

A Gene–Environment Interaction Between Smoking and Shared Epitope Genes in HLA–DR Provides a High Risk of Seropositive Rheumatoid Arthritis

Leonid Padyukov,¹ Camilla Silva,² Patrik Stolt,² Lars Alfredsson,³ and Lars Klareskog,² for the Epidemiological Investigation of Rheumatoid Arthritis Study Group

Objective. The main genetic risk factor for rheumatoid arthritis (RA) is the shared epitope (SE) of HLA–DR, while smoking is an important environmental risk factor. We studied a potential gene–environment interaction between SE genes and smoking in the etiology of the 2 major subgroups of RA: rheumatoid factor (RF)–seropositive and RF–seronegative disease.

Methods. A population-based case–control study involving incident cases of RF–seropositive and RF–seronegative RA (858 cases and 1,048 controls) was performed in Sweden. Cases and controls were classified according to their cigarette smoking status and HLA–DRB1 genotypes. The relative risk of developing RA was calculated for different gene/smoking combinations and was compared with the relative risk in never smokers without SE genes.

Results. The relative risk of RF–seropositive RA was 2.8 (95% confidence interval [95% CI] 1.6–4.8) in never smokers with SE genes, 2.4 (95% CI 1.3–4.6) in current smokers without SE genes, and 7.5 (95% CI 4.2–13.1) in current smokers with SE genes. Smokers carrying double SE genes displayed a relative risk of

RF–seropositive RA of 15.7 (95% CI 7.2–34.2). The interaction between smoking and SE genes was significant, as measured by the attributable proportion due to interaction, which was 0.4 (95% CI 0.2–0.7) for smoking and any SE, and 0.6 (95% CI 0.4–0.9) for smoking and a double SE. Neither smoking nor SE genes nor the combination of these factors increased the risk of developing RF–seronegative RA.

Conclusion. The disease risk of RF–seropositive RA associated with one of the classic genetic risk factors for immune-mediated diseases (the SE of HLA–DR) is strongly influenced by the presence of an environmental factor (smoking) in the population at risk.

Rheumatoid arthritis (RA), similar to other multifactorial diseases, is believed to occur as a result of the interaction between genetic constitution and environmental triggers. However, as in most other complex diseases, few such interactions have been described, and it has been assumed that very large studies will be needed to describe significant gene–environment interactions in these diseases. Nevertheless, the few existing examples, with the best one being a study of coeliac disease (1), have demonstrated how the definition of such interactions may open the field for new etiologic and pathogenetic studies.

The current investigation was initiated with the purpose of studying the interaction between the major genetic risk factor so far defined for RA, the shared epitope (SE) of HLA–DR (2), and smoking, which appears to be an environmental factor of major importance (3–5). We also performed a separate analysis on the 2 major phenotypes that are known in the disease, i.e., rheumatoid factor (RF)–seropositive and RF–seronegative RA. No studies using a case–control design have previously been performed to address these issues,

Supported by the Swedish Medical Research Council, the Swedish Council for Working Life and Social Research, the AFA Insurance Company, the Swedish Rheumatism Association, King Gustav V's 80-Year Foundation, the Börje Dahlin Fund, and the Stockholm County Council.

¹Leonid Padyukov, MD, PhD: Karolinska Institutet, Stockholm, Sweden, and Mechnikov Research Institute, Moscow, Russia; ²Camilla Silva, BSc, Patrik Stolt, MD, Lars Klareskog, MD, PhD: Karolinska Institutet, Stockholm, Sweden; ³Lars Alfredsson, PhD: Karolinska Institutet, and Stockholm Center for Public Health, Stockholm County Council, Stockholm, Sweden.

Dr. Padyukov and Ms Silva contributed equally to this work. Address correspondence and reprint requests to Leonid Padyukov, MD, PhD, Rheumatology Unit, Department of Medicine, KI, CMM L8:O4, Karolinska Hospital, S-17176 Stockholm, Sweden. E-mail: leonid.padyukov@cmm.ki.se.

Submitted for publication December 16, 2003; accepted in revised form June 22, 2004.

but a recent study, which involved RA cases only, suggested that interactions between smoking and the SE may exist (6).

Our study took advantage of a nationwide program that was initiated in Sweden in 1996 for identification and early treatment of patients with recent-onset RA, from whom we have continuously collected extensive information on environmental exposures, including smoking, that occurred before the onset of the disease. In conjunction with the first clinic visit, we also subphenotyped the patients according to RF status, and collected blood samples for subsequent genotyping. In parallel to the identification of cases, we collected similar information from control individuals, who were randomly selected from the study base.

