

EXTENDED REPORT

Antibodies against cyclic citrullinated peptides of IgG, IgA and IgM isotype and rheumatoid factor of IgM and IgA isotype are increased in unaffected members of multicase rheumatoid arthritis families from northern Sweden

Lisbeth Ärlestig,¹ Mohammed Mullazehi,² Heidi Kokkonen,¹ Joacim Rocklöv,¹ Johan Rönnelid,² Solbritt Rantapää Dahlqvist¹

► An additional figure is published online only. To view the files, please visit the journal online (<http://ard.bmj.com/content/71/6.toc>).

¹Departments of Public Health and Clinical Medicine/ Rheumatology and Epidemiology and Global Medicine, Umeå University, Umeå, Sweden
²Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

Correspondence to

Solbritt Rantapää-Dahlqvist, Department of Public Health and Clinical Medicine, Rheumatology, University Hospital, SE-901 85 Umeå, Sweden; solbritt.rantapaa.dahlqvist@medicin.umu.se

Received 25 August 2011
Accepted 21 October 2011
Published Online First
29 November 2011

ABSTRACT

Background Rheumatoid factors (RFs) and antibodies against cyclic citrullinated peptides (CCPs) of IgG, IgA and IgM isotype have been shown to precede disease onset by years.

Objective To evaluate serological risk markers in first-degree relatives from multicase families in relation to genetic and environmental risk factors.

Methods 51 multicase families consisting of 163 individuals with rheumatoid arthritis (RA) (mean \pm SD age, 60 \pm 14 years; disease duration 21 years; 71.8% female) and with 157 first-degree relatives unaffected by RA (54 \pm 17 years; 59.9% female) were recruited. Isotypes of antibodies against CCPs (IgG, IgA and IgM) and RFs (IgM and IgA) were determined using automated enzyme immunoassays. Cut-off levels were established using receiver operating characteristic curves based on values for 100 unrelated healthy controls.

Results The concentrations and frequencies of all anti-CCP and RF isotypes were significantly increased in first-degree relatives and patients with RA compared with unrelated healthy controls. The relative distribution of IgA and IgM isotypes was higher than IgG in the relatives, whereas the IgG isotype dominated in patients with RA. The patients carried human leucocyte antigen-shared epitope (HLA-SE) significantly more often than the relatives (71.4% vs 53.9%, $p=0.01$), while the frequency of the *PTPN22* T variant was similar. HLA-SE, combined with smoking, was significantly related to all combinations of anti-CCP and RF isotypes in patients with RA. No such relationships were found for the first-degree relatives.

Conclusions All anti-CCP and RF isotypes analysed occurred more commonly in unaffected first-degree relatives from multicase families than in controls, but with different isotype distribution from patients with RA.

INTRODUCTION

A genetic component of susceptibility to rheumatoid arthritis (RA) has long been suggested by data from twin and familial studies.^{1,2} Twin studies have estimated the inheritability of RA to be between 50% and 60%,³ although the role of genetic factors has been questioned.⁴ Consistent associations implicate a role for genes located in the human leucocyte antigen (HLA) region in the risk of developing RA.⁵ Linkage studies support a role for the

HLA locus in genetic susceptibility to RA.⁶ A large number of genes have, in genome-wide association studies, been particularly associated with anti-citrullinated protein/peptide antibody (ACPA) positive patients with RA.^{7,8} Another robust association with RA is the R620W polymorphism of the *PTPN22* locus.^{9,10} The association between the T variant of the *PTPN22* gene and RA was more evident in patients seropositive for rheumatoid factors (RFs) and/or ACPAs.^{11–13} We have previously shown that a combination of antibodies against CCP with either HLA-shared epitope (SE) or the T variant of *PTPN22* analysed before the onset of symptoms of disease strongly predicted future onset of RA with a high relative risk of developing RA.^{14,15}

Smoking is one of the aetiological factors identified as a risk factor for RA.¹⁶ Data from a number of studies have suggested that a combination of HLA-SE alleles and exposure to tobacco interact in the development of ACPA-positive RA.^{17,18}

In addition to the previously identified IgG isotype, we have shown that ACPAs of either the IgA or IgM isotype predate the onset of RA by a number of years.^{19,20} All three isotypes were also significantly increased after development of the disease. In addition, ACPAs have been shown to contribute to progression of arthritis in collagen-induced arthritis in mice, appearing 7 days after immunisation before clinical disease.²¹ Taken together, these results suggest a pathogenic role for ACPAs in the development of RA.

In a previous study of native North Americans with a high prevalence of severe RA, ACPAs of different isotypes were present with increased frequency in the unaffected relatives of that cohort.²² In another study of first-degree relatives of patients with RA, the frequency of RFs and/or anti-CCP IgG increased to 16%.²³

The aim of the present study was to analyse the isotypes of ACPAs and RFs in unaffected first-degree relatives of patients with RA from multicase families from northern Sweden in relation to HLA-SE alleles and *PTPN22* polymorphism and smoking habits.

MATERIALS AND METHODS

Familial clustering of RA was identified using a questionnaire distributed to patients with RA attending



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://ard.bmj.com/info/unlocked.dtl>

Table 1 Concentration of the different isotypes of CCP2 antibodies (AU/ml) and RFs (IU/ml) in healthy controls, patients with RA and first-degree relatives

	IgG-CCP2	IgA-CCP2	IgM-CCP2	IgM-RF	IgA-RF
Controls (n=100)	1.5	0.6	18.5	0.8	3.4
IQR	(1.0–2.2)	(0.4–0.7)	(9.6–30.9)	(0.6–1.2)	(2.7–4.7)
First-degree relatives (n=157)	2.2***	1.0***	27.9***	1.3***	4.7***
IQR	(1.4–3.1)	(0.7–1.5)	(13.7–53.3)	(0.9–2.4)	(3.4–8.0)
RA patients (n=163)	235.0***	2.8***	51.7***	29.9***	24.6***
IQR	(27.0–569.0)	(1.3–13.0)	(22.5–140.0)	(4.6–134.3)	(8.8–60.3)
p Value [†]	<0.001	<0.001	<0.001	<0.001	<0.001

Values are median (Q1–Q3).

***p<0.001, refers to post-hoc comparisons between first-degree relatives or RA patients versus controls.

[†]Kruskal–Wallis test for three independent samples.

CCP, cyclic citrullinated peptide; RA, rheumatoid arthritis; RF, rheumatoid factor.

departments of rheumatology in the four northern-most counties of Sweden. All of the reported relatives from families who wished to participate were interviewed, by a second questionnaire, about symptoms and signs of joint disease. Affected relatives were evaluated clinically by a rheumatologist and inspection of their medical records, and personal interviews were used to confirm the diagnosis of RA (defined by the ACR 1987 criteria).²⁴ Sixty-one multicase families, with a total of 196 individuals with RA, were willing to participate. From 51 of these families, 157 first-degree relatives (59.9% female) who reported no symptoms of inflammatory joint disease on a questionnaire evaluated by us were willing to donate blood samples. These 51 families included between two and 11 affected members with a total of 163 RA cases; 71.8% were female. Of the patients with RA, 62.6% had erosive disease and 21.5% had extra-articular manifestations.

IgG, IgA and IgM anti-CCP isotypes were determined in plasma using EliA anti-CCP assay on an ImmunoCAP 250 instrument (Phadia Diagnostic AB, Uppsala, Sweden), according to the manufacturer's instructions and as previously described.²⁰ All samples above the upper limit were diluted further to obtain precise values. Within-study reference ranges were defined by receiver operating characteristic (ROC) curves, based on the patients with RA and healthy controls (n=100), whereby the optimal cut-off was defined as the antibody level yielding the highest sum of sensitivity and specificity. The cut-offs determined were 3.3 AU/ml for IgG, 1.41 AU/ml for IgA and 59.9 AU/ml for IgM. Repeated analysis of individual samples during the study period showed stable values. Cut-off values for RF IgA and IgM were defined by ROC curves to be 9.05 IU/ml and 4.8 IU/ml, respectively.

Genomic DNA was extracted using standard methods and analysed for a single-nucleotide polymorphism in *PTPN22* 1858C/T (rs2476601) with a 5'-nuclease assay according to the manufacturer's instructions (Applied Biosystems, Foster City, California, USA). The detection protocol used a 7900 HT Sequence Detection System (Applied Biosystems) in combination with SDS2.1 software (Applied Biosystems) to perform the genotyping. HLA-DRB1* analysis was performed using a DR low-resolution kit and DRB1*04 subtyping kit (Olerup SSP AB, Saltsjöbaden, Sweden). The HLA-SE was defined by DRB1*0401, 0404 and 0408. Controls for genetic analyses were 193 people (age- and sex-matched for the patients with RA) identified from the Medical Biobank, northern Sweden.

