Gene–environment interaction between *HLA-DRB1* shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis

E W Karlson,[1] S-C Chang,[2] J Cui,[1] L B Chibnik,[1] P A Fraser,[1,3,4] I De Vivo,[2,5] K H Costenbader[1]

► Additional data (supplementary information) are published online only at http://ard.bmj.com/content/vol69/issue1

[1] Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [2] Harvard School of Public Health, Boston, Massachusetts, USA; [3] Immune Disease Institute, Boston, Massachusetts, USA; [4] Genzyme Corporation, Boston, Massachusetts, USA; [5] Channing Laboratory, Brigham and Women's Hospital, Boston, Massachusetts, USA

Correspondence to: E W Karlson, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA; ekarlson@partners.org

Accepted 2 January 2009
Published Online First 16 January 2009

ABSTRACT

Background: Previous studies have reported an interaction between ever cigarette smoking and the presence of the human leukocyte antigen (*HLA*)-*DRB1* shared epitope (SE) genotype and rheumatoid arthritis (RA) risk. To address the effect of dosage, a case-control study nested within two prospective cohorts to determine the interaction between heavy smoking and the *HLA-SE* was conducted.

Methods: Blood was obtained from 32 826 women in the Nurses' Health Study and 29 611 women in the Nurses' Health Study II. Incident RA diagnoses were validated by chart review. Controls were matched for age, menopausal status and postmenopausal hormone use. High-resolution *HLA-DRB1* genotyping was performed for SE alleles. *HLA-SE*, smoking, *HLA-SE** smoking interactions and RA risk, were assessed using conditional logistic regression models, adjusted for age and reproductive factors. Additive and multiplicative interactions were tested.

Results: In all, 439 Caucasian matched pairs were included. Mean age at RA diagnosis was 55.2 years; 62% of cases were seropositive. A modest additive interaction was observed between ever smoking and *HLA-SE* in seropositive RA risk. A strong additive interaction (attributable proportion due to interaction (AP) = 0.50; $p<0.001$) and significant multiplicative interaction ($p = 0.05$) were found between heavy smoking (>10 pack-years) and any *HLA-SE* in seropositive RA risk. The highest risk was in heavy smokers with double copy *HLA-SE* (odds ratio (OR) 7.47, 95% CI 2.77 to 20.11).

Conclusions: A strong gene–environment interaction was observed between *HLA-SE* and smoking when stratifying by pack-years of smoking rather than by ever smoking. Future studies should assess cumulative exposure to cigarette smoke when testing for gene–smoking interactions.

Rheumatoid arthritis (RA), an autoimmune disease of unknown aetiology, affects approximately 1% of the adult population.[1] Genetic and environmental factors are thought to interact in RA development. Epidemiological research has demonstrated a strong association between cigarette smoking and RA risk.[2–14] In the Nurses' Health Study (NHS), we have found that RA risk is significantly elevated among women with >10 pack-years of smoking, and a strong dose-response exists.[12 14]

The strongest genetic risk factor for RA is found within the human leukocyte antigen (*HLA*) complex (or major histocompatibility complex antigen (MHC)). Within the *HLA* class II region, multiple *HLA-DRB1* alleles are associated with RA.[15–17] In individuals of European ancestry, the associated *HLA-DRB1* alleles share a region of sequence similarity or "shared epitope" (SE) at amino acid positions 70–74 in the third hypervariable region of the *HLA-DRB1* molecule.[18] Smoking and HLA shared epitope (*HLA-SE*) genotypes interact to increase risk of seropositive but not seronegative RA in several studies.[13 19–21] However, the dose effect aspects of this interaction have not been studied. One study of three North American RA cohorts[22] did not demonstrate a significant interaction between ever smoking and *HLA-SE* in predicting anti-cyclic citrullinated protein (CCP) antibodies or rheumatoid factor (RF) among RA cases.[19]

We studied the interaction between *HLA-SE* alleles and smoking dose among women in a case-control study nested within two large prospective cohort studies, the Nurses' Health Studies. We aimed to determine whether heavier smoking was associated with a stronger gene–environment interaction than was ever smoking.

METHODS

Study population

The NHS is a prospective cohort of 121 700 female nurses, aged between 30–55 years in 1976. From 1989 to 1990, 32 826 (27%) NHS participants ages 43–70 provided blood samples. The Nurses' Health Study II (NHSII) is a similar cohort, with 116 608 female nurses aged between 25–42 in 1989. Between 1996 and 1999, 29 611 (25%) of the women participating in NHSII cohort, aged 32–52 at that time, provided blood samples. The demographics and exposure characteristics of participants who provided blood samples are similar to those of the overall cohorts.[23 24] All aspects of this study were approved by the Partners' HealthCare Institutional Review Board.

Identification of rheumatoid arthritis

As previously described,[14] we confirmed self-reports of RA based on presence of RA symptoms on a connective tissue disease screening questionnaire (CSQ),[25] and, medical record review for four or more of the seven American College of Rheumatology (ACR) classification criteria for RA.[26] We included a small number of subjects (n = 14) with agreement by two rheumatologist reviewers on diagnosis of RA, three documented ACR criteria for RA and a diagnosis of RA by their

doctor. The response rate to requests for the CSQ among RA self-reports was 77%, and 96% to requests for medical records.

Population for analysis

For cases and controls, we excluded women who reported any cancer (except non-melanoma skin cancer) at baseline or during follow-up. Each participant with RA was matched by year of birth, race/ethnicity, menopausal status and postmenopausal hormone use to a single healthy woman in the same cohort without RA. To minimise potential population stratification, we limited the analyses to Caucasian women.

DNA extraction and amplification

DNA was extracted from buffy coats and processed via the QIAmp (Qiagen, Chatsworth, California, USA) 96-spin blood kit protocol as previously described.[27] All genomic DNA samples had an aliquot put through a whole genome amplification protocol using the GenomPhi DNA amplification kit (GE Healthcare, Piscataway, New Jersey, USA) to yield high quality DNA sufficient for HLA genotyping.

Seropositive phenotyping

We collected information on RF from medical records reviewed from the date of RA diagnosis. We did not have records from later in the disease course, or information on CCP as cases were diagnosed prior to its widespread use. For a subset of 180 NHS and 41 NHSII RA cases, plasma samples were collected in 1989 and a second set of plasma samples collected in 2000. In all, 98 samples were collected before RA diagnosis (incident samples) and 123 were collected after diagnosis (prevalent samples). We used the DIASTAT CCP (Axis-Shield Diagnostics, Dundee, UK) second-generation test, a semiquantitative/qualitative ELISA for detection of IgG CCP antibodies. A CCP antibody titre >5 U/ml was considered positive according to the manufacturer's established threshold. Since prior work from Sweden has demonstrated gene–environment interactions between *HLA-SE* and smoking for RF-positive[13] and CCP-positive RA,[19] we created a combination phenotype based on RF results from the medical record supplemented by CCP results from plasma samples where available, as ''ever seropositive'' versus ''never seropositive''.

HLA-SE determination

Low-resolution *HLA-DRB1* genotyping was performed using polymerase chain reaction with sequence specific primers (PCR-SSP) using OLERUP SSP kits (Qiagen, West Chester, Pennsylvania, USA). We used primers to amplify DNA samples that contained sequences for *HLA-DRB1*04, *01,*10* and **14*, along with consensus primers and appropriate positive and negative control samples. For samples with positive two-digit *HLA* signals, sequence specific primers were used for high-resolution four-digit shared epitope allele detection of *DRB1*0401, *0404, *0405, *0101, *0102, *1402* and **1001*. OLERUP SSP computer software (Qiagen) was used to determine four-digit *HLA* types.

Covariates

Information was collected via prospective biennial subject questionnaires regarding diseases, lifestyle and health practices. Reproductive covariates were chosen based on associations between reproductive factors and the RA risk in this cohort.[28] Lifetime smoking history was collected at baseline and updated data concerning current smoking and number of cigarettes smoked a day were collected every 2 years. Data on smoking, parity, total duration of breast feeding, menopausal status and postmenopausal hormone use were selected from the cycle prior to the RA diagnosis date (or index date in controls). Smoking was categorised as: (1) never versus ever and (2) pack-years of smoking (product of years of smoking and packs of cigarettes per day). Pack-years were dichotomised as never or light smoking versus heavy smoking $\leqslant 10$ vs >10 pack-years based on epidemiological data from this cohort that demonstrate increased RA risk with >10 pack-years of smoking.[14] We further investigated three smoking categories (never, past, current) and three categories of pack-years ($\leqslant 10$, 10–20 and >20 pack-years).

Statistical methods

We verified the Hardy–Weinberg equilibrium for each genotype among controls in each nested case-control dataset. We calculated means with standard deviation and medians with range for continuous covariates stratified by cohort and case/control status. For categorical covariates, we calculated frequencies and percentages. SAS V.9.1 was used for all analyses (SAS Institute, Cary, North Carolina, USA). Distributions for *HLA-SE* among cases and controls were compared using the χ^2 test of independence. Conditional logistic regression analyses, conditioning on matching factors and adjusting for age at menarche, menstrual regularity, parity, breastfeeding duration, menopausal status and postmenopausal hormone use, tested the association between *HLA-SE* alleles and RA risk in a general model and in a dominant model. Unconditional logistic regression analyses, adjusting for matching factors and covariates, were used to examine the risks of seropositive and seronegative RA.

Analyses of interaction

We used an additive models of interaction based on disease rates connected to the ''pie model''.[29] Rothman showed that independent risk factors adhere to an additive model and that biological interaction results in departure from additivity of the disease rates (see supplementary material). To test for additive interactions we followed the methods outlined by Lundberg[30] and Andersson,[31] using a 2×2 factorial design to calculate the attributable proportion due to interaction (AP), the relative excess risk due to interaction (RERI) and the synergy index (SI). A p value of <0.05 for AP was considered as departure from an additive model of association. For models where *HLA-SE*, pack-years of smoking and smoking status were categorised into three categories, we calculated indices of additive interaction for each stratum of exposure compared to the referent category of non-exposure. The 95% confidence intervals (CIs) were calculated using the delta method as described previously,[32] which is a straightforward Taylor expansion of the variances and covariances to derive a probability distribution. Multiplicative interaction was assessed by adding an interaction variable (*HLA-SE** smoking) to the regression models. A p value of <0.05 was considered as evidence for departure from a multiplicative model of association.

RESULTS

A total of 439 pairs of Caucasian women, each pair being 1 RA case and a matched control, were included. The cases in the NHS had a mean (SD) age of 56.7 (9.4) years, compared to 43.1 (5.1) years in the younger NHSII cohort, due to the different ages targeted for enrolment in each of the cohorts (table 1). Otherwise, the cases were similar in terms of RA characteristics with 61% seropositive RA in NHS and 65% seropositive in NHSII.

