

# A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis

Leonid Padyukov,<sup>1</sup> Mark Seielstad,<sup>2,3</sup> Rick T H Ong,<sup>2</sup> Bo Ding,<sup>4</sup> Johan Rönnelid,<sup>5</sup> Maria Seddighzadeh,<sup>1</sup> Lars Alfredsson,<sup>4</sup> Lars Klareskog,<sup>1</sup> the Epidemiological Investigation of Rheumatoid Arthritis (EIRA) Study Group

► Additional data are published online only. To view these files please visit the journal online (<http://ard.bmj.com>).

<sup>1</sup>Rheumatology Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden

<sup>2</sup>The Genome Institute of Singapore, Singapore

<sup>3</sup>The Institute for Human Genetics and Department of Laboratory Medicine, University of California, San Francisco School of Medicine, San Francisco, California, USA

<sup>4</sup>The Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden

<sup>5</sup>The Unit for Clinical Immunology, Uppsala University, Uppsala, Sweden

## Correspondence to

Dr Seielstad, at the Genome Institute of Singapore, 60 Biopolis St, Singapore 138672; [seielstadm@gis.a-star.edu.sg](mailto:seielstadm@gis.a-star.edu.sg)

LP and MS contributed equally to this work.

Accepted 26 July 2010  
Published Online First  
14 December 2010

## ABSTRACT

**Background** Rheumatoid arthritis (RA) can be divided into two major subsets based on the presence or absence of antibodies to citrullinated peptide antigens (ACPA). Until now, data from genome-wide association studies (GWAS) have only been published from ACPA-positive subsets of RA or from studies that have not separated the two subsets. The aim of the current study is to provide and compare GWAS data for both subsets. **Methods and results** GWAS using the Illumina 300K chip was performed for 774 ACPA-negative patients with RA, 1147 ACPA-positive patients with RA and 1079 controls from the Swedish population-based case-control study EIRA. Imputation was performed which allowed comparisons using 1 723 056 single nucleotide polymorphisms (SNPs). No SNP achieved genome-wide significance ( $2.9 \times 10^{-8}$ ) in the comparison between ACPA-negative RA and controls. A case-case association study was then performed between ACPA-negative and ACPA-positive RA groups. The major difference in this analysis was in the HLA region where 768 HLA SNPs passed the threshold for genome-wide significance whereas additional contrasting SNPs did not reach genome-wide significance. However, one SNP close to the *RPS12P4* locus in chromosome 2 reached a p value of  $2 \times 10^6$  and this locus can thus be considered as a tentative candidate locus for ACPA-negative RA.

**Conclusions** ACPA-positive and ACPA-negative RA display significant risk allele frequency differences which are mainly confined to the HLA region. The data provide further support for distinct genetic aetiologies of RA subsets and emphasise the need to consider them separately in genetic as well as functional studies of this disease.

## INTRODUCTION

Rheumatoid arthritis (RA) is a common inflammatory joint disease caused by a complex interplay of genetic variants and environmental exposures.<sup>1-3</sup> Disease outcomes in RA are highly variable, and the presence or absence of antibodies to citrullinated peptide antigens (ACPA) has proved to be one of the best clinical predictors of the severity of disease course.<sup>4 5</sup> In addition, ACPA-positive patients with unspecified arthritis respond differently from ACPA-negative patients with RA to early methotrexate therapy.<sup>6</sup>

In recent years a number of candidate genes have been shown to associate differently with ACPA-positive and ACPA-negative RA. Several genetic

variants within the HLA region, specifically the shared epitope-containing HLA-DRB1 alleles,<sup>7-9</sup> PTPN22 alleles,<sup>10 11</sup> as well as a variant in the C5-TRAF1 region,<sup>12 13</sup> TNFAIP3,<sup>14 15</sup> CD40, CCL21 and many other loci<sup>16 17</sup> have been shown to associate with ACPA-positive RA but have not been tested for ACPA-negative RA. By contrast, variations in IRF5<sup>18</sup> and C-type lectin genes<sup>19</sup> appear to be associated with ACPA-negative RA. Association of ACPA-negative disease with HLA-DRB1\*03 haplotype was previously suggested<sup>20 21</sup> but has not been replicated in a larger study.<sup>22</sup> In at least one study<sup>23</sup> an indication for an association of PTPN22 marker with ACPA-negative RA was presented based on 65 cases of RA. On the other hand, STAT4 variant has been shown in a meta-analysis to be a risk factor for both subgroups of RA.<sup>24</sup> Smoking is the only environmental risk factor unambiguously associated with the risk of RA, but it too appears to affect risk only for ACPA-positive patients.<sup>8</sup>

These data on different risk factors for ACPA-positive and ACPA-negative RA have been used to propose a new aetiological model for ACPA-positive RA, whereas no such model yet exists for ACPA-negative disease.<sup>18 25 26</sup> A major implication of these observations is that genetic and immunological studies of RA should consider this heterogeneity of RA.

So far, the most powerful technique to analyse the effects of genetic variation on disease susceptibility—that is, the genome-wide association study (GWAS)—has not addressed this heterogeneity and genome-wide data published on RA to date have either considered ACPA-positive disease alone<sup>13 17 27</sup> or grouped both subtypes together.<sup>14 28</sup>

In order to provide a more complete picture of genetic risk factors for RA, we have performed genome-wide association analyses in both RA subsets in two different collections of RA cases and controls defined by ACPA status (Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) and North American RA Consortium (NARAC)), and in data from the Wellcome Trust Case-Control Consortium (WTCCC) which contains patients from both subsets but where subdivision according to ACPA status has not yet been performed.

## METHODS

### Subjects

EIRA is a population-based case-control study enrolling incident (predominantly <1 year after



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

clinical onset) cases of RA. The study base comprises residents aged 18–70 years in a geographically-defