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

Bo Ding,[1] Leonid Padyukov,[2] Emeli Lundström,[2] Mark Seielstad,[3] Robert M. Plenge,[4] Jorge R. Oksenberg,[5] Peter K. Gregersen,[6] Lars Alfredsson,[1] and Lars Klareskog[2]

Objective. **To identify additional variants in the major histocompatibility complex (MHC) region that independently contribute to risk in 2 disease subsets of rheumatoid arthritis (RA) defined according to the presence or absence of antibodies to citrullinated protein antigens (ACPAs).**

Methods. **In a multistep analytical strategy using unmatched as well as matched analyses to adjust for HLA–DRB1 genotype, we analyzed 2,221 single-nucleotide polymorphisms (SNPs) spanning 10.7 Mb, from 6p22.2 to 6p21.31, across the MHC. For ACPA-positive RA, we analyzed samples from the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) and the North American Rheumatoid Arthritis Consortium (NARAC) studies (totaling 1,255 cases and 1,719 controls). For ACPA-negative RA, we used samples from the EIRA study (640 cases and 670 controls). Plink and SAS statistical packages were used to conduct all statistical analyses.**

Results. **A total of 299 SNPs reached locus-wide significance ($P < 2.3 \times 10^{-5}$) for ACPA-positive RA, whereas surprisingly, no SNPs reached this significance for ACPA-negative RA. For ACPA-positive RA, we adjusted for known DRB1 risk alleles and identified additional independent associations with SNPs near HLA–DPB1 (rs3117213; odds ratio 1.42 [95% confidence interval 1.17–1.73], $P_{combined} = 0.0003$ for the strongest association).**

Conclusion. **There are distinct genetic patterns of MHC associations in the 2 disease subsets of RA defined according to ACPA status. HLA–DPB1 is an independent risk locus for ACPA-positive RA. We did not identify any associations with SNPs within the MHC for ACPA-negative RA.**

Rheumatoid arthritis (RA) is a complex disease that is influenced by both genetic and environmental factors (1,2). It can be divided into 2 major clinical subtypes that are defined by the presence or absence of

antibodies to citrullinated protein antigens (ACPA) (3,4). The major histocompatibility complex (MHC) locus has long been known to contain genes that confer substantial risk of the disease (5). Much of the genetic signal maps to class II HLA–DRB1 alleles (DRB1*01, DRB1*04, and DRB1*10, which are known as shared epitope [SE] alleles), and this information has been crucial to the working hypothesis concerning the role of class II MHC–dependent immune activation in the pathogenesis of RA (6). Previous studies have also suggested the presence of other MHC risk factors (7–10), although their precise localization has been hampered by linkage disequilibrium and insufficient single-nucleotide polymorphism (SNP) genotyping over the entire region.

It was recently demonstrated that the SE alleles are associated only with the risk of ACPA-positive RA but not with ACPA-negative disease (11–13). Two studies showed that HLA–DRB1*03 is associated with ACPA-negative RA (14,15), thus emphasizing the need to consider the subdivision of RA according to ACPA status in genetic studies. A crucial question for future understanding of the immune pathogenesis of RA is whether polymorphisms in other MHC loci are associated with only the ACPA-positive form of RA or whether different MHC loci are associated with different disease variants.

PATIENTS AND METHODS

Epidemiological Investigation of Rheumatoid Arthritis (EIRA) study population. This study was approved by ethics review boards at the Karolinska Institutet, and informed consent was obtained from all participating subjects. EIRA is a population-based case–control study of incident cases of RA defined according to the American College of Rheumatology (ACR; formerly, the American Rheumatism Association) 1987 criteria (16), which enrolls 85% of patients within 1 year after initial arthritis symptoms. Most subjects were born in Sweden, and 97% report a white ancestry (for details, see ref. 17).

North American Rheumatoid Arthritis Consortium (NARAC) study population. Cases consisted of RA patients of self-reported white ancestry that were randomly drawn from 4 different collections of patients (the NARAC population, the National Data Bank for Rheumatic Diseases, the National Inception Cohort of Rheumatoid Arthritis, and the Study of New-Onset Rheumatoid Arthritis) (for details, see ref. 18). All patients selected for the present study were ACPA-positive and met the ACR 1987 criteria for RA.

Control subjects were selected based on similar self-reported ancestry from among 20,000 persons who were part of the New York Cancer Project (18).

Genotyping. *Genome-wide association study (GWAS).* Genotyping was performed as previously described (18). Briefly, the Illumina Human Hap300 version 1.0 chip (Illu-

mina, San Diego, CA) containing probes for 317,503 SNPs was used. Samples included in the analysis had call rates >96%, and more than 93% of samples had a SNP call rate >99% (mean 99.7% for all completed samples). Forty-one samples were genotyped twice, with a mean concordance prior to any SNP quality control filtering of 99.96% (median 99.98% [range 99.24–100%]).

IMAGEN analysis. Genotyping was also performed as described by the International MHC and Autoimmunity Genetics Network (IMAGEN) Consortium (submitted for publication). Briefly, the Illumina GoldenGate assay was used to genotype a panel of 1,230 SNPs. Replication genotyping was performed by iPlex Gold chemistry (Sequenom, San Diego, CA).