Table 1 Characteristics of rheumatoid arthritis (RA) cases and matched controls in the Nurses' Health Study (NHS; 1976–2002) and the Nurses' Health Study II (NHSII; 1989–2003)

	Nurses' Health Study (390 matched pairs)		Nurses' Health Study II (49 matched pairs)	
	RA cases	Controls	RA cases	Controls
Matching factors				
Age, mean (SD)	56.7 (9.4)	57.0 (9.3)	43.1 (5.1)	43.1 (5.1)
Postmenopausal (%)	265 (68.0%)	258 (66.2%)	14 (28.6%)	14 (28.6%)
Current PMH use (%)*	114 (37.1%)	106 (35.9%)	11 (68.8%)	12 (75.0%)
Other characteristics				
Ever cigarette smokers (%)	246 (63.1%)	223 (57.2%)	19 (38.8%)	22 (44.9%)
Pack-years among smokers, mean (SD)	23.9 (16.7)	24.1 (20.3)	16.2 (9.1)	10.8 (5.5)
Parous (%)	355 (91.0%)	368 (94.4%)	46 (93.9%)	42 (85.7%)
Breastfed babies ⩾12 months total (%)†	46 (13.7%)	65 (18.4%)	18 (42.9%)	18 (45.0%)
RA features				
Mean age at diagnosis, (SD)	56.7 (9.4)	–	43.1 (5.1)	–
Rheumatoid factor positive (%)	224 (57.4%)	–	27 (55.1%)	–
Anti-CCP‡ positive	86 (47.8%)	–	19 (46.3%)	–
Seropositive (%)	238 (61.0%)		32 (65.3%)	
Rheumatoid nodules, (%)	50 (12.8%)	–	6 (12.2%)	–
Radiographic changes, (%)	118 (30.3%)	–	17 (34.7%)	–
Diagnosed by a member of ACR, (%)	322 (84.1%)	–	47 (95.9%)	–

*Percentage is calculated among postmenopausal women or parous women, with unknown/missing group excluded; for the rest of the variables, percentage was calculated with missing category included.
†Among parous women.
‡Anti-cyclic citrullinated peptide (CCP) antibodies assayed in subset of cases (NHS = 180, NHSII = 41) with stored blood samples at collected at different points with respect to RA onset, up to 12 years prior to onset or after diagnosis.
ACR, American College of Rheumatology; PMH, postmenopausal hormone.

Table 1 shows the distribution of covariates for the RA cases and their matched controls at the time of RA diagnosis (or index date for the controls). A higher proportion of RA cases and controls were postmenopausal at RA diagnosis in NHS compared to NHSII cohorts. In NHSII a slightly higher percentage of women with RA were parous compared to their matched controls (93.9% and 85.7%), but not in the NHS cohort (91.0% of RA cases and 94.4% of controls).

HLA-SE genotype distributions did not deviate from Hardy–Weinberg equilibrium. Overall, genotyping call rates were 98.5% for *HLA-SE*. The frequency of the *HLA-SE* was significantly higher among RA cases than controls (χ^2 with 1 degree of freedom, $p<0.001$ for pooled NHS/NHSII cohorts). In all, 49 (12.8%) NHS cases had 2 copies of the HLA SE allele as compared to 24 (6.3%) controls ($p<0.001$). Similar results were seen in NHSII participants with nine (18.4%) cases having two copies of the *HLA-SE* allele versus one (2.1%) in the controls ($p = 0.03$). The most common *HLA-SE* alleles in RA cases were *0401* (13.9%), *0404* (5.4%) and *0101* (9%).

Table 2 includes the results of conditional logistic regression analyses for RA risk associated with *HLA-SE* for all RA and from unconditional logistic regression analyses stratified by seropositivity. The adjusted model includes pack-years of cigarette smoking, age at menarche, regularity of menses, parity, breast feeding, menopausal status and postmenopausal hormone use. RA risk associated with a single copy of *HLA-SE* was elevated (odds ratio (OR) 1.60, 95% CI 1.16 to 2.22) and with a double copy was markedly elevated (OR 3.78, 95% CI 2.13 to 6.71). These effects of *HLA-SE* were limited to seropositive RA (double *HLA-SE* OR 4.41, 95% CI 2.53 to 7.68) with no significant association with seronegative RA.

Interaction results

Table 3 shows the results of analyses in which we tested for additive and multiplicative interactions between *HLA-SE* and smoking, categorised as ever/never smoking or dichotomised at ⩽10 or >10 pack-years of smoking. There was a 2.14-fold increased risk of RA (95% CI 1.39 to 3.29) for ever smokers who carried any *HLA-SE* compared to the referent group, never smokers with no *HLA-SE*, however, there was no evidence for gene–environment interaction. There was a modest additive but not multiplicative gene–environment interaction between ever smoking and the presence of the *HLA-SE* allele, with the proportion of risk due to additive interaction (AP) of 0.38 (95% CI 0.05 to 0.70, $p = 0.02$) for seropositive RA. In contrast, a 2.75-fold elevated risk of RA (95% CI 1.75 to 4.31) and a 3.6-fold elevated risk of seropositive RA (95% CI 2.26 to 5.78) were observed among heavy smokers (>10 pack-years) with any *HLA-SE* compared to the referent group (⩽10 pack-years without *HLA-SE*). We observed a significant additive, but not multiplicative, interaction between heavy cigarette smoking and the presence of the *HLA-SE* allele, with the proportion of risk due to additive interaction (AP) of 0.39 (95% CI 0.08 to 0.69, $p = 0.01$) for RA. A stronger additive interaction between heavy cigarette smoking and *HLA-SE* allele was observed for seropositive RA, with AP of 0.50 (95% CI 0.24 to 0.77, $p<0.002$), and a significant multiplicative interaction term ($p = 0.05$).

When stratified by number of copies of *HLA-SE* there was strong evidence for increasing risk of RA with each copy of *HLA-SE* among heavy smokers, with a 6.6-fold increased risk of RA, (95% CI 2.49 to 17.46), for heavy smokers with 2 copies of *HLA-SE* as compared to the reference group, with evidence for additive interaction (table 4). The strongest evidence for additive interaction was for *HLA-SE* with heavy smoking in seropositive RA (OR 7.47, 95% CI 2.77 to 20.11), with a borderline multiplicative interaction ($p = 0.07$). Among *HLA-SE* subtypes, the only interaction was between the *0401* allele and heavy smoking, with evidence for additive interaction ($p = 0.005$) for all RA. The *HLA-SE* subtypes of *0401* and *0101* demonstrated significant additive interaction with heavy smoking ($p = 0.003$ and 0.01, respectively) for seropositive RA. There was no evidence for multiplicative interactions between *HLA-SE* subtypes and heavy smoking (data not shown).

Table 2 Association between *HLA-SE* and rheumatoid arthritis (RA) risk in the Nurses' Health Studies with stratification by serological status

Model	*HLA-SE*	All RA, unadjusted* OR (95% CI)	All RA, adjusted† OR (95% CI)	Seropositive RA, adjusted‡ OR (95% CI)	Seronegative RA, adjusted‡ OR (95% CI)
General	None	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)
	Single SE	1.54 (1.13 to 2.10)	1.60 (1.16 to 2.22)	1.88 (1.34 to 2.65)	1.07 (0.72 to 1.59)
	Double SE	3.19 (1.88 to 5.43)	3.78 (2.13 to 6.71)	4.41 (2.53 to 7.68)	1.55 (0.75 to 3.21)
Dominant	None	1.0 (reference)	1.0 (reference)	1.0 (reference)	1.0 (reference)
	Any SE	1.80 (1.35 to 2.41)	1.90 (1.40 to 2.58)	2.25 (1.64 to 3.10)	1.14 (0.78 to 1.65)

*Conditional logistic regression.
†conditional logistic regression, adjusting for smoking (pack-years), parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use at diagnosis.
‡Unconditional logistic regression adjusting for year of birth, smoking (pack-years), parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use at blood draw and at diagnosis.
HLA-SE, human leukocyte antigen shared epitope; OR, odds ratio.

When stratifying pack-years into three categories, the highest odds for RA were in the 10–20 pack-year and *HLA-SE* group in all RA and seropositive RA analyses (table 5). Comparing each stratum to the referent demonstrated significant additive interactions for 10–20 pack-year stratum ($p = 0.008$) for all RA and for 10–20 ($p = 0.003$) and for >20 year strata ($p = 0.002$) for seropositive RA with a borderline multiplicative interaction ($p = 0.09$). When stratifying smoking status into never, past, or current (table 6), there was little evidence for interactions except a modest additive interaction between past smoking and *HLA-SE* for seropositive RA ($p = 0.04$).

DISCUSSION

In this nested case-control study of women, we demonstrate significant additive and multiplicative interaction between the *HLA-DRB1* shared epitope and heavy cigarette smoking of at least 10 pack-years. The observed interaction between *HLA-SE* and smoking was strongest for seropositive RA with little evidence for association with seronegative RA. Evidence for interaction was less evident when smoking status was analysed as never/ever smoking or never/past/current, suggesting the importance of considering cumulative "dose" of smoking when testing for gene–environment interaction in RA.

Interactions between *HLA-SE* alleles and smoking in RA risk have been demonstrated in several large epidemiological studies. In the Epidemiologic Investigations in Rheumatoid Arthritis (EIRA) study, a strong additive interaction between *HLA-SE* and smoking was demonstrated for RF-positive and anti-CCP-positive RA, but not for seronegative RA[13 19]; a 21-fold increased risk of CCP-positive RA was observed among smokers carrying a double

Table 3 Gene–environment interaction of human leukocyte antigen shared epitope (*HLA-SE*) and smoking in the Nurses' Health Studies

HLA-SE	Smoking status	All RA	Seropositive RA	Seronegative RA
		OR (95% CI)*	OR (95% CI)†‡	OR (95% CI)†§¶
None	Never	1.00 (reference)	1.00 (reference)	1.00 (reference)
None	Ever	0.99 (0.65 to 1.51)	1.05 (0.66 to 1.68)	1.04 (0.64 to 1.69)
Any	Never	1.65 (1.05 to 2.61)	1.78 (1.08 to 2.95)	1.18 (0.67 to 2.08)
Any	Ever	2.14 (1.39 to 3.29)	2.94 (1.84 to 4.69)	1.23 (0.72 to 2.10)
Additive interaction:				
AP (95% CI)		0.23 (−0.14 to 0.61)	0.38 (0.05 to 0.70)	0.004 (−0.69 to 0.70)
p For interaction		$p_{add} = 0.23$	$p_{add} = 0.02$	$p_{add} = 0.99$
RERI		0.49 (−0.35 to 1.33)	1.10 (−0.02 to 2.22)	0.005 (−0.85 to 0.86)
Synergy Index		1.77 (0.50 to 6.25)	2.32 (0.72 to 7.47)	1.02 (0.02 to 47.34)
Multiplicative interaction:				
p For interaction		$p_{multi} = 0.38$	$p_{multi} = 0.18$	$p_{multi} = 1.00$
		OR (95% CI)*	OR (95% CI)†‡	OR (95% CI)
None	⩽10 pack-years smoking	1.00 (reference)	1.00 (reference)	1.00 (reference)
None	>10 pack-years smoking	1.10 (0.73 to 1.67)	1.08 (0.67 to 1.72)	1.16 (0.72 to 1.89)
Any	⩽10 pack-years smoking	1.58 (1.08 to 2.33)	1.72 (1.13 to 2.61)	1.09 (0.68 to 1.77)
Any	>10 pack-years smoking	2.75 (1.75 to 4.31)	3.61 (2.26 to 5.78)	1.45 (0.81 to 2.56)
Additive interaction:				
AP (95% CI)		0.39 (0.08 to 0.69)	0.50 (0.24 to 0.77)	0.13 (−0.49 to 0.75)
p For interaction		$p_{add} = 0.01$	$p_{add} < 0.001$	$p_{add} = 0.68$
RERI		1.06 (−0.06 to 2.17)	1.82 (0.35 to 3.30)	0.19 (−0.75 to 1.13)
Synergy index		2.54 (0.81 to 7.97)	3.30 (1.06 to 10.33)	1.74 (0.07 to 42.22)
Multiplicative interaction:				
p For interaction		$p_{multi} = 0.14$	$p_{multi} = 0.05$	$p_{multi} = 0.74$

p_{add} is the p value for the attributable proportion due to interaction, p_{multi} is the p value for multiplicative interaction term between binary smoking variable and binary *HLA-SE* genotype with 1 degree of freedom.
*Conditional logistic regression, adjusting for parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use.
†Unconditional logistic regression adjusting for year of birth, parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use.
‡Seropositive RA cases and all controls; §seronegative RA cases and all controls.
AP, attributable proportion due to interaction, an index of additive interactions between binary smoking variable and binary *HLA-SE* genotype; RA, rheumatoid arthritis; RERI, relative excess risk due to interaction.