PATIENTS AND METHODS

Study design. This study was designed as a population-based case-control study involving incident cases of RA, which were derived from the population ages 18–70 years in a geographically defined area in the middle and southern parts of Sweden. The recruitment period for the cases and controls was May 1996 to February 2001. The general structure of the study has been extensively described in a previous report, which presented findings regarding the influence of smoking on the risk of developing RA (5).

Identification of cases and controls. A case was defined as a person in the study base who fulfilled the American College of Rheumatology (ACR; formerly, the American Rheumatism Association) 1987 criteria for the classification of RA (7) and who had never had this diagnosis before (newly diagnosed cases). All potential cases were examined and diagnosed by a rheumatologist who was located at the unit in which the case was entered. All hospital-based rheumatology units as well as most of the privately run rheumatology units (very few) in the study area participated in the study. RF status was determined locally using agglutination tests, in which the cutoff levels were selected to yield positive results in not more than 5% of healthy controls. Results were reported as RF-seropositive or RF-seronegative.

For each potential case, a control subject was randomly selected from among a stratified random sample of the study base, taking into consideration the subject's age, sex, and residential area. The selection of controls was conducted using the national population register, which is continuously updated (see ref. 7). If information on a control subject was incomplete, then another control was selected using the same principles. Initially, some of the centers also reported cases of arthritis that did not fulfill the ACR criteria for RA, which enabled investigations of undifferentiated arthritis. These subjects were eventually excluded from this study, but their controls were kept in order to improve the precision of the study.

Data collection. Cases and controls were asked to answer a questionnaire. The patients answered the questionnaires shortly after they had received their diagnosis. All questionnaires were supposed to be answered at home. Incom-

pletely answered questionnaires were completed by mail or by telephone conversations conducted by specifically trained individuals.

Of the 900 identified patients with RA, 858 (95%) completed the questionnaire (612 women and 246 men), and of these, 64% of the women and 66% of the men were RF-seropositive. The total number of identified controls was 1,263, and the overall response rate concerning completion of the questionnaire for these individuals was 83%, producing a total of 1,048 controls (736 women and 312 men). For each case, a blood sample was obtained at the clinical center at the time of the first visit. The blood samples from participating control individuals were obtained in local medical wards and were sent to us by mail. We received blood samples from 843 (98%) of all of the cases who answered the questionnaires and from 627 (60%) of the controls (448 women and 179 men).

All patients and control individuals consented to participate in the study after receiving written information. All aspects of the study were approved by the ethics committee of the Karolinska Institutet.

Definition of smoking habits. The questionnaire contains a wide spectrum of questions regarding demographic and reproductive factors, heredity, previous health, body weight and length, lifestyle factors, occupational exposures, and psychosocial and socioeconomic circumstances. The questions posed regarding smoking, which are described in detail elsewhere (5), permit stratification of cases and controls into categories such as current smokers and never smokers. Information was also obtained regarding the type of tobacco smoked. The few subjects who smoked a pipe or cigars were excluded, leaving a study group restricted to cigarette smokers to be compared with the group of never smokers. Former smokers of cigarettes were also excluded, thus restricting the analysis to a comparison of current smokers of cigarettes with those who had never smoked.

For each case, the time point at the initial appearance of symptoms indicative of RA was used as an estimation of the disease onset. The year in which this time point occurred was defined as the index year, which was also used as the index year for the corresponding control. Only data on smoking up to the index year have been analyzed in the present study. Thus, individuals who reported that they were regularly smoking during the index year were defined as current smokers, and individuals who reported that they had never smoked prior to or during the index year were defined as never smokers.

Genotyping. DNA from blood samples in EDTA was extracted by the salting-out method (8), and analysis of HLA-DRB1 genotypes was made using the sequence-specific primer-polymerase chain reaction method (DR low-resolution kit; Olerup SSP, Saltsjöbaden, Sweden) as previously described (9). Among the HLA-DRB1 genes, DRB1*01, DRB1*04, and DRB1*10 were defined as SE genes. Any genotype with a combination of 2 of these genes was considered to be a double SE genotype. At the beginning of the study, some patients (81 cases) were subtyped for identification of HLA-DRB1*01 and 04 alleles. We determined a frequency of DRB1*0101 of 89% and a frequency of DRB1*0401,*0404,*0405,*0408 alleles of 98%, and for practical reasons, we decided to restrict the genotyping to only DR low-resolution analysis.

Potential confounding factors. We adjusted the data for age, sex, and residential area according to the principle of

Table 1. Distribution of smoking status and other characteristics among rheumatoid arthritis cases and controls, by sex and information source*

	Women				Men			
	Cases		Controls		Cases		Controls	
	Questionnaire (n = 612)	Questionnaire + genetic data (n = 598)	Questionnaire (n = 736)	Questionnaire + genetic data (n = 448)	Questionnaire (n = 246)	Questionnaire + genetic data (n = 245)	Questionnaire (n = 312)	Questionnaire + genetic data (n = 179)
Age, mean ± SD year†	49.7 ± 13	49.7 ± 13.1	48.6 ± 13.3	50.3 ± 12.2	52.7 ± 11.6	52.7 ± 11.7	51.6 ± 12	53.8 ± 10.9
Smoking status								
Current smoker	174 (28)	170 (28)	173 (24)	100 (22)	61 (25)	60 (25)	63 (20)	31 (17)
Former smoker	166 (27)	164 (27)	179 (24)	118 (26)	87 (35)	87 (36)	105 (34)	70 (39)
Non-regular smoker	27 (4)	25 (4)	46 (6)	29 (7)	6 (2)	6 (2)	18 (6)	10 (6)
Non-cigarette smoker	40 (7)	39 (7)	45 (6)	28 (6)	38 (15)	37 (15)	35 (11)	20 (11)
Never smoker	205 (34)	201 (34)	292 (40)	173 (39)	54 (22)	54 (22)	91 (29)	48 (27)
Information missing	0	0	1 (0.1)	0	0	0	0	0
Blue collar worker	190 (31)	185 (31)	212 (29)	121 (27)	96 (39)	95 (39)	100 (32)	52 (29)
BMI ≥25 kg/m²	266 (43)	261 (44)	284 (39)	181 (40)	135 (55)	133 (55)	183 (59)	110 (61)
Married	448 (73)	437 (73)	573 (78)	342 (76)	208 (85)	206 (84)	251 (80)	148 (83)
Parity	501 (82)	491 (82)	597 (81)	378 (84)	—	—	—	—
Oral contraceptive use	379 (62)	369 (62)	480 (65)	293 (65)	—	—	—	—
Born in Sweden	545 (89)	533 (89)	626 (85)	387 (86)	220 (89)	218 (89)	278 (89)	159 (89)

* Except where indicated otherwise, values are the no. (%). Genetic data were derived from blood samples. BMI = body mass index.

† Age when attending the Epidemiological Investigation of Rheumatoid Arthritis study.

Table 2. Relative risk of developing RF-seropositive RA, RF-seronegative RA, and total RA among current cigarette smokers, by sex and information source*

	Women†		Men†		All‡	
	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)
Questionnaire						
RF-seropositive RA	126/173	2.2 (1.6–3.0)	47/63	3.3 (1.7–6.3)	173/236	2.2 (1.7–3.0)
RF-seronegative RA	48/173	0.8 (0.6–1.3)	14/63	0.6 (0.3–1.3)	62/236	0.8 (0.6–1.2)
Total RA	174/173	1.5 (1.1–2.0)	61/63	1.7 (1.0–2.9)	235/236	1.5 (1.2–2.0)
Blood sample						
RF-seropositive RA	123/100	2.3 (1.6–3.4)	47/31	3.5 (1.6–7.4)	170/131	2.4 (1.7–3.3)
RF-seronegative RA	47/100	0.9 (0.5–1.3)	13/31	0.5 (0.2–1.2)	60/131	0.8 (0.6–1.2)
Total RA	170/100	1.6 (1.1–2.2)	60/31	1.8 (0.9–3.3)	230/131	1.6 (1.2–2.2)

* Cases/controls are the number of exposed cases/number of exposed controls. The relative risk (RR) and 95% confidence interval (95% CI) are versus never smokers. RF = rheumatoid factor; RA = rheumatoid arthritis.

† RR adjusted for age (10 strata) and residential area.

‡ RR adjusted for age (10 strata), sex, and residential area.

control selection. Furthermore, socioeconomic group, body mass index (BMI), marital status, parity, and oral contraceptive use were considered to be potential confounding factors. In the analysis, socioeconomic group was categorized into 8 strata, and the other variables were dichotomized as follows: BMI as ≥ 25 kg/m² versus BMI < 25 kg/m², marital status as married and/or cohabiting with an adult versus other status, parity as yes or no, and oral contraceptive use as ever or never.

Statistical analysis. In the data analysis, subjects with different genotypes and smoking habits were compared with regard to the incidence of RF-seropositive RA, RF-seronegative RA, and total RA, by calculating the odds ratios (ORs) with 95% confidence intervals (95% CIs) by means of logistic regression. ORs were interpreted as the relative risk, because the study was population-based and the control population was a random sample from the study base (10). We performed matched analyses (conditional logistic regression) based on all available case–control pairs, as well as unmatched analyses of the data (unconditional logistic regression) based on all available cases and all available controls. Herein, we present only the results from the unmatched analyses, since these were in close agreement with those from the matched analyses but had higher precision.

Adjustment for socioeconomic class, BMI, marital status, parity, and oral contraceptive use had minor influences on the results of the study, and was therefore not retained in the final analyses. Interaction between genotype and smoking habits was evaluated, using departure from additivity of effects as the criterion of interaction, as suggested by Rothman et al (11). To quantify the amount of interaction, the attributable proportion (AP) due to interaction was calculated together with the 95% CI (12). The AP due to interaction, which is expressed as a value between 0 and 1, is the proportion of the incidence among persons exposed to 2 interacting factors that is attributable to the interaction per se (i.e., reflecting their combined effect beyond the sum of their independent effects). Missing values occurred for only 1 variable (for 33 subjects, the value for socioeconomic group was missing). All analyses were conducted using the SAS software package, version 8.2 (SAS Institute, Cary, NC).

RESULTS

Characteristics of the subjects. The study involved mainly individuals born in Sweden (Table 1). Among all of the study subjects, including those born outside Sweden, 97% were of Caucasian origin. The proportion of blue collar workers was somewhat higher among male cases than among male controls. The BMI was similar between the cases and controls. Of the cases, ~66% had seropositive RA.

Cigarette smoking as a risk factor for RA. Among all cases and controls who responded to the questionnaire and provided information on smoking habits, the relative risk of developing RA was 1.5 (95% CI 1.2–2.0) for current smokers compared with never smokers (relative risk 1.5 in women and 1.7 in men) (Table 2). After subdividing these RA cases according to RF status at inclusion, the relative risk of RF-seropositive RA in current smokers was 2.2 (95% CI 1.7–3.0), but only 0.8 (95% CI 0.6–1.2) for RF-seronegative RA. The pattern was similar between the women and men. Very similar relative risk values were obtained when the analysis was restricted to those individuals from whom blood samples were available for subsequent genetic analysis (Table 2).

SE genes as a risk factor for RA. The relative risk for development of RA associated with the presence of SE genes was calculated both for the presence of single SE genes and for the presence of double SE genes. Women and men were analyzed separately, and subjects were also analyzed according to RF status (RF-seropositive RA versus RF-seronegative RA). As is apparent in Table 3, both single and double SE genes

Table 3. Relative risk of developing RF-seropositive RA, RF-seronegative RA, and total RA among carriers of SE genes (single, double, or any), by sex*

	Single SE		Double SE		Any SE	
	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)
Women						
RF-seropositive RA	183/192	2.2 (1.6–3.0)	114/53	5.0 (3.3–7.5)	297/245	2.8 (2.1–3.8)
RF-seronegative RA	103/192	1.3 (0.9–1.9)	28/53	1.3 (0.8–2.2)	131/245	1.3 (0.9–1.8)
Total RA	286/192	1.8 (1.3–2.3)	142/53	3.2 (2.2–4.6)	428/245	2.1 (1.6–2.7)
Men						
RF-seropositive RA	85/72	3.7 (2.2–6.1)	48/14	10.6 (5.2–21.9)	133/86	4.8 (2.9–7.8)
RF-seronegative RA	38/72	1.5 (0.9–2.7)	11/14	2.3 (0.9–5.5)	49/86	1.7 (1.0–2.8)
Total RA	123/72	2.6 (1.7–4.0)	59/14	6.3 (3.2–12.3)	182/86	3.2 (2.1–4.8)
All†						
RF-seropositive RA	268/264	2.5 (1.9–3.3)	162/67	6.0 (4.2–8.5)	430/331	3.2 (2.5–4.2)
RF-seronegative RA	141/264	1.4 (1.0–1.8)	39/67	1.5 (1.0–2.3)	180/331	1.4 (1.1–1.9)
Total RA	409/264	2.0 (1.6–2.5)	201/67	3.8 (2.7–5.2)	610/331	2.3 (1.9–2.9)

* Cases/controls are the number of exposed cases/number of exposed controls. The relative risk (RR) and 95% confidence interval (95% CI) are versus noncarriers of shared epitope (SE) genes. RF = rheumatoid factor; RA = rheumatoid arthritis.