HLA–DRB1 type was determined by sequence-specific primer–polymerase chain reaction analysis using low-resolution HLA–DR and DR4 kits from Olerup SSP (Saltsjö-baden, Sweden). High-resolution (4 digits) genotype data were used to define the categories of SE. HLA–DRB1*01, *0401, *0404, *0405, *0408, and *10 were defined as SE alleles. Any genotype with a combination of 2 of these alleles was considered to be a double SE.

Grouping of HLA–DRB1 SE alleles and genotypes. The rationale for our classification of SE alleles into different groups was to adjust the associations between RA risk and non-DRB1 genetic characteristics for confounding by DRB1 as much as possible, especially confounding by the DRB1 SE alleles. We tried to achieve this in different ways. We used 3 different classifications of SE alleles (models 1, 2, and 3). The traditional SE grouping (model 1, consisting of 3 categories) was derived from the number of copies of SE alleles (0, 1, and 2 copies). The other groupings (model 2 and model 3) were created based on our data. We used the odds ratio (OR) for each SE allele combination in our study as the basis for grouping and tried to 1) group SE allele combinations that had similar ORs, 2) group the less-frequent SE alleles, and 3) separate the DRB1*01 allele and the DRB1*10 allele from the DRB1*04 allele. In addition, we categorized the SE alleles as proposed by Tezenas du Montcel et al (19).

HLA–DRB1 genotype matching. For the EIRA study, high-resolution DRB1*04 (4 digits) typing and low-resolution (2 digits) typing for other allelic groups were used to perform case–control matching at a ratio of 1:1. For the NARAC study, high-resolution DRB1*01 (4 digits), DRB1*04 (4 digits), DRB1*09 (4 digits), DRB1*15 (4 digits), and DRB1*16 (5 digits) typing and low-resolution (2 digits) typing for other allelic groups were used. Using this stringent selection strategy, we identified 358 pairs of cases and controls from the EIRA and 264 pairs from the NARAC study.

Quality control procedures. All genotype data were processed using the statistical software package Plink (20). We combined data from the MHC region in the data sets from the GWAS and IMAGEN analyses and then did filtering. To quantify and control for population stratification, we used a principal components analysis approach implemented in the EigenStrat software (21). Using the genome-wide SNP data (18), EigenStrat identified 53 significant outliers ($\sigma = 4$; iterations = 5, with no outliers identified after the third iteration) from the ACPA-positive RA patient and control association analysis, and 59 significant outliers from the ACPA-negative RA patient and control association analysis.
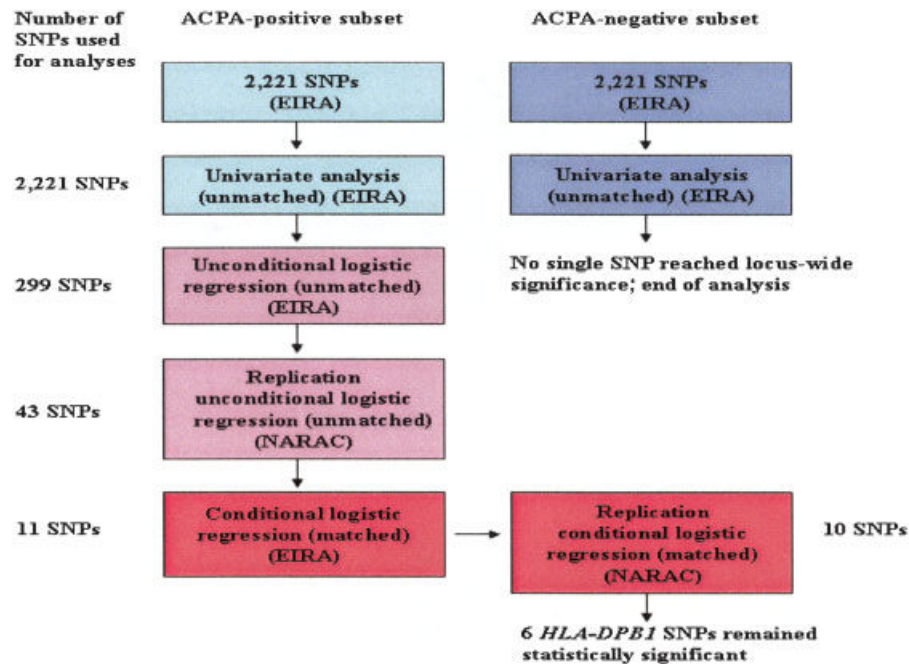
Figure 1. Analytical approach used in our study of variants in the major histocompatibility complex region that contribute independently to risk in anti–citrullinated protein antibody (ACPA)–positive rheumatoid arthritis (RA) and ACPA-negative RA, as well as the number of single-nucleotide polymorphisms (SNPs) analyzed at each step. EIRA = Epidemiological Investigation of Rheumatoid Arthritis; NARAC = North American Rheumatoid Arthritis Consortium.

These outliers and subjects with self-reported non-Swedish ancestry were removed from the final analysis.

Data sets were filtered as follows: SNPs with >5% missing data (n = 8 for the ACPA-positive and n = 10 for the ACPA-negative subgroups), control Hardy-Weinberg equilibrium at $P < 2.251 \times 10^{-5}$ (n = 9), and minor allele frequency <0.01 (n = 85 for ACPA-positive and n = 74 for ACPA-negative subgroup) were excluded. We found no individuals with >5% missing genotypes in both the ACPA-positive and ACPA-negative subgroups. There were 2,122 SNPs that passed quality control filters in the ACPA-positive group and 2,131 in the ACPA-negative group.