Table 4 Gene–environment interaction between *HLA-SE* (none, single and double copies) and heavy smoking in all rheumatoid arthritis (RA) groups and in groups stratified by serological status in the Nurses' Health Studies

HLA-SE	Pack-years smoking	All RA			Seropositive RA		Seronegative RA	
		Cases/controls	OR* (95% CI)	OR† (95% CI)	Cases/controls	OR‡ (95% CI)	Cases/controls	OR§ (95% CI)
0	⩽10	119/161	1.00 (reference)	1.00 (reference)	64/161	1.00 (reference)	55/161	1.00 (reference)
1	⩽10	86/94	1.26 (0.85 to 1.88)	1.33 (0.88 to 2.00)	52/94	1.36 (0.87 to 2.15)	34/94	1.08 (0.65 to 1.79)
2	⩽10	32/19	2.61 (1.37 to 4.97)	2.96 (1.51 to 5.80)	25/19	3.56 (1.81 to 7.02)	7/19	1.21 (0.47 to 3.08)
0	>10	86/103	1.13 (0.75 to 1.69)	1.07 (0.70 to 1.62)	44/103	1.08 (0.68 to 1.73)	42/103	1.16 (0.72 to 1.89)
1	>10	78/46	2.26 (1.42 to 3.62)	2.24 (1.38 to 3.62)	57/46	3.14 (1.91 to 5.17)	21/46	1.25 (0.67 to 2.32)
2	>10	24/6	5.43 (2.14 to 13.77)	6.60 (2.49 to 17.46)	18/6	7.47 (2.77 to 20.11)	6/6	2.97 (0.90 to 9.79)
Additive:								
AP1 (95% CI)			0.39 (0.05 to 0.72)	0.38 (0.03 to 0.73)		0.54 (0.26 to 0.82)		0.01 (−0.75 to 0.77)
p For interaction			0.02	0.04	<0.001		0.98	
RERI			0.88 (−0.11 to 1.86)	0.84 (−0.16 to 1.85)		1.70 (0.29 to 3.11)		0.01 (−0.94 to 0.97)
Synergy index			3.26 (0.52 to 20.25)	3.15 (0.48 to 20.54)		4.81 (0.71 to 32.54)		1.06 (0.02 to 54.35)
AP2 (95% CI)			0.50 (−0.03 to 1.02)	0.54 (0.05 to 1.04)		0.51 (−0.03 to 1.06)		0.54 (−0.12 to 1.20)
p For interaction			0.06	0.03	0.07		0.11	
RERI			2.70 (−2.34 to 7.73)	3.57 (−2.75 to 9.90)		3.83 (−3.59 to 11.24)		1.60 (−2.00 to 5.20)
Synergy index			2.56 (0.63 to 10.40)	2.77 (0.69 to 11.06)		2.45 (0.62 to 9.72)		5.32 (0.11 to 252.71)
Multiplicative:								
p For interaction§			0.13	0.20	0.07		0.52	

*Conditional logistic regression.
†Conditional logistic regression, adjusting for parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use at diagnosis.
‡Unconditional logistic regression adjusting for year of birth, parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use at blood draw and at diagnosis.
§p Value calculated from HLA-SE×pack-years interaction term.
AP, attributable proportion due to interaction, an index of additive interactions between binary smoking variable and binary *HLA-SE* genotype; RERI, relative excess risk due to interaction.

copy of *HLA-SE*.[19] This finding suggests that these two important risk factors may interact along one or more biological pathways.[29] The statistical interaction between smoking and *HLA-SE* alleles in CCP-positive RA is consistent with the hypothesis that cigarette smoking modulates the immunogenicity of citrulline and related peptides in individuals with specific HLA alleles. This hypothesis has been strengthened by demonstration that smoking can cause citrullination of peptides in lung macrophages[19] and is associated with an increased expression of peptidyl arginine deiminase 2 (PADI2) in bronchoalveolar cells.[33] In *HLA-DRB1 0401* transgenic mice, citrullination of certain peptides increases binding to HLA class II molecules with the SE, triggering immune responses to citrullinated peptides.[34]

Table 5 Gene–environment interaction between *HLA-SE* (none, any) and pack-years of smoking (⩽10, 10–20, >20) in all rheumatoid arthritis (RA) groups and in groups stratified by serological status in the Nurses' Health Studies