The regional ethics committee at the University Hospital, Umeå, Sweden approved the study, and all participants gave their written informed consent.

Statistical analysis

The χ^2 test was used for categorical data, and non-parametric analysis for continuous data. Binary conditional logistic

regression models were used to estimate predictive variables for the presence of RA in the families. Statistical analysis was performed using SPSS V.18.0 software. To allow distinction between first-degree relatives and patients with RA, a multivariate classification algorithm termed the Random Forest method was used.²⁵ Discrimination of the Random Forest models can be visualised graphically by using multidimensional scaling of the proximity matrix. The proximity is a relative measure of pair-wise relations derived from the set of decision trees in the Random Forest method. The prediction classification performance of the Random Forest model can be translated to sensitivity and specificity. Principal component scores were computed for the predictive variables to assess their multivariate relationship. The principal component method combines the information in several correlated variables and produces a few new independent scores based on linear combinations of the risk factors. It has the advantage of explaining most of the variability in the original data.

RESULTS

The mean±SD age of the family members with RA was significantly higher than that of their relatives (60±14 years and 54±17 years, respectively). They had a mean disease duration of 21±15.6 years. The frequency of current or past smoking habits was higher among the patients with RA (57.9% vs 46.9%, respectively), but this difference was not significant. The frequency of HLA-DRB1*0401/0404/0408 was 53.9% in first-degree relatives and 71.4% in patients with RA (p<0.01); both of these frequencies were significantly increased compared with that of the control population (35.8%). Of the patients with RA, 12.8% had double SE alleles compared with 1.3% of the relatives (p<0.001) and 5.6% of the controls. The *PTPN22* T variant was significantly more prevalent in both first-degree relatives (46.8%) and patients with RA (47.8%) compared with controls (19.5%).

The concentrations of ACPA isotypes IgG, IgM and IgA and RF isotypes IgM and IgA were significantly increased in first-degree relatives compared with controls, and were further increased in patients with RA (table 1). These concentrations were also significantly increased in patients with RA compared with their first-degree relatives (p<0.001 for all isotypes). High concentrations (defined as three times above cut-off) of ACPA isotypes IgM, IgA and IgG were found in 5.1%, 4.5% and 5.7%, respectively, of first-degree relatives compared with 21.5%, 42.9% and 81.0% of patients with RA (online supplementary figure S1A–C).

In the first-degree relatives, the frequency of the different ACPA isotypes was 21.7% for IgG, 26.8% for IgA and 22.3% for IgM, with specificities of 98%, 99% and 94%, respectively

Table 2 Frequency and diagnostic sensitivity and specificity of ACPA isotypes and RF isotypes in first-degree relatives and RA patients from multicase families

	Frequency	Sensitivity	Specificity
	First-degree relatives, n (%)	Patients with RA, n (%)	%
IgG anti-CCP	34 (21.7)***	140 (85.9)***	98
IgA anti-CCP	42 (26.8)***	118 (72.4)***	99
IgM anti-CCP	35 (22.3)***	74 (45.4)***	94
IgM-RF	22 (14.0)*	121 (74.2)***	95
IgA-RF	35 (22.3)***	123 (75.5)***	94

* $p < 0.05$, *** $p < 0.001$, patients with RA and first-degree relatives versus healthy controls.

ACPA, anti-citrullinated protein/peptide antibody; CCP, cyclic citrullinated peptide; RA, rheumatoid arthritis; RF, rheumatoid factor.

(table 2). All of the ACPA isotypes were significantly increased compared with healthy controls. The frequency of the IgM-RF isotype was increased significantly to 14.0%, and that of IgA-RF to 22.3% at specificities of 95% and 94%, respectively. The sensitivity for all ACPA and RF isotypes was highly significant in the patients with RA compared with controls (table 2).