† RR adjusted for sex.

were associated with an increased risk of RA in men and in women, with double SE genes conferring a higher risk than a single SE gene. When analyzed further, both single and double SE genes were related to an increased risk of RF-seropositive RA, but not of RF-seronegative RA.

Interaction between smoking and SE genes. Table 4 depicts the results from the analysis of interaction between cigarette smoking and occurrence of SE genes as it relates to the development of RA (RF-seropositive and RF-seronegative RA analyzed together). The risk of RA associated with SE genes among never smokers (women and men together) was only moderately increased (relative risk 1.5, 95% CI 1.0–2.2). In smoking subjects with any SE gene, the relative risk was 2.9 (95%

CI 1.9–4.5). For this group, the AP due to interaction was 0.4 (95% CI 0.2–0.7), indicating that the interaction between cigarette smoking and SE genes is statistically significant. An even stronger interaction was observed between smoking and double SE genes (relative risk 5.6, 95% CI 2.9–11.1; AP due to interaction 0.7, 95% CI 0.4–0.9). The relative risk in smokers carrying only a single SE gene was intermediate.

Interaction between smoking and SE genes in relation to RF-seropositive RA or RF-seronegative RA. In Table 5, with women and men analyzed together, results of the same interaction analysis as described above are presented, but the association with regard to RF-seropositive RA and RF-seronegative RA is exam-

Table 4. Relative risk of developing rheumatoid arthritis among subjects exposed to different combinations of current cigarette smoking habits and SE genes and attributable proportion due to interaction between cigarette smoking and SE genes (single, double, or any), by sex*

	No SE		Single SE		Double SE		Any SE	
	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)
Women†								
Never smokers	63/73	1.0 –	98/75	1.7 (1.0–2.6)	40/25	1.9 (1.0–3.6)	138/100	1.7 (1.1–2.7)
Current smokers	44/43	1.3 (0.7–2.2)	79/46	2.4 (1.4–4.0)	47/11	5.8 (2.7–12.5)	126/57	3.0 (1.9–4.9)
Men‡								
Never smokers	16/13	1.0 –	29/29	0.7 (0.3–1.9)	9/6	1.3 (0.3–5.3)	38/35	0.8 (0.3–2.1)
Current smokers	16/17	0.7 (0.2–2.1)	28/11	1.9 (0.6–5.9)	16/3	5.5 (1.2–26.3)	44/14	2.7 (0.9–7.5)
All								
Never smokers	79/86	1.0 –	127/104	1.4 (0.9–2.1)	49/31	1.8 (1.0–3.1)	176/135	1.5 (1.0–2.2)
Current smokers	60/60	1.1 (0.7–1.9)	107/57	2.3 (1.4–3.7)	63/14	5.6 (2.9–11.1)	170/71	2.9 (1.9–4.5)
AP	–	–	–	0.3 (0.0–0.7)	–	0.7 (0.4–0.9)	–	0.4 (0.2–0.7)

* Values are the relative risk (RR) and 95% confidence interval (95% CI) as compared with never smokers without shared epitope (SE) genes. Cases/controls are the number of exposed cases/number of exposed controls. AP = attributable proportion.

† RR adjusted for age (10 strata) and residential area.

‡ RR adjusted for age (10 strata), residential area, and sex.

Table 5. Relative risk of developing RF-seropositive RA and RF-seronegative RA among all male and female subjects exposed to different combinations of cigarette smoking habits and SE genes, and attributable proportion due to interaction between cigarette smoking and SE genes (single, double, or any)*