HLA-SE	Pack-years smoking	All RA			Seropositive RA		Seronegative RA	
		Cases/controls	OR* (95% CI)	OR† (95% CI)	Cases/controls	OR‡ (95% CI)	Cases/controls	OR‡ (95% CI)
None	⩽10	119/161	1.00 (reference)	1.00 (reference)	64/161	1.00 (reference)	55/161	1.00 (reference)
None	10 to 20	25/33	0.98 (0.54 to 1.77)	1.04 (0.56 to 1.93)	15/33	1.14 (0.57 to 2.26)	10/33	0.93 (0.43 to 2.04)
None	>20	61/70	1.22 (0.78 to 1.91)	1.13 (0.71 to 1.80)	29/70	1.04 (0.61 to 1.79)	32/70	1.27 (0.74 to 2.17)
Any	⩽10	118/113	1.49 (1.03 to 2.17)	1.58 (1.08 to 2.32)	77/113	1.72 (1.13 to 2.60)	41/113	1.09 (0.68 to 1.77)
Any	10 to 20	29/12	3.42 (1.60 to 7.28)	3.55 (1.62 to 7.78)	21/12	4.31 (1.97 to 9.42)	8/12	1.76 (0.67 to 4.64)
Any	>20	73/40	2.50 (1.54 to 4.07)	2.52 (1.53 to 4.14)	54/40	3.39 (2.02 to 5.69)	19/40	1.35 (0.71 to 2.57)
Additive:								
AP1 (95% CI)			0.57 (0.20 to 0.94)	0.54 (0.14 to 0.94)		0.57 (0.20 to 0.95)		0.42 (−0.29 to 1.13)
p For interaction			0.003	0.008	0.003		0.25	
RERI			1.94 (−0.61 to 4.49)	1.93 (−0.81 to 4.66)		2.46 (−0.81 to 5.72)		0.74 (−1.06 to 2.53)
Synergy index			5.10 (0.72 to 36.36)	4.09 (0.74 to 22.56)		3.89 (0.86 to 17.64)		31.55 (0.00 to 2.072E19)
Multiplicative:								
p For interaction			0.09	0.13	0.13		0.39	
Additive:								
AP2 (95% CI)			0.32 (−0.05 to 0.68)	0.32 (−0.05 to 0.70)		0.48 (0.17 to 0.79)		−0.01 (−0.80 to 0.78)
p For interaction			0.09	0.09	0.002		0.98	
RERI			0.79 (−0.36 to 1.93)	0.81 (−0.37 to 1.99)		1.63 (0.04 to 3.22)		−0.02 (−1.07 to 1.04)
Synergy index			2.10 (0.65 to 6.80)	2.15 (0.64 to 7.23)		3.16 (0.89 to 11.21)		0.96 (0.05 to 19.07)
Multiplicative:								
p For interaction			0.33	0.30	0.09		0.95	

*Conditional logistic regression.
†Conditional logistic regression, adjusting for smoking (pack-years), parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use at diagnosis.
‡Unconditional logistic regression adjusting for year of birth, smoking (pack-years), parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use at blood draw and at diagnosis.
AP, attributable proportion due to interaction, an index of additive interactions between binary smoking variable and binary *HLA-SE* genotype; RERI, relative excess risk due to interaction.

Table 6 Gene–environment interaction between *HLA-SE* (none, any) and smoking status (never, past, current) in all rheumatoid arthritis (RA) groups and in groups stratified by serological status in the Nurses' Health Studies

HLA-SE	Smoking status	All RA			Seropositive RA		Seronegative RA	
		Cases/controls	OR* (95% CI)	OR† (95% CI)	Cases/controls	OR‡ (95% CI)	Cases/controls	OR‡ (95% CI)
None	Never	81/111	1.00 (reference)	1.00 (reference)	42/111	1.00 (reference)	39/111	1.00 (reference)
None	Past	92/110	1.13 (0.74 to 1.73)	1.03 (0.66 to 1.61)	50/110	1.11 (0.68 to 1.83)	42/110	1.05 (0.62 to 1.77)
None	Current	32/46	0.90 (0.51 to 1.59)	0.90 (0.49 to 1.64)	15/46	0.90 (0.45 to 1.81)	17/46	1.03 (0.52 to 2.06)
Any	Never	85/77	1.60 (1.03 to 2.49)	1.67 (1.05 to 2.63)	54/77	1.78 (1.08 to 2.95)	31/11	1.18 (0.70 to 2.09)
Any	Past	99/60	2.25 (1.43 to 3.52)	2.23 (1.40 to 3.56)	71/60	3.03 (1.83 to 5.03)	28/60	1.28 (0.71 to 2.33)
Any	Current	38/27	1.89 (1.03 to 3.47)	1.91 (1.03 to 3.55)	27/27	2.71 (1.41 to 5.19)	11/27	1.10 (0.49 to 2.48)
Additive:								
AP1 (95% CI)			0.23 (−0.15 to 0.61)	0.24 (−0.15 to 0.63)		0.38 (0.03 to 0.73)		0.04 (−0.68 to 0.74)
p For interaction			0.23	0.22	0.04		0.91	
RERI			0.52 (−0.40 to 1.44)	0.54 (−0.41 to 1.48)		1.14 (−0.17 to 2.44)		0.06 (−1.31 to 1.08)
Synergy index			1.71 (0.54 to 5.40)	1.77 (0.52 to 6.08)		2.27 (0.71 to 7.27)		1.24 (0.02 to 67.47)
Multiplicative:								
p For interaction			0.48	0.41	0.24		0.93	
Additive:								
AP2 (95% CI)			0.21 (−0.35 to 0.77)	0.18 (−0.42 to 0.78)		0.38 (−0.10 to 0.85)		−0.10 (−1.22 to 1.02)
p For interaction			0.47	0.55	0.12		0.86	
RERI			0.39 (−0.80 to 1.59)	0.35 (−0.91 to 1.60)		1.02 (−0.66 to 2.70)		−0.11 (−0.88 to 0.99)
Synergy index			1.80 (0.26 to 12.58)	1.62 (0.25 to 10.36)		2.49 (0.46 to 13.63)		0.49 (0.00 to 2327.19)
Multiplicative:								
p For interaction			0.51	0.58	0.28		0.85	

*Conditional logistic regression.
†Conditional logistic regression, adjusting for parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use at diagnosis.
‡Unconditional logistic regression adjusting for year of birth, parity, breast feeding and menstrual irregularity, and age at menarche, menopausal status and postmenopausal hormone use at blood draw and at diagnosis.
AP, attributable proportion due to interaction, an index of additive interactions between binary smoking variable and binary *HLA-SE* genotype; RERI, relative excess risk due to interaction.