Statistical analysis. We used the Armitage trend test for the initial univariate test of association for both ACPA-positive and ACPA-negative subsets implemented in the package Plink (20). P values less than 0.05 after Bonferroni correction were considered statistically significant for the univariate analysis. Unconditional logistic regression and conditional logistic regression were conducted using the SAS statistical package (version 9.1.3; SAS Institute, Cary, NC). Raw genotypes were recoded as a score variable (0, 1, and 2), counting the number of common alleles using Plink. The genotype variable was entered into the logistic regression models. Associations are reported as ORs and 95% confidence intervals (95% CIs), which were calculated from the models.

RESULTS

MHC genetic patterns of ACPA-positive and ACPA-negative RA. In order to identify variants in the MHC region that might contribute to risk of the 2 forms of RA that are defined by presence and absence of ACPA, we selected tag SNPs to capture common genetic variation across the MHC region, using both a set of 1,230 SNPs selected for a combined analysis of 7 different inflammatory diseases (the IMAGEN study) (IMAGEN Consortium: submitted for publication) and a set of 1,298 additional SNPs covering the MHC region that were included in a GWAS (18). There were 307 SNPs that overlapped between the IMAGEN and the GWAS data, leaving a total of 2,221 SNPs for analysis.

For the exploratory analysis, we genotyped a total of 1,291 RA patients (cases), who were selected equally from the 2 major RA subsets (651 ACPA-positive and 640 ACPA-negative), and 670 controls; all of these study subjects were from the EIRA population (18,22). For replication, we used data from the NARAC study (18). In the NARAC population, all RA cases were ACPA-positive; therefore, we used these data for
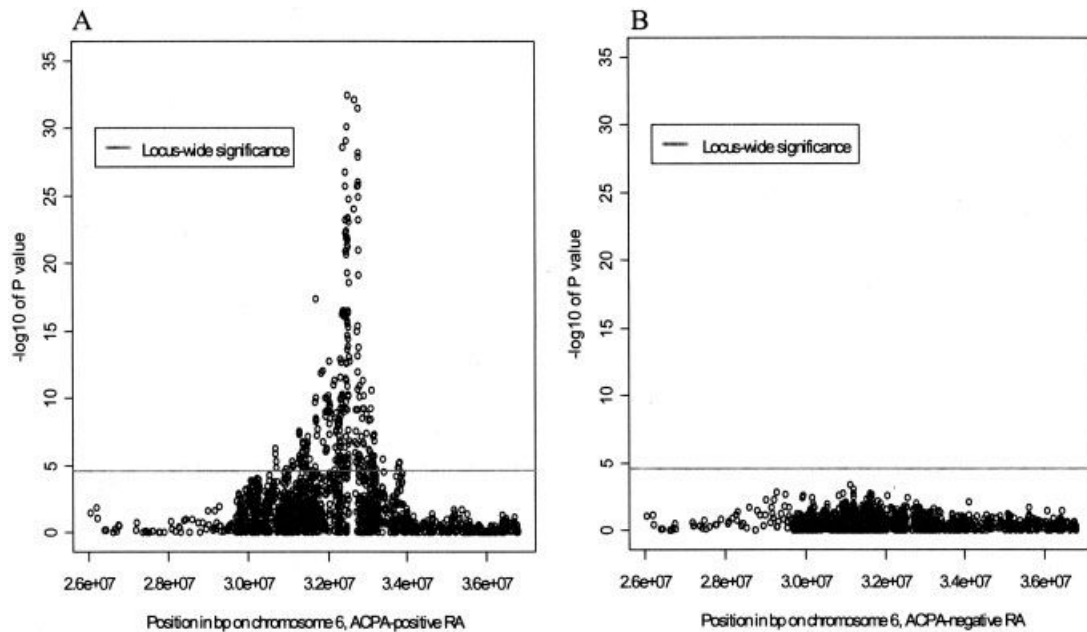
Figure 2. Single-nucleotide polymorphisms found to be significant (locus-wide significance defined as $P < 2.3 \times 10^{-5}$) in case–control association analyses, as determined by Armitage test for trend. Rheumatoid arthritis (RA) cases with anti–citrullinated protein antibody (ACPA) (**A**) and without ACPA (**B**) and control groups from the Epidemiological Investigation of Rheumatoid Arthritis population sample were analyzed.

replication and extension of the findings in the ACPA-positive RA cases.

The analytical strategy is illustrated in Figure 1. Due to the availability of equal numbers of samples from ACPA-positive and ACPA-negative RA patients as well as matched controls in the EIRA population, we performed the initial analyses in this group. In the initial univariate analysis of EIRA cases and controls, 299 SNPs reached locus-wide significance (defined here as $P < 2.3 \times 10^{-5}$) when the ACPA-positive RA cases were compared with the controls (Figure 2A). In contrast, no single SNP was found to be statistically significant at this level when the ACPA-negative RA cases were compared with the controls (Figure 2B), despite similar statistical power for the 2 subsets of RA cases in the EIRA study. This provides strong evidence of genetically distinct etiologies behind these 2 forms of RA. Subsequently, we analyzed only ACPA-positive RA cases and controls.

Associations of MHC loci independent of HLA–DRB1 in ACPA-positive RA cases using unmatched analysis. To adjust for the influence of the SE alleles, we initially used unconditional logistic regression analyses, including all ACPA-positive cases and controls in the EIRA study. We first investigated which HLA–DR genotypes were dependent on known HLA–DRB1 SE

alleles, as typed by conventional HLA–DR typing (high resolution 4-digit genotype) (Table 1).

We confirmed that different DRB1 SE genotypes confer different levels of risk of developing RA (23). We tried 4 different models to adjust for DRB1 SE alleles, based on categorization of the various DRB1 SE variants into 3, 5, or 7 different categories (Table 1), in order to analyze whether any of the initially identified 299 SNPs conferred risk of ACPA-positive RA independently of the known DRB1 SE risk alleles. In addition, we also included the classification of DRB1 alleles according to the method described by Tezenas du Montcel et al (19). We found that our third model, which included 7 DRB1 SE categories, gave the smallest number of SNPs (70 SNPs) at the $\sigma = 0.05$ level and that among the 70 significant SNPs, 61 were common among these 4 classification methods. In further analyses, we chose model 3 to control for the effect of DRB1 SE alleles and found that 70 of the 299 SNPs that were associated with ACPA-positive RA were independent of DRB1 SE alleles.

In order to replicate these findings, we used data from 604 ACPA-positive RA cases and 1,049 controls in the NARAC study. Here, we had access to information on DRB1 type and SNP genotypes for 43 of the 70 SNPs

Table 1. OR of developing ACPA-positive RA in association with the HLA–DRB1 genotype, and 4 different methods of categorizing HLA–DRB1 in analyses of ACPA-positive RA*

HLA–DRB1 genotype	No. of cases	No. of controls	OR (95% CI)	Model 1 (3 categories)	Model 2 (5 categories)	Model 3 (7 categories)	Tezenas du Montcel classification†
No SE alleles/no SE alleles	130	474	Reference	I (0 copies of SE allele)	I	I	L/L
*01/*01	9	14	1.97 (0.76–5.15)	III (2 copies of SE allele)	II	S_{3p}/S_{3p}	
*01/*0401	78	24	12.9 (7.75–21.56)	III	III	S_2/S_{3p}	
*01/*0404	14	7	8.51 (3.25–22.30)	III	II	II	S_{3p}/S_{3p}
*01/*0408	3	2	6.46 (0.83–50.19)	III	II	S_{3p}/S_{3p}	
*01/*10	4	3	4.74 (1.03–21.92)	III	II	II	S_{3p}/S_{3p}
*01/no SE alleles	120	126	3.53 (2.55–4.89)	II (1 copy of SE allele)	II	II	S_{3p}/L
*0401/*0401	63	13	17.71 (9.36–33.49)	III	IV	IV	S_2/S_2
*0401/*0404	63	8	28.79 (13.34–62.11)	III	V	V	S_2/S_{3p}
*0401/*0408	14	5	9.44 (3.29–27.13)	III	V	VI	S_2/S_{3p}
*0401/no SE alleles	222	167	5.07 (3.80–6.77)	II	V	VI	S_2/L
*0404/*0404	10	4	5.02 (1.35–18.62)	III	V	VI	S_{3p}/S_{3p}
*0404/no SE alleles	73	48	5.90 (3.85–9.04)	II	V	VI	S_{3p}/L
*10/*0401	11	3	13.23 (3.57–49.03)	III	II	VII	S_2/S_{3p}
*10/*0404	4	3	4.46 (0.97–20.59)	III	II	VII	S_{3p}/S_{3p}
*10/no SE alleles	11	11	2.93 (1.17–7.33)	II	II	VII	S_{3p}/L

* Odds ratios (ORs) were adjusted for age, sex, and geographic location. The shared epitope (SE) was defined as follows: HLA–DRB1*01 (*0103 was excluded), *0401, *0404, *0405, *0408, and *10; the *0401/*0408 genotype consists of *0401/*0405 (n = 5) and *0401/*0408 (n = 14); the *0404/*0404 genotype also includes *0404/*0408 (n = 2); and the *10/*0404 genotype also includes *10/*0405 (n = 2). See Patients and Methods for an explanation of the groupings of models 2 and 3. ACPA = anti–citrullinated protein antibody; RA = rheumatoid arthritis; 95% CI = 95% confidence interval.
† DRB1 alleles are divided into susceptibility (S) and nonsusceptibility (X) alleles based on the presence or absence of RAA at amino acid positions 72–74. S alleles are further divided according to the amino acid at position 71 (S_2 = KRAA and S_3 = RRAA), and S_3 alleles are further divided according to the amino acid at position 70 (S_{3P} = QRRAA or RRRAA). All other alleles (S_{1A} = ARAA at amino acids 71–74, S_{1E} = ERAA at amino acids 71–74, S_{3D} = DRRAA at amino acids 70–74, and X) are designated as low-risk (L) alleles (see ref. 19).

that had been selected from the initial EIRA analysis. We adjusted for DRB1 SE alleles in the same way, and only SNPs that in the NARAC sample were statistically significantly associated with ACPA-positive RA at $P < 0.05$ and had the same direction of association were considered to be replicated. Using these criteria, 11 of the 43 SNPs were replicated (Table 2), representing 4 different loci: HLA–DPB1, C2-DOM3Z, MICA, and HLA–DQA1. The position and linkage disequilibrium structure of these 11 SNPs are shown in Figure 3.

Table 2. P values and ORs of developing ACPA-positive RA in association with HLA genotypes after adjustment for HLA–DRB1 categorized according to model 3 (7 categories)*

SNP	Position	Gene	EIRA P	EIRA OR (95% CI)	NARAC P	NARAC OR (95% CI)	Combined P	Combined OR (95% CI)
rs2064476	33181300	HLA–DPB1	5.4×10^{-4}	1.46 (1.18–1.81)	8.2×10^{-4}	1.39 (1.14–1.68)	2.3×10^{-6}	1.41 (1.22–1.62)
rs3091282	33165176	HLA–DPB1	6.6×10^{-4}	1.45 (1.17–1.80)	1.1×10^{-3}	1.38 (1.14–1.66)	3.6×10^{-6}	1.40 (1.21–1.61)
rs9277378	33158257	HLA–DPB1	6.5×10^{-4}	1.47 (1.18–1.83)	2.0×10^{-3}	1.35 (1.12–1.64)	6.7×10^{-6}	1.39 (1.21–1.61)
rs3117213	33172583	HLA–DPB1	1.2×10^{-3}	1.44 (1.15–1.79)	2.7×10^{-3}	1.36 (1.11–1.66)	1.5×10^{-5}	1.38 (1.19–1.60)
rs3117234	33181962	HLA–DPB1	3.0×10^{-3}	1.43 (1.13–1.81)	2.8×10^{-3}	1.37 (1.12–1.69)	3.9×10^{-5}	1.39 (1.19–1.62)
rs2395314	33170651	HLA–DPB1	1.5×10^{-3}	1.42 (1.14–1.77)	3.5×10^{-3}	1.35 (1.10–1.64)	2.3×10^{-5}	1.37 (1.18–1.59)
rs3117226	33165637	HLA–DPB1	1.8×10^{-3}	1.43 (1.14–1.79)	3.9×10^{-3}	1.34 (1.10–1.64)	3.1×10^{-5}	1.37 (1.18–1.59)
rs3128930	33183644	HLA–DPB1	6.1×10^{-4}	1.50 (1.19–1.90)	4.0×10^{-3}	1.35 (1.10–1.65)	1.5×10^{-5}	1.40 (1.20–1.63)
rs544167	31998137	C2-DOM3Z	7.1×10^{-5}	3.76 (1.96–7.22)	0.0249	1.57 (1.06–2.33)	2.1×10^{-5}	2.06 (1.48–2.87)
rs2523451	31477130	MICA	0.0125	0.79 (0.65–0.95)	0.0261	0.81 (0.68–0.98)	8.8×10^{-4}	0.80 (0.70–0.91)
rs6457617	32771829	HLA–DQA1	8.7×10^{-3}	1.38 (1.09–1.77)	0.0333	1.26 (1.02–1.57)	1.6×10^{-3}	1.29 (1.10–1.52)

* Results are from replication in the North American Rheumatoid Arthritis Consortium (NARAC) sample and from a combined analysis of the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) and NARAC samples for the 11 single-nucleotide polymorphisms (SNPs) initially identified in the EIRA study. P values and odds ratios (ORs) per copy of the risk allele were calculated by unconditional logistic regression modeling. ACPA = anti–citrullinated protein antibody; RA = rheumatoid arthritis; 95% CI = 95% confidence interval.
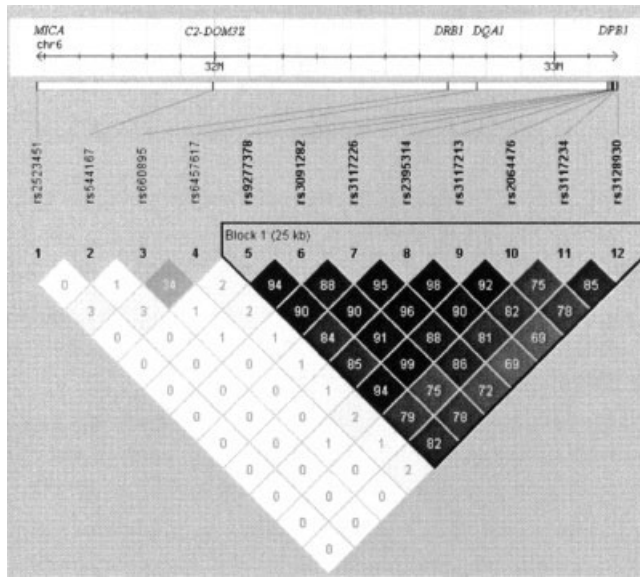
Figure 3. Linkage disequilibrium structure for the 11 replicated single-nucleotide polymorphisms (SNPs) after adjustment for HLA–DRB1 in the Epidemiological Investigation of Rheumatoid Arthritis population sample. Correlation coefficients (r^2; $\times$ 100) are shown in the individual boxes; white denotes $r^2 = 0$, black denotes $r^2 = 1$, and various shades of gray denote $0 < r^2 < 1$. SNP positions (in million bp) are indicated across the top. To illustrate the correlation between HLA–DRB1 and these SNPs, the position of HLA–DRB1 (rs660895) is also shown.