	No SE		Single SE		Double SE		Any SE	
	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)	Cases/controls	RR (95% CI)
RF-seropositive RA†								
Never smokers	26/86	1.0 –	68/104	2.4 (1.4–4.2)	35/31	4.2 (2.1–8.3)	103/135	2.8 (1.6–4.8)
Current smokers	40/60	2.4 (1.3–4.6)	77/57	5.5 (3.0–10.0)	53/14	15.7 (7.2–34.2)	130/71	7.5 (4.2–13.1)
AP	–	–		0.3 (0.0–0.7)		0.6 (0.4–0.9)		0.4 (0.2–0.7)
RF-seronegative RA†								
Never smokers	53/86	1.0 –	59/104	0.9 (0.6–1.5)	14/31	0.7 (0.4–1.6)	73/135	0.9 (0.6–1.4)
Current smokers	20/60	0.6 (0.3–1.1)	30/57	0.9 (0.5–1.7)	10/14	1.2 (0.5–3.0)	40/71	1.0 (0.6–1.7)

* Values are the relative risk (RR) and 95% confidence interval (95% CI) as compared with never smokers without shared epitope (SE) genes. Cases/controls are the number of exposed cases/number of exposed controls. RF = rheumatoid factor; RA = rheumatoid arthritis; AP = attributable proportion.

† RR adjusted for age (10 strata), sex, and residential area.

ined separately. Compared with never smokers without SE genes, the relative risk of developing RF-seropositive RA among never smokers with SE genes was 2.8 (95% CI 1.6–4.8). The corresponding relative risk in current cigarette smokers without SE genes was 2.4 (95% CI 1.3–4.6). Thus, it is evident that smoking and SE genes are independently related to the development of RF-seropositive RA. Among current smokers with SE genes, the relative risk of developing RF-seropositive RA was 7.5 (95% CI 4.2–13.1). Therefore, an interaction between smoking and SE genes was observed in association with RF-seropositive RA, which also was reflected in the AP due to interaction, which was 0.4 (95% CI 0.2–0.7). The interaction was even more pronounced in smoking subjects with double SE genes, whose relative risk of RF-seropositive RA was 15.7 (95% CI 7.2–34.2). The AP due to interaction for this group was 0.6 (95% CI 0.4–0.9). Neither smoking nor SE genes nor the combination of these factors increased the risk of developing RF-seronegative RA.

DISCUSSION

Two principal findings are presented in this study. First, a striking gene–environment interaction between smoking and HLA-DRB1 genotypes was seen for seropositive, but not seronegative, RA, something that should have implications for formulations of pathogenetic hypotheses in these 2 conditions. Second, the data demonstrate that the risk associated with one of the classic genetically defined risk factors for an autoimmune disease is strongly influenced by the presence of an environmental factor—smoking.

Our study was designed as a population-based case–control study with incident cases, in which infor-

mation regarding smoking habits was collected retrospectively. One inherent problem in this case–control design is that recall bias may introduce systematic error in the calculation of the association between smoking and RA. In order to minimize such a recall bias, we only included incident cases of RA with a duration of symptoms mainly shorter than 1 year, in order to diminish the time between the etiologically relevant exposure and the time of response to our questionnaire. We also took great effort to obtain the information on smoking in an identical way between cases and controls. The fact that only seropositive RA cases, but not seronegative cases, more often reported smoking habits than did controls indicated to us that recall bias was most probably not a significant problem, since there is no reason to believe that seropositive patients would recall their smoking habits differently from seronegative patients.

RF status was only determined once. Thus, some cases might have been misclassified with regard to the presence of RF. This potential misclassification, however, would most likely be unrelated to the studied exposures, which in turn means that the potential bias in estimated relative risks is toward the null value (both with regard to RF-seropositive RA and RF-seronegative RA).

Another potential methodologic problem is that the recruitment of cases and controls may introduce selection biases. In order to minimize such recruitment bias, we took advantage of the fact that almost all health care in Sweden is provided within the general health care system, and that all such units in the area that defined the study base contributed to our study, as did almost all of the few privately run rheumatology units. Nevertheless, some cases may have been unidentified in

our study, for instance, cases diagnosed in primary health care facilities that were never referred to a rheumatology unit. However, we know, on the basis of population-based studies aimed at identifying RA cases directly in primary care, that almost all cases of RA in our current Swedish system are indeed referred to rheumatology units (13). It is therefore not likely that the relatively few unidentified cases would cause a substantial bias in our calculations.

The response rate with regard to participation in the study was high, with 95% for cases and 83% for controls. Only control subjects who had responded to the questionnaire were asked to give a blood sample, and 60% of these controls provided a sample. We compared the controls with and without a blood sample, respectively, with regard to smoking habits, and no differences were found. We cannot find any reason to believe that the prevalence of SE genes should differ between controls with and without a blood sample, since there were no differences between these groups with respect to other parameters such as age, sex, and residential area.