Evidence for gene–environment interaction between *HLA-SE* and smoking in seropositive RA risk was seen in patients with undifferentiated arthritis at the Leiden Early Arthritis Clinic. Among the participants who were *HLA-SE* positive, ever smoking significantly increased the OR for development of CCP-positive RA from 3.3 to 8.0 ($p = 0.002$ for multiplicative interaction)[21]; *HLA-SE* and ever smoking also increased the OR of CCP antibodies, with evidence for additive but not multiplicative interaction.[20 35] A Danish case-control study of *HLA-SE* interactions with RA epidemiological risk factors, did not demonstrate any significant multiplicative interaction for *HLA-SE** smoking, however, *HLA-SE* homozygotes had a 52-fold increased risk of CCP-positive RA compared to non-carrier/never smokers. Testing for additive interaction was not performed. In contrast, a significant *HLA-SE** smoking interaction was demonstrated in the Iowa Women's Health Study, an older Caucasian female cohort in which smoking was associated with increased risk of RA only among subjects who were *HLA-SE* negative, but not among subjects who were *HLA-SE* positive.[36] The reasons for this discrepancy are unknown, although it may relate to the older age at RA onset or the small sample size (116 cases) in that cohort. A case-only analysis combining data from three large US populations was unable to confirm an interaction between *HLA-SE* alleles and cigarette smoking in relation to presence of CCP antibodies but smoking status was defined only as never/ever smoking.[22] In one cohort with information on pack-years of smoking, there was an independent effect of smoking on the presence of CCP among heavy smokers (>20 pack-years).

Strengths of this study include the prospective collection of exposure information prior to the onset of RA, the detailed smoking data collected every 2 years and availability of high-resolution HLA genotyping. Limitations include the lack of data on CCP antibody status, as most cases were diagnosed prior to the widespread usage of this test, and absence of plasma samples for CCP testing in about half the cases. However, RF status was available from medical record reviews, and other gene–environment interaction studies demonstrate similar relationships for RF-positive and CCP-positive phenotypes.[13 19] The rate of seropositive RA in this study (60%) is similar to that reported from a large US registry study, the National Databank (n = 14 000) with patients recruited from rheumatology practices across the US.[37] Limited generalisability of NHS a concern, as the NHS cohorts are comprised of middle to older aged women with high educational levels and with primarily Caucasian heritage. However, an advantage of similar ethnic background in genetic studies is a lower potential for population stratification.

This study of gene–environment interactions in RA in a cohort of Caucasian US women demonstrates a significant additive and multiplicative interaction between the strongest genetic risk factor for RA, the *HLA-SE*, and heavy cigarette smoking >10 pack-years, the strongest environmental risk factor for RA, for seropositive but not seronegative RA. We demonstrate only additive interaction between smoking categorised as never/ever smoked or as three categories (never, past, current) with the *HLA-SE* and seropositive RA; however, if smoking is classified by dose (as pack-years), we demonstrate additive and multiplicative interaction for seropositive RA. This illustrates the importance of considering the dose effects of environmental and genetic factors in gene–environment interaction studies. Additionally, it lends evidence to the theory that seropositive and seronegative RA have different risk factors and pathogenic pathways.

Acknowledgements: The authors gratefully acknowledge the participants in the NHS for their continuing participation. The authors also thank Gideon Aweh, Karen Corsano, Wei-Zi Ding, Lingsheng Dong and Brendan Keenan for their technical assistance.

Funding: Supported by NIH grants R01 AR49880, CA87969, P60 AR047782, K24 AR0524-01 and BIRCWH K12 HD051959 (supported by NIMH, NIAID, NICHD and OD).

KHC is the recipient of an Arthritis Foundation/American College of Rheumatology Arthritis Investigator Award and a Katherine Swan Ginsburg Memorial Award.

Competing interests: None declared.

Ethics approval: Ethics approval was granted by the Partners Human Subjects Committee.

Provenance and peer review: Not commissioned; externally peer reviewed.