Associations of MHC loci independent of HLA–DRB1 in ACPA-positive RA using pairwise matched analysis. There may still exist residual confounding by HLA–DRB1 SE alleles, since the categorization into 7 groups might not entirely capture all the complexity of the DRB1 locus. Therefore, we analyzed the replicated 11 SNPs using a data set with pairwise matching of cases and controls on DRB1 genotypes (both alleles). A total of 358 case–control pairs with identical DRB1 genotypes were identified in the EIRA sample and used in the conditional logistic regression analysis. The results showed that associations with markers at 10 of the 11 SNPs were statistically significant at $P < 0.05$ (MICA was marginally significant at $P = 0.0497$) and that the HLA–DQA1 locus was not statistically significant (Table 3).

We replicated our findings by including 264 case–control pairs matched for DRB1 genotype from the NARAC sample in the conditional logistic regression model and confirmed associations with 6 HLA–DPB1 SNPs (Table 3). Analysis adjusting for study by combining the pairwise-matched EIRA and NARAC data sets (622 matched case–control pairs) showed statistically significant associations with all 8 of the HLA–DPB1 SNPs and the C2-DOM3Z SNP (rs544167; $P_{combined} = 0.01$, OR 1.75 [95% CI 1.13–2.70]) (Table 3). The strongest association in DPB1 was seen for SNP rs3117213 ($P_{combined} = 0.0003$, OR 1.42 [95% CI 1.17–1.73]). The 8 strongly linked DPB1 SNPs were in linkage disequilibrium and are independent on HLA–DRB1 and HLA–DQA1. The C2-DOM3Z SNP was independent of all DPB1 SNPs (Figure 3).

Analysis of ACPA-negative RA. We did not find any SNPs that associated with ACPA-negative RA when the data were adjusted for multiple testing for the 2,131 SNPs. SNP rs2040410, which tags the HLA–DRB1*03 allele, had a P value of only 0.17 (Armitage test for

Table 3. P values and ORs of developing ACPA-positive RA in association with HLA genotypes after adjustment for HLA–DRB1 by pairwise matching on HLA–DRB1 genotype*

SNP	Position	Gene	EIRA		NARAC		Combined	
			P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)
rs2064476	33181300	HLA–DPB1	0.0025	1.49 (1.15–1.94)	0.0444	1.32 (1.01–1.74)	0.0004	1.41 (1.17–1.70)
rs3091282	33165176	HLA–DPB1	0.0029	1.49 (1.15–1.93)	0.0475	1.32 (1.00–1.73)	0.0004	1.40 (1.16–1.69)
rs9277378	33158257	HLA–DPB1	0.0030	1.50 (1.15–1.95)	0.0643	1.29 (0.98–1.70)	0.0006	1.39 (1.15–1.68)
rs3117213	33172583	HLA–DPB1	0.0069	1.44 (1.11–1.88)	0.0174	1.41 (1.06–1.86)	0.0003	1.42 (1.17–1.73)
rs3117234	33181962	HLA–DPB1	0.0127	1.44 (1.08–1.92)	0.0249	1.40 (1.04–1.87)	0.0008	1.42 (1.16–1.74)
rs2395314	33170651	HLA–DPB1	0.0094	1.42 (1.09–1.85)	0.0174	1.41 (1.06–1.86)	0.0004	1.41 (1.17–1.71)
rs3117226	33165637	HLA–DPB1	0.0142	1.41 (1.07–1.85)	0.0296	1.37 (1.03–1.81)	0.0010	1.39 (1.14–1.69)
rs3128930	33183644	HLA–DPB1	0.0032	1.53 (1.15–2.03)	0.1837	1.21 (0.91–1.62)	0.0021	1.37 (1.12–1.67)
rs544167	31998137	C2-DOM3Z	0.0017	3.46 (1.60–7.51)	0.6243	1.15 (0.65–2.03)	0.0124	1.75 (1.13–2.70)
rs2523451	31477130	MICA	0.0497	0.78 (0.61–1.00)	0.7075	1.06 (0.79–1.41)	0.2227	0.89 (0.74–1.07)
rs6457617	32771829	HLA–DQA1	0.7841	0.93 (0.54–1.59)	0.0019	2.40 (1.38–4.16)	0.0608	1.36 (0.99–1.88)

* Results are from analysis of the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) and the North American Rheumatoid Arthritis Consortium (NARAC) samples separately and combined for the 11 single-nucleotide polymorphisms (SNPs) initially identified in the EIRA study. P values and odds ratios (ORs) per copy of the risk allele were calculated by conditional logistic regression modeling. ACPA = anti–citrullinated protein antibody; RA = rheumatoid arthritis; 95% CI = 95% confidence interval.

trend). Since there have been previous reports on an association between DRB1*03 and ACPA-negative RA (14,15), we further performed a separate analysis regarding risk of ACPA-negative RA, using the EIRA data concerning an association with DRB1*03 when using conventional DRB1 genotyping. This analysis showed an OR of 1.15 (95% CI 0.54–2.44) for DRB1*03/ DRB1*03 (double-dose DRB1*03), an OR of 1.16 (95% CI 0.89–1.52) for DRB1*03/x (single-dose DRB1*03), and an OR of 1.16 (95% CI 0.90–1.51) for DRB1*03/ DRB1*03 or DRB1*03/x, using the x/x (no DRB1*03 alleles) as the reference group and adjusting for age, sex, and geographic location. Thus, there was a trend toward an association between DRB1*03 and ACPA-negative RA also in our data, but the association did not reach statistical significance.