Adjustments for socioeconomic class, BMI, marital status, parity, and oral contraceptive use had minor influences on the results of the study and were therefore not retained in the final analyses of interactions between SE genes and smoking. Our conclusion with regard to the methodology is, therefore, that there is always a risk of selection bias in the current type of case-control study, but we consider this risk to be well controlled.

The effects of smoking that were observed in our case-control study are in concordance with those previously reported by other groups. The results are consistent both with regard to the overall effect on the development of RA and with regard to the finding that smoking is primarily associated with seropositive RA (3,4,6,13).

In the present report, we restricted the analysis to comparing current and never smokers. Former smokers were excluded from the analysis. Previously we have extensively analyzed the effects of various time courses and dosages of smoking on the risk of RA, as presented in a recently published report (5). Briefly, there was a relationship between the risk of RA and the cumulative dose of smoking. In addition, the risk diminished with time after cessation of smoking, even if a long time was required for the risk to approach zero. Because we did not consider it possible to analyze the interactions between genes and smoking according to different time courses and dosages of smoking (due to low numbers), we had to choose one way to categorize smokers. We considered current smokers to be the best category in

this context, since current smokers also had a high cumulative smoking history, whereas former smokers had a wide variation in cumulative smoking history. We hope to be able to update our findings with more information on the effects of different time courses and dosages of smoking in future studies, when our database has grown larger.

With regard to the findings in the genetic analyses, our observation of a relatively modest risk of RA in individuals with a single SE gene is consistent with the findings in several studies of patients with early RA who were recruited directly from primary care (3). Less information has been published on the risk conferred by double SE genes, although there are data indicating an increased risk of RA associated with double SE genes (14,15), which is in accordance with our present findings.

We used the attributable proportion due to interaction (the AP value) as a measure to quantify the interaction between smoking and SE genes (12) and found significant gene-environment interactions concerning seropositive RA, both for smoking and a single SE gene and even more for smoking and double SE genes. The magnitude of the risk conferred by smoking and a defined genetic trait is striking, with relative risk values of 5.5 and 15.7 with single and double SE genes, respectively.

The molecular mechanisms responsible for the observed interaction are not yet known, but several interesting possibilities exist which now require further attention. One possible mechanism is that smoking may cause a modification of potential autoantigens being recognized by T cells, which are restricted by major histocompatibility complex (MHC) class II antigens carrying the SE structures (16). Of particular interest is the possibility that the process involved in citrullination of antigens might be influenced by events triggered by smoking. Another possibility is that the smoke may deliver neoantigens that might be bound to SE-containing MHC class II molecules, with subsequent T cell activation toward such antigens. A third possibility is that substances in the smoke (such as char) might act as adjuvants, and thereby trigger the innate immune system to contribute to arthritis development, similar to what has been shown in animal models of adjuvant-induced arthritis (17,18). A number of additional possibilities exist, including the notion that the gene involved in the gene-environment interaction is not HLA-DR itself but rather a gene in linkage disequilibrium with HLA-DR, and these possibilities might now be scrutinized.

In addition, we cannot exclude the possibility that the apparent gene-environment interaction that we

observed may be due, in part, to genetic influences on smoking behavior. We also need additional information on the potential influence of sex on the smoking-related risk of RA; our study showed this risk to be present both in men and in women, whereas 3 previous studies have demonstrated an increased risk among men only (4,19,20). Finally, a major finding of this study is that disease mechanisms dependent on SE gene–smoking interactions obviously are active only in RF-seropositive, but not RF-seronegative, RA, thus further emphasizing the necessity for subphenotyping of RA (i.e., RF status) in all pathogenetic and genetic studies of this disease.

Extending from the field of RA, a few general comments may also be made concerning the implications of this study in terms of gene–environment interactions in complex diseases. Our study emphasizes the need for careful subphenotyping of diseases with an unknown etiology (in this case, in RF-seropositive and RF-seronegative RA). Our study also emphasizes the need to include data on environmental exposures in genetic analyses of a complex disease, since it is now shown that even one of the most classic genetic risk factors for an autoimmune disease, the shared epitope in RF-seropositive RA, is strongly influenced by the presence of a defined environmental risk factor, smoking, in the population at risk.