The numbers of ACPA isotypes identified in the first-degree relatives, patients with RA and controls were significantly different (Kruskal–Wallis, $p < 0.0001$, between the three groups; Mann–Whitney, $p < 0.001$ for both relatives vs RA patients and vs controls; data not shown). Sixty-five (38.7%) of the patients with RA were positive for all three isotypes, as were six (7.1%) of the symptom-free first-degree relatives, while 49 (31.2%) of the latter group showed the presence of one isotype. The distribution of each isotype and their combinations presented in figure 1 show that the number of isolated IgA- and IgM-positive ACPA isotypes in relation to isolated IgG isotype were significantly higher in first-degree relatives than in patients with RA ($p < 0.0001$ for each type). There were no significant differences in the distribution of RFs in first-degree relatives compared with RA patients (figure 1).

Stratification of the patients with RA and their first-degree relatives for ACPA and RF status showed, in simple logistic regression analyses, that there was a significant relationship between HLA-SE and IgG isotype of ACPAs in the patients with RA (table 3). There were significant relationships between smoking and the IgM isotype and the combination of all three ACPA isotypes, and with IgA-RF and a combination of both RFs measured. However, the combination of HLA-SE and current or past smoking habits was related to all three ACPA isotypes both separately and in combination, and also in combination with RFs, while none of these associations were found in the first-degree relatives (table 3). In fact, all patients with RA with SE and being current or past smokers were positive for the ACPA IgG isotype. To permit calculation, one of these patients was recorded as being IgG-ACPA negative. There were no relationships between any of the isotypes and *PTPN22* polymorphism in patients with RA or their first-degree relatives (data not shown).

The ACPA isotypes (with the highest OR for the IgG isotype followed by IgA), RF isotypes, HLA-SE and age all differentiated, in simple conditional logistic regression analyses, patients with RA from first-degree relatives ($p < 0.0001$ – 0.05), while ever-smoking, T variant of *PTPN22* and sex did not (table 4). In a multiple conditional backwards regression analysis including all of the significant variables from the simple regression analyses, IgG ACPAs (OR 3.46, 95% CI 1.94 to 6.18, $p < 0.0001$) and IgM RF (OR 1.71, 95% CI 1.06 to 2.76, $p < 0.05$) were the remaining significant variables that distinguished patients with RA from first-degree relatives within the families. The OR was higher for

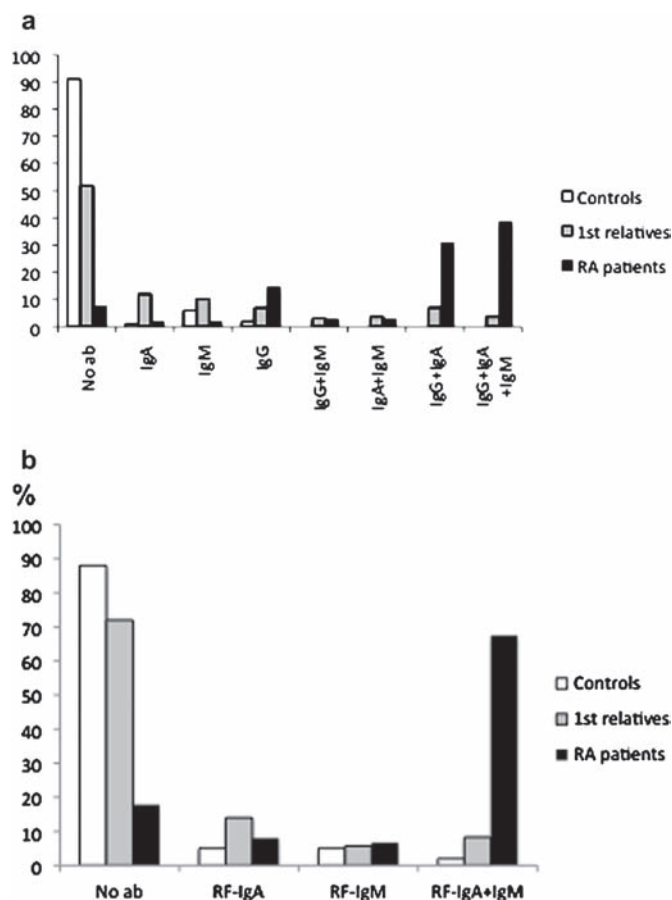


Figure 1 (A) Distribution of each anti-citrullinated protein/peptide antibody (ACPA) isotype and combinations in controls, first-degree relatives and patients with rheumatoid arthritis (RA). The number of isolated IgA- and IgM-positive ACPA isotypes in relation to isolated IgG isotype was significantly higher in first-degree relatives than in patients with RA ($p < 0.0001$ for each type). (B) Distribution of each rheumatoid factor isotype in controls, first-degree relatives and patients with RA. No ab, no antibody.