DISCUSSION

Our study suggests that at least 1 independent locus from the classic HLA–DRB1 locus in the MHC region (HLA–DPB1) contributes significantly to the risk of ACPA-positive RA. A similar finding related to the DPB1 locus in ACPA-positive RA was made independently of the present study in a North American Caucasian population (24). Our study provides an indication that there may also be an independent effect of HLA–DQ, but this requires further study in even larger populations. The C2-DOM3Z locus showed suggestive evidence of association, although not yet fully replicated in the NARAC sample using the matching technique. Finally, with regard to clinical subgroups of RA, we did not identify any associations with SNPs within the MHC for ACPA-negative RA.

Two studies have previously reported an association of DRB1*03 with ACPA-negative RA (14,15). Since the SNPs used in our study do not identify specific HLA–DRB1 alleles (such as DRB1*03), we performed an analysis to investigate the relationship between DRB1*03 and ACPA-negative RA using the DRB1 allele data in the EIRA study. Despite the observed trend for an association between DRB1*03 and the risk of ACPA-negative RA, this association was not significant in our data. We must acknowledge, however, that our data, as well as the previous data showing a positive association, are still of limited size, and even larger studies will be needed to resolve the question of whether and in which populations DRB1*03 is associated with RA and, in particular, with ACPA-negative RA. More detailed analyses and data concerning DRB1 allele distribution and association with the 2 disease subsets of

RA using the EIRA data have been described in a separate article (Lundström E, et al: submitted for publication).

To test the robustness of our findings, we also analyzed the relationship between all the primarily identified 299 SNPs and the risk of RA in a matched manner, with case–control pairs matched according to DRB1 genotype. In terms of bias, the matched analysis is the optimal way to adjust for confounding by DRB1 genotype, but it will often be at the expense of power, since the number of useful subject pairs in the analysis often is relatively small. Using this strategy in the EIRA sample, we found 47 SNPs that were statistically significant at $P < 0.05$. The NARAC study had access to data on 24 of these 47 SNPs, and the EIRA results were confirmed in 6 of these. All 6 of these SNPs were identical to those identified from our first strategy using unconditional logistic regression. We considered the possibility that other non-SE alleles of DRB1 might be in linkage disequilibrium with HLA–DPB1 alleles and might potentially explain the association between DPB1 alleles and susceptibility to ACPA-positive RA. The pairwise matching process, however, eliminated this risk and demonstrated that DPB1 variations are independently associated with the risk of ACPA-positive RA.

In order to remove the confounding by DRB1 SE alleles as much as possible without losing sufficient statistical power, we classified the DRB1 alleles into different groups (defined here as models 2 and 3) based on a statistically oriented approach, in which the OR was used as the basis for grouping the DRB1 alleles (see Patients and Methods and Table 1 for details). Another approach to classifying DRB1 alleles according to their amino acid sequences at positions 70–74 was proposed by Tezenas du Montcel and colleagues (19). The Tezenas du Montcel model has been cross-validated (25,26) and shown to be superior to other classification systems for SE alleles in predicting radiologic progression to erosive disease and in supporting the identification of a protective effect on disease progression (27,28). We compared this approach with our models and found that it yielded results that were very similar to those of our best model, model 3, with 71 significant SNPs for the Tezenas du Montcel model versus 70 significant SNPs for our model 3; among these 70 SNPs, 61 were in common with those in our model 3, as compared with 62 SNPs in common among our 3 models. Therefore, we think our model 3 is, in the present context, similar to the Tezenas du Montcel model in terms of controlling for confounding by HLA–DRB1 risk alleles. For exploratory purposes, we also performed an unconditional

logistic regression analysis using recursive partitioning, an alternative risk-categorization method of grouping (29), in the EIRA data, which replicated our findings very well.

These findings support a model in which the role of MHC-dependent adaptive immunity in a broad manner is restricted only or mainly to ACPA-positive RA and in which at least 2, and possibly more, different class II MHC loci (HLA–DR, HLA–DQ, and HLA–DP) may be involved in the pathogenesis of the ACPA-positive RA. Our observations may help to open up the field for new immunologic studies aiming to determine how different MHC-restricted immune reactions, tentatively directed toward citrullinated peptides and proteins, may be involved in the pathogenesis of a serologically defined subset of RA (11,30).

AUTHOR CONTRIBUTIONS

Dr. Ding had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study design. Ding, Padyukov, Gregersen, Alfredsson, Klareskog.
Acquisition of data. Ding, Padyukov, Lundström, Seielstad, Gregersen, Alfredsson, Klareskog.
Analysis and interpretation of data. Ding, Padyukov, Lundström, Seielstad, Plenge, Oksenberg, Gregersen, Alfredsson, Klareskog.
Manuscript preparation. Ding, Padyukov, Lundström, Seielstad, Plenge, Oksenberg, Gregersen, Alfredsson, Klareskog.
Statistical analysis. Ding, Padyukov, Alfredsson.

REFERENCES

1. Gregersen PK, Behrens TW. Genetics of autoimmune diseases—disorders of immune homeostasis. Nat Rev Genet 2006;7:917–28.