ACKNOWLEDGMENTS

We thank Marie-Louise Serra for her excellent work with the data collection, Eva Jemseby for her invaluable assistance with the biobank, Olle Olerup for providing us with HLA typing kits, Diab Diab for performing HLA typing, and Associate Professor Robert A. Harris for linguistic advice. We also acknowledge the members of the Epidemiological Investigation of Rheumatoid Arthritis Study Group, which includes one of us (PS): Ingeli Andréasson, Landvetter; Eva Baecklund, Akademiska Hospital, Uppsala; Ann Bengtsson and Thomas Skogh, Linköping Hospital; Johan Bratt and Ingiöld Hafström, Huddinge University Hospital; Kjell Huddénus, rheumatology clinic in Stockholm; Shirani Jayawardene, Bollnäs Hospital; Ann Knight, Hudiksvall Hospital; Ido Leden, Kristianstad Hospital; Thomas Lerndal and Göran Lindahl, Danderyd Hospital; Bengt Lindell, Kalmar Hospital; Christin Lindström and Gun Sandahl, Sophiahemmet; Björn Löfström, Katrineholm Hospital; Birgitta Nordmark, Karolinska Hospital; Ingemar Petersson, Spenshult Hospital; Christoffer Schaufelberger, Sahlgrenska University Hospital; Berit Sverdrup, Eskilstuna Hospital; Olle Svernell, Västervik Hospital; and Tomas Weitoff, Gävle Hospital.

REFERENCES

1. Sollid LM. Coeliac disease: dissecting a complex inflammatory disorder. *Nat Rev Immunol* 2002;2:647–55.
2. 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.
3. Silman AJ, Newman J, MacGregor AJ. Cigarette smoking increases the risk of rheumatoid arthritis: results from a nationwide study of disease-discordant twins. *Arthritis Rheum* 1996;39:732–5.
4. Uhlig T, Hagen KB, Kvien TK. Current tobacco smoking, formal education, and the risk of rheumatoid arthritis. *J Rheumatol* 1999;26:47–54.
5. 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. *Ann Rheum Dis* 2003;62:835–41.
6. Matvey DL, Hutchinson D, Dawes PT, Nixon NB, Clarke S, Fisher J, et al. Smoking and disease severity in rheumatoid arthritis: association with polymorphism at the glutathione S-transferase M1 locus. *Arthritis Rheum* 2002;46:640–6.
7. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
8. Aldener-Cannava A, Olerup O. HLA-DPA1 typing by PCR amplification with sequence-specific primers and distribution of DPA1 alleles in Caucasian, African and Oriental populations. *Tissue Antigens* 1996;48:153–60.
9. Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. *Tissue Antigens* 1992;39:225–35.
10. Miettinen O. Estimability and estimation in case-referent studies. *Am J Epidemiol* 1976;103:226–35.
11. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *Am J Epidemiol* 1980;112:467–70.
12. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology* 1992;3:452–6.
13. Symmons DP, Bankhead CR, Harrison BJ, Brennan P, Barrett EM, Scott DG, 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.
14. Harrison B, Thomson W, Symmons D, Ollier B, Wiles N, Payton T, et al. The influence of HLA-DRB1 alleles and rheumatoid factor on disease outcome in an inception cohort of patients with early inflammatory arthritis. *Arthritis Rheum* 1999;42:2174–83.
15. Meyer JM, Evans TI, Small RE, Redford TW, Han J, Singh R, et al. HLA-DRB1 genotype influences risk for and severity of rheumatoid arthritis. *J Rheumatol* 1999;26:1024–34.
16. Klareskog L, McDevitt H. Rheumatoid arthritis and its animal models: the role of TNF- α and the possible absence of specific immune reactions. *Curr Opin Immunol* 1999;11:657–62.
17. Lorentzen JC, Glaser A, Jacobsson L, Galli J, Fakhrai-rad H, Klareskog L, et al. Identification of rat susceptibility loci for adjuvant-oil-induced arthritis. *Proc Natl Acad Sci U S A* 1998;95:6383–7.
18. Wilder RL, Griffiths MM, Remmers EF, Cannon GW, Caspi RR, Kawahito Y, et al. Localization in rats of genetic loci regulating susceptibility to experimental erosive arthritis and related autoimmune diseases. *Transplant Proc* 1999;31:1585–8.
19. Heliovaara M, Aho K, Aromaa A, Knekt P, Reunanen A. Smoking and risk of rheumatoid arthritis. *J Rheumatol* 1993;20:1830–5.
20. Krishnan E, Sokka T, Hannonen P. Smoking-gender interaction and risk for rheumatoid arthritis. *Arthritis Res Ther* 2003;5:R158–62.