presence of IgG ACPAs than for presence of IgM-RF, but stratification for either antibody did not give consistent results for an association with ACPA irrespective of seropositivity for IgM-RF (data not shown). Principal component analysis including the isotypes of ACPA and RF, HLA-SE and T variant of *PTPN22*, smoking habits, age and sex showed that IgG-ACPA isotype and IgM-RF correlated highly, and these two variables alone can describe 98% of the information from all factors together.

Random Forest modelling including the concentrations of ACPA and RF isotypes, genes, smoking habits and also age and sex gave a sensitivity of 90.7% for having RA, with a specificity of 81.8% for discriminating between first-degree relatives and patients with RA. A plot illustrating the relationships between the patients and the relatives according to the discriminating variables (ACPA, RF isotypes, genes, smoking habits, age and sex) is presented in figure 2. Thus observations that are close together are more closely related than those that are further apart. The first dimension of the multidimensional scaling variable shows the best discrimination according to the proximities.

An individual summation of these eight factors (ie, ACPA and RF isotypes, HLA-SE, *PTPN22* T variant and smoking habits) as risk markers showed that patients with RA had a median (Q1–Q3) number of 6 (4–7) of these factors, whereas their relatives had 2 (1–3) (Mann–Whitney U test, $p < 0.001$).

Table 3 Risk of antibody-positivity predicted by HLA-SE, smoking and HLA-SE+smoking in patients with RA and their first-degree relatives

Antibody	Patients with RA			First-degree relatives		
	SE	Smoking ever	SE+smoking ever	SE	Smoking ever	SE+smoking ever
IgG-ACPA	4.71 (1.84 to 12.05)	2.17 (0.90 to 5.23)	19.33 (2.54 to 147.21)*	0.55 (0.25 to 1.21)	1.37 (0.63 to 2.95)	0.71 (0.28 to 1.81)
IgA-ACPA	2.03 (0.96 to 4.30)	1.89(0.94 to 3.79)	2.81 (1.27 to 6.24)	0.47 (0.22 to 0.97)	1.36 (0.65 to 2.85)	0.76 (0.31 to 1.86)
IgM-ACPA	1.30 (0.64 to 2.63)	2.33(1.21 to 4.48)	2.17 (1.07 to 4.00)	0.76 (0.35 to 1.64)	0.79 (0.36 to 1.73)	0.79 (0.31 to 2.02)
IgG+IgA+IgM-ACPA	1.78 (0.84 to 3.77)	2.39 (1.21 to 4.71)	2.98 (1.51 to 5.89)	0.20 (0.02 to 1.87)	1.14 (0.22 to 5.82)	0.71 (0.08 to 6.48)
IgM-RF	1.85 (0.86 to 4.0)	1.76 (0.87 to 3.58)	2.45 (1.10 to 5.46)	0.75 (0.30 to 1.88)	0.75 (0.30 to 1.88)	0.43 (0.12 to 1.54)
IgA-RF	1.36 (0.62 to 2.96)	2.31 (1.11 to 4.78)	2.23 (0.99 to 4.98)	1.11 (0.51 to 2.38)	1.61 (0.73 to 3.55)	1.21 (0.50 to 2.94)
IgM+IgA-RF	1.47 (0.71 to 3.05)	2.16 (1.11 to 4.23)	2.53 (1.21 to 5.28)	0.58 (0.18 to 1.93)	0.97 (0.31 to 3.03)	0.24 (0.03 to 1.91)

Values are OR (95% CI).

*All ever-smoking patients with RA with SE had IgG isotype antibodies. To permit this calculation, one of these patients was recorded as IgG negative. ACPA, anti-citrullinated protein/peptide antibody; HLA, human leucocyte antigen; RA, rheumatoid arthritis; SE, shared epitope.