2. Klareskog L, Padyukov L, Ronnelid J, Alfredsson L. Genes, environment and immunity in the development of rheumatoid arthritis. Curr Opin Immunol 2006;18:650–5.

3. Schellekens GA, de Jong BA, van den Hoogen FH, van de Putte LB, van Venrooij WJ. Citrulline is an essential constituent of antigenic determinants recognized by rheumatoid arthritis-specific autoantibodies. J Clin Invest 1998;101:273–81.

4. Avouac J, Gossec L, Dougados M. Diagnostic and predictive value of anti-cyclic citrullinated protein antibodies in rheumatoid arthritis: a systematic literature review. Ann Rheum Dis 2006;65:845–51.

5. Stastny P, Fink CW. HLA-Dw4 in adult and juvenile rheumatoid arthritis. Transplant Proc 1977;9:1863–6.

6. 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.

7. Mulcahy B, Waldron-Lynch F, McDermott MF, Adams C, Amos CI, Zhu DK, et al. Genetic variability in the tumor necrosis factor-lymphotoxin region influences susceptibility to rheumatoid arthritis. Am J Hum Genet 1996;59:676–83.

8. Ota M, Katsuyama Y, Kimura A, Tsuchiya K, Kondo M, Naruse T, et al. A second susceptibility gene for developing rheumatoid arthritis in the human MHC is localized within a 70-kb interval telomeric of the TNF genes in the HLA class III region. Genomics 2001;71:263–70.

9. Jawaheer D, Li W, Graham RR, Chen W, Damle A, Xiao X, et al. Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. Am J Hum Genet 2002;71:585–94.

10. Newton JL, Harney SM, Wordsworth BP, Brown MA. A review of the MHC genetics of rheumatoid arthritis. Genes Immun 2004;5:151–7.

11. Klareskog L, Stolt P, Lundberg K, Kallberg H, Bengtsson C, Grunewald J, et al, and the Epidemiological Investigation of Rheumatoid Arthritis Study Group. 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.

12. Huizinga TW, Amos CI, van der Helm-van Mil AH, Chen W, van Gaalen FA, Jawaheer D, et al. Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA–DRB1 shared epitope for antibodies to citrullinated proteins. Arthritis Rheum 2005;52:3433–8.

13. Van der Helm-van Mil AH, Verpoort KN, Breedveld FC, Huizinga TW, Toes RE, de Vries RR. The HLA–DRB1 shared epitope alleles are primarily a risk factor for anti–cyclic citrullinated peptide antibodies and are not an independent risk factor for development of rheumatoid arthritis. Arthritis Rheum 2006;54:1117–21.

14. Verpoort KN, van Gaalen FA, van der Helm-van Mil AH, Schreuder GM, Breedveld FC, Huizinga TW, et al. Association of HLA-DR3 with anti–cyclic citrullinated peptide antibody–negative rheumatoid arthritis. Arthritis Rheum 2005;52:3058–62.

15. 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 Rheum 2005;52:3813–8.

16. 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.

17. 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.

18. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF,

Ding B, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. N Engl J Med 2007;357:1199–209.

19. Tezenas du Montcel S, Michou L, Petit-Teixeira E, Osorio J, Lemaire I, Lasbleiz S, et al. New classification of HLA–DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. Arthritis Rheum 2005;52:1063–8.

20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559–75.

21. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.

22. Kallberg H, Padyukov L, Plenge RM, Ronnelid J, Gregersen PK, van der Helm-van Mil AH, et al. Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. Am J Hum Genet 2007;80: 867–75.

23. Van der Helm-van Mil AH, Verpoort KN, le Cessie S, Huizinga TW, de Vries RR, Toes RE. 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.

24. Lee HS, Lee AT, Criswell LA, Seldin MF, Amos CI, Carulli JP, et al. Several regions in the major histocompatibility complex confer risk for anti-CCP-antibody positive rheumatoid arthritis, independent of the DRB1 locus. Mol Med 2008;14:293–300.

25. Michou L, Croiseau P, Petit-Teixeira E, Tezenas du Montcel S, Lemaire I, Pierlot C, et al, for the European Consortium on Rheumatoid Arthritis Families. Validation of the reshaped shared epitope HLA-DRB1 classification in rheumatoid arthritis. Arthritis Res Ther 2006;8:R79.

26. Barnetche T, Constantin A, Cantagrel A, Cambon-Thomsen A, Gourraud PA. New classification of HLA-DRB1 alleles in rheumatoid arthritis susceptibility: a combined analysis of worldwide samples. Arthritis Res Ther 2008;10:R26.

27. Gourraud PA, Boyer JF, Barnetche T, Abbal M, Cambon-Thomsen A, Cantagrel A, et al. A new classification of HLA–DRB1 alleles differentiates predisposing and protective alleles for rheumatoid arthritis structural severity. Arthritis Rheum 2006;54:593–9.

28. Winchester R. Reshaping Cinderella's slipper: the shared epitope hypothesis. Arthritis Res Ther 2006;8:109.

29. Nejentsev S, Howson JM, Walker NM, Szeszko J, Field SF, Stevens HE, et al, Wellcome Trust Case Control Consortium. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature 2007;450:887–92.

30. Holers VM. Are anti-cyclic citrullinated peptide antibodies pathogenic in rheumatoid arthritis? Nat Clin Pract Rheumatol 2006;2: 400–1.