	Low	223 (41.14)	319 (58.86)	1.0 ref.	
	>10	Low	Low	40 (85.11)	7 (14.89)	8.76 (3.80-20.19)	0.709 (0.357-1.060) p=7.79E-05

Amino acid	GRS	Pack-year	GRS level*	Cases	Controls	OR (95%CI)	AP (95%CI)** and P
DRB1_P13	0-10	High	High	1003 (71.24)	405 (28.76)	2.87 (2.28-3.62)	0.721 (0.383-1.060) p=2.89E-05
		>10	High	124 (96.88)	4 (3.13)	44.12 (15.70-123.96)	
	0-10	Low	Low	266 (43.68)	343 (56.32)	1.0 ref.	
		>10	Low	37 (84.09)	7 (15.91)	6.88 (2.99-15.83)	
	0-10	High	High	960 (71.59)	381 (28.41)	2.48 (1.98-3.11)	
		>10	High	127 (96.95)	4 (3.05)	34.30 (12.40-94.92)	
	0-10	Low	Low	403 (53.03)	357 (46.97)	1.0 ref.	
		>10	Low	65 (89.04)	8 (10.96)	7.72 (3.54-16.83)	
DRB1_P71	0-10	High	High	823 (69.16)	367 (30.84)	1.17 (0.91-1.50)	0.588 (0.016-1.160) p=0.044
		>10	High	99 (97.06)	3 (2.94)	19.22 (5.75-64.20)	
	0-10	Low	Low	493 (57.73)	361 (42.27)	1.0 ref.	
		>10	Low	63 (94.03)	4 (5.97)	12.55 (4.42-35.65)	
DRB1_P74	0-10	High	High	733 (66.88)	363 (33.12)	1.17 (0.93-1.47)	-0.290 (-1.964-1.384) p=0.734
		>10	High	101 (93.52)	7 (6.48)	7.93 (3.44-18.27)	
	0-10	Low	Low	563 (57.16)	422 (42.84)	1.0 ref.	
		>10	Low	75 (90.36)	8 (9.64)	7.75 (3.56-16.86)	
B_P9***	0-10	High	High	663 (68.70)	302 (31.30)	1.37 (1.12-1.67)	0.628 (0.116-1.139) p=0.016
		>10	High	89 (96.74)	3 (3.26)	20.74 (6.39-67.32)	
	0-10	Low	Low	602 (63.57)	345 (36.43)	1.0 ref.	
		>10	Low	82 (97.62)	2 (2.38)	30.55 (7.18-130.07)	
DRB1_P9	0-10	High	High	624 (62.21)	379 (37.79)	0.99 (0.81-1.22)	NC
	>10	High	High	82 (90.11)	9 (9.89)	6.85 (3.28-14.33)	

* Amino acid GRS was calculated by weighing each amino acid on its reported RA-risk effect size. The calculated GRS were dichotomized at the median in controls of each population (High GRS group was defined by GRS median GRS).

** Dichotomized two main factors, GRS and smoking pack-year, were used in the AP model adjusting for the top 5 principal components and GRS from other amino acid positions.

*** The high GRS group for HLA-B amino-acid position 9 in the Korean population was defined by GRS > median GRS because the median GRS is equivalent to the lowest GRS.

Significant P values after the Bonferroni correction are in bold.