Table 4 Results of simple conditional logistic regression analyses of the different isotypes of anti-CCP2 antibodies, RFs, HLA-SE, *PTPN22* 1858T and smoking with first-degree relatives and patients with RA as dependent variable

Factor	OR (95%CI)	p Value
IgG-CCP	5.80 (3.61 to 9.30)	<0.0001
IgA-CCP	2.47 (1.74 to 3.51)	<0.0001
IgM-CCP	1.46 (1.07 to 1.99)	<0.05
IgA-RF	3.12 (2.16 to 4.50)	<0.0001
IgM-RF	3.64 (2.50 to 5.29)	<0.0001
HLA-SE	1.56 (1.00 to 2.42)	<0.05
<i>PTPN22</i> 1858T	1.23 (0.83 to 1.82)	NS
Smoking	1.28 (0.92 to 1.78)	NS
Age	1.02 (1.01 to 1.03)	<0.01
Sex	1.28(0.92 to 1.79)	NS

CCP, cyclic citrullinated peptide; RA, rheumatoid arthritis; RF, rheumatoid factor; HLA-SE, human leucocyte antigen shared epitope.

DISCUSSION

In this study of patients with RA and their first-degree asymptomatic relatives from multicase families, we found significantly increased concentrations of both ACPA and RF isotypes compared with healthy controls. The frequencies of these isotypes were also significantly increased, based on ROC curves constructed using data from patients with RA and healthy controls. Our analyses detected a surprisingly high frequency of ACPA isotypes in symptom-free first-degree relatives. The prevalence of the IgG isotype of ACPA in our study was slightly higher than in studies from the USA and Canada despite a number of the first-degree relatives in the former study having swollen and/or tender joints.^{22 23} The frequencies of RF isotypes were similar to those in the Canadian study (15.4% for IgM-RF and 21.6% for IgA-RF), but slightly higher than in the US study (4.7% and 3.8%, respectively).^{22 23} As we had also analysed the two genes most commonly related to RA and collected data on smoking habits in most of the subjects, we were able to compare these asymptomatic relatives with their family members who had RA with respect to these factors. We were unable to relate the presence of ACPA or RF isotypes in the first-degree relatives to either of the two gene polymorphisms, to smoking habits or to demographic data such as age or sex. We could not confirm that HLA-SE alleles were a risk factor for development of ACPAs in the first-degree relatives, which was the case for the IgG isotype in the patients with RA in this study and as has been suggested by others.²⁶ All patients with RA who were SE positive and had ever smoked had ACPAs of the IgG isotype. However,

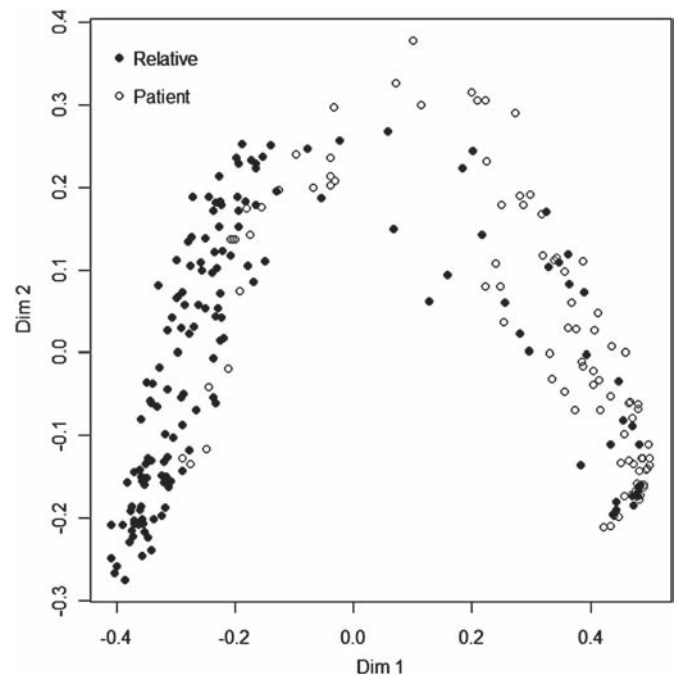


Figure 2 Visualisation of multidimensional scaling demonstrating clustering of first-degree relatives and patients with rheumatoid arthritis according to a set of predictive variables in a Random Forest model. The two axes represent the dominant clustering directions between the groups scaled as proximities. Dim, dimension.

our findings on RFs are consistent with the results presented by MacGregor *et al* in 1995,³ who found in twin studies that seropositivity for RF had a genetic contribution independent of HLA-DR. Also, in another large study of 1058 first-degree relatives, no significant association between autoantibodies and the presence of HLA-SE or *PTPN22* polymorphism was found.²³