REFERENCES

1. **Gabriel SE,** Crowson CS, O'Fallon WM. The epidemiology of rheumatoid arthritis in Rochester, Minnesota, 1955-1985. *Arthritis Rheum* 1999;**42**:415–20.
2. **Vessey MP,** Villard-Mackintosh L, Yeates D. Oral contraceptives, cigarette smoking and other factors in relation to arthritis. *Contraception* 1987;**35**:457–64.
3. **Hernandez Avila M,** Liang MH, Willett WC, *et al*. Reproductive factors, smoking, and the risk for rheumatoid arthritis. *Epidemiology* 1990;**1**:285–91.
4. **Hazes JM,** Dijkmans BA, Vandenbroucke JP, *et al*. Lifestyle and the risk of rheumatoid arthritis: cigarette smoking and alcohol consumption. *Ann Rheum Dis* 1990;**49**:980–2.
5. **Heliovaara M,** Aho K, Aromaa A, *et al*. Smoking and risk of rheumatoid arthritis. *J Rheumatol* 1993;**20**:1830–5.
6. **Voigt LF,** Koepsell TD, Nelson JL, *et al*. Smoking, obesity, alcohol consumption, and the risk of rheumatoid arthritis. *Epidemiology* 1994;**5**:525–32.
7. **Symmons DP,** Bankhead CR, Harrison BJ, *et al*. Blood transfusion, smoking, and obesity as risk factors for the development of rheumatoid arthritis: results from a primary care-based incident case-control study in Norfolk, England. *Arthritis Rheum* 1997;**40**:1955–61.
8. **Karlson EW,** Lee IM, Cook NR, *et al*. A retrospective cohort study of cigarette smoking and risk of rheumatoid arthritis in female health professionals. *Arthritis Rheum* 1999;**42**:910–7.
9. **Uhlig T,** Hagen KB, Kvien TK. Current tobacco smoking, formal education, and the risk of rheumatoid arthritis. *J Rheumatol* 1999;**26**:47–54.
10. **Criswell LA,** Merlino LA, Cerhan JR, *et al*. Cigarette smoking and the risk of rheumatoid arthritis among postmenopausal women: results from the Iowa Women's Health Study. *Am J Med* 2002;**112**:465–71.
11. **Krishnan E,** Sokka T, Hannonen P. Smoking-gender interaction and risk for rheumatoid arthritis. *Arthritis Res Ther* 2003;**5**:R158–62.
12. **Stolt P,** Bengtsson C, Nordmark B, *et al*. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. *Ann Rheum Dis* 2003;**62**:835–41.
13. **Padyukov L,** Silva C, Stolt P, *et al*. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum* 2004;**50**:3085–92.
14. **Costenbader KH,** Feskanich D, Mandl LA, *et al*. Smoking intensity, duration, and cessation, and the risk of rheumatoid arthritis in women. *Am J Med* 2006;**119**:503e1–9.
15. **Jawaheer D,** Gregersen PK. Rheumatoid arthritis. The genetic components. *Rheum Dis Clin North Am* 2002;**28**:1–15.
16. **Newton JL,** Harney SM, Wordsworth BP, *et al*. A review of the MHC genetics of rheumatoid arthritis. *Genes Immun* 2004;**5**:151–7.
17. **Jawaheer D,** Seldin MF, Amos CI, *et al*. Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. *Arthritis Rheum* 2003;**48**:906–16.
18. **Gregersen PK,** Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987;**30**:1205–13.
19. **Klareskog L,** Stolt P, Lundberg K, *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 Rheum* 2006;**54**:38–46.
20. **Linn-Rasker SP,** van der Helm-van Mil AH, van Gaalen FA, *et al*. Smoking is a risk factor for anti-CCP antibodies only in rheumatoid arthritis patients who carry HLA-DRB1 shared epitope alleles. *Ann Rheum Dis* 2006;**65**:366–71.
21. **van der Helm-van Mil AH,** Verpoort KN, le Cessie S, *et al*. The HLA-DRB1 shared epitope alleles differ in the interaction with smoking and predisposition to antibodies to cyclic citrullinated peptide. *Arthritis Rheum* 2007;**56**:425–32.
22. **Lee HS,** Irigoyen P, Kern M, *et al*. Interaction between smoking, the shared epitope, and anti-cyclic citrullinated peptide: a mixed picture in three large North American rheumatoid arthritis cohorts. *Arthritis Rheum* 2007;**56**:1745–53.
23. **Hankinson SE,** Colditz GA, Hunter DJ, *et al*. Reproductive factors and family history of breast cancer in relation to plasma estrogen and prolactin levels in postmenopausal women in the Nurses' Health Study (United States). *Cancer Causes Control* 1995;**6**:217–24.
24. **Tworoger SS,** Sluss P, Hankinson SE. Association between plasma prolactin concentrations and risk of breast cancer among predominately premenopausal women. *Cancer Res* 2006;**66**:2476–82.
25. **Karlson EW,** Sanchez-Garo J, Wright EA, *et al*. A connective tissue disease screening questionnaire for population studies. *Ann Epidemiol* 1995;**5**:297–302.
26. **Arnett FC,** Edworthy SM, Bloch DA, *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;**31**:315–24.
27. **Costenbader KH,** Chang SC, De Vivo I, *et al*. PTPN22, PADI-4 and CTLA-4 genetic polymorphisms and risk of rheumatoid arthritis in two longitudinal cohort studies: evidence of gene-environment interactions with heavy cigarette smoking. *Arthritis Res Ther* 2008;**10**:R52.
28. **Karlson EW,** Mandl LA, Hankinson SE, *et al*. Do breast-feeding and other reproductive factors influence future risk of rheumatoid arthritis? Results from the Nurses' Health Study. *Arthritis Rheum* 2004;**50**:3458–67.
29. **Rothman KJ.** *Epidemiology. An introduction*. New York, USA: Oxford University Press, 2002.
30. **Lundberg M,** Fredlund P, Hallqvist J, *et al*. A SAS program calculating three measures of interaction with confidence intervals. *Epidemiology* 1996;**7**:655–6.
31. **Andersson T,** Alfredsson L, Kallberg H, *et al*. Calculating measures of biological interaction. *Eur J Epidemiol* 2005;**20**:575–9.
32. **Hosmer DW,** Lemeshow S. Confidence interval estimation of interaction. *Epidemiology* 1992;**3**:452–6.
33. **Makrygiannakis D,** Hermansson M, Ulfgren AK, *et al*. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Ann Rheum Dis* 2008;**67**:1488–92.
34. **Hill JA,** Southwood S, Sette A, *et al*. Cutting edge: the conversion of arginine to citrulline allows for a high-affinity peptide interaction with the rheumatoid arthritis-associated HLA-DRB1*0401 MHC class II molecule. *J Immunol* 2003;**171**:538–41.
35. **Costenbader KC,** Chibnik LB, Mandl LA, *et al*. Testing for gene-environment interaction. http://ard.bmj.com/cgi/eletters/65/3/366#484 (accessed 6 Jan 2006).
36. **Criswell LA,** Saag KG, Mikuls TR, *et al*. Smoking interacts with genetic risk factors in the development of rheumatoid arthritis among older Caucasian women. *Ann Rheum Dis* 2006;**65**:1163–7.
37. **Finckh A,** Choi HK, Wolfe F. Progression of radiographic joint damage in different eras: trends towards milder disease in rheumatoid arthritis are attributable to improved treatment. *Ann Rheum Dis* 2006;**65**:1192–7.