The lack of association found in this study could also be a result of too few individuals being antibody and HLA-SE positive compared with RA patients, making it more difficult to show relationships. In a number of recent studies, genetic factors have been identified in the HLA region that protect against the development of RA.²⁷ In particular, the presence of HLA-DRB1*1301/1302 has been identified as being protective.²⁷ In the present study, we were unable to analyse for HLA-DRB1 alleles associated with protective effects on the development of ACPAs or RA, although this could be a tentative explanation of why the relatives with SE and/or *PTPN22* T allele and ACPAs, with predominantly immature isotypes, have not developed RA.

The sensitivity of all ACPA isotypes was higher for patients with RA in this study than in our previous study of patients with early RA, when we used the same laboratory procedure and ROC curve calculations to determine cut-off values, in which the sensitivity was 71.7% for IgG, 43.3% for IgA and 32.8% for IgM.²⁰ In that study the sensitivity for the IgG isotype was higher (33.8%) among individuals (referred to as prepatients) who subsequently developed RA compared with the first-degree relatives of patients in the present study. The frequency for the IgM isotype was higher among first-degree relatives than among prepatients (22.4% vs 11.8%, $\chi^2=3.69$, $p=0.058$). For the IgA isotype, the frequencies were similar (26.9% vs 23.9%) compared with prepatients. The number of people ($n=156$) providing samples for the present study was double that of the previous study based on samples predating disease onset ($n=71$). The median predating time for the prepatients was 2.5 years before they developed symptoms of RA disease, while in the present study, we do not yet know if the first-degree relatives will ever develop RA. One argument against the later appearance of RA in at least some of the autoantibody-positive first-degree relatives is the age distribution. The mean age of the patients with RA was 60 years compared with 54 years for the first-degree relatives. However, as the mean disease duration was 21 years, the patients with RA had a mean age at diagnosis of 39 years—that is, 15 years younger than their relatives included in the study.

An interesting observation among the first-degree relatives was the relative increase in IgA and IgM antibodies compared with ACPAs of the IgG isotype. This antibody pattern differed distinctly from that in the patients with RA. One can speculate whether this is a predisease pattern that will change when the disease develops or represents an abortive form of autoimmunity that will never lead to a switch to the IgG isotype traditionally associated with RA. The fact that we found a higher percentage of IgM anti-CCP positive subjects among the first-degree relatives in the present study compared with among prepatients in our earlier study²⁰ argues for the latter possibility, as does the fact that our group of symptom-free relatives were 15 years older at inclusion than their relatives with RA when they were diagnosed.

In conclusion, in this study on multicase families with RA, first-degree asymptomatic relatives have increased concentrations of all ACPA and RF isotypes without any significant relationship of these antibodies to HLA-SE, *PTPN22* polymorphism or smoking habits. The presence of protective factors may prevent these relatives from developing the disease, but a later disease onset should also be considered a possibility, which could be investigated with appropriate follow-up protocols.

Contributors LÅ was the main investigator, identified the patients with RA and the first-degree relatives and arranged for blood sampling, carried out the statistical analysis, and contributed to preparation of the manuscript. MM established the IgA and IgM anti-CCP and RF analyses, carried out and interpreted all anti-CCP antibody analyses on the Phadia platform HK, and performed the genetic analyses. JR is the statistician and performed the Random Forest analyses and the principal component analyses. JR participated in design of the study, was responsible for the anti-CCP antibody and RF analyses, and participated in finalising the manuscript. SRD is the principal investigator, responsible for the patients, designed the investigation and participated in data collection, statistical analysis and drafting of the manuscript. All authors have read and approved the final manuscript.

Funding This study was supported by grants from the Swedish Research Council (K2010-52X-20307-04-3), the Swedish Rheumatism Association and King Gustav Vth 80-Year Foundation, the Groschinsky foundation and by research funding from the European Community FP6 funding project 018661 'Autocure'.

Competing interests None.

Patient consent Obtained.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- Aho K, Koskenvuo M, Tuominen J, *et al*. Occurrence of rheumatoid arthritis in a nationwide series of twins. *J Rheumatol* 1986;**13**:899–902.
- Silman AJ, MacGregor AJ, Thomson W, *et al*. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. *Br J Rheumatol* 1993;**32**:903–7.
- MacGregor AJ, Snieder H, Rigby AS, *et al*. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000;**43**:30–7.
- Svendsen AJ, Holm NV, Kyvik K, *et al*. Relative importance of genetic effects in rheumatoid arthritis: historical cohort study of Danish nationwide twin population. *BMJ* 2002;**324**:264–6.
- 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.
- Etzel CJ, Chen WW, Shepard N, *et al*. Genome-wide meta-analysis for rheumatoid arthritis. *Hum Genet* 2006;**119**:634–41.
- De Rycke L, Peene I, Hoffman IE, *et al*. Rheumatoid factor and anticitrullinated protein antibodies in rheumatoid arthritis: diagnostic value, associations with radiological progression rate, and extra-articular manifestations. *Ann Rheum Dis* 2004;**63**:1587–93.
- van der Helm-van Mil AH, Huizinga TW. Advances in the genetics of rheumatoid arthritis point to subclassification into distinct disease subsets. *Arthritis Res Ther* 2008;**10**:205.
- Begovich AB, Carlton VE, Honigberg LA, *et al*. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (*PTPN22*) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004;**75**:330–7.
- Gregersen PK. Gaining insight into *PTPN22* and autoimmunity. *Nat Genet* 2005;**37**:1300–2.
- Lee HS, Li W, Lee A, *et al*. Microsatellite typing for DRB1 alleles: application to the analysis of HLA associations with rheumatoid arthritis. *Genes Immun* 2006;**7**:533–43.
- Plenge RM, Padyukov L, Remmers EF, *et al*. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with *PTPN22*, *CTLA4*, and *PADI4*. *Am J Hum Genet* 2005;**77**:1044–60.
- Kokkonen H, Johansson M, Innala L, *et al*. The *PTPN22* 1858C/T polymorphism is associated with anti-cyclic citrullinated peptide antibody-positive early rheumatoid arthritis in northern Sweden. *Arthritis Res Ther* 2007;**9**:R56.
- Berglin E, Padyukov L, Sundin U, *et al*. A combination of autoantibodies to cyclic citrullinated peptide (CCP) and HLA-DRB1 locus antigens is strongly associated with future onset of rheumatoid arthritis. *Arthritis Res Ther* 2004;**6**:R303–8.
- Johansson M, Arlestig L, Hallmans G, *et al*. *PTPN22* polymorphism and anti-cyclic citrullinated peptide antibodies in combination strongly predicts future onset of rheumatoid arthritis and has a specificity of 100% for the disease. *Arthritis Res Ther* 2005;**8**:R19(1–6).
- Sugiyama D, Nishimura K, Tamaki K, *et al*. Impact of smoking as a risk factor for developing rheumatoid arthritis: a meta-analysis of observational studies. *Ann Rheum Dis* 2010;**69**:70–81.
- 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.
- 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.
- Rantapää-Dahlqvist S, de Jong BA, Berglin E, *et al*. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. *Arthritis Rheum* 2003;**48**:2741–9.
- Kokkonen H, Mullazehi M, Berglin E, *et al*. Antibodies of IgG, IgA and IgM isotypes against cyclic citrullinated peptide precede the development of rheumatoid arthritis. *Arthritis Res Ther* 2011;**13**:R13.
- Kuhn KA, Kulik L, Tomooka B, *et al*. Antibodies against citrullinated proteins enhance tissue injury in experimental autoimmune arthritis. *J Clin Invest* 2006;**116**:961–73.
- Ioan-Facsinay A, Willemze A, Robinson DB, *et al*. Marked differences in fine specificity and isotype usage of the anti-citrullinated protein antibody in health and disease. *Arthritis Rheum* 2008;**58**:3000–8.
- Kolfenbach JR, Deane KD, Derber LA, *et al*. A prospective approach to investigating the natural history of preclinical rheumatoid arthritis (RA) using first-degree relatives of probands with RA. *Arthritis Rheum* 2009;**61**:1735–42.
- 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.
- Breiman L. Random Forests. *Machine Learning* 2001;**45**:5–32.
- van der Helm-van Mil AH, Verpoort KN, Breedveld FC, *et al*. 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.
- van der Woude D, Lie BA, Lundström E, *et al*. Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: a meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in four European populations. *Arthritis Rheum* 2010;**62**:1236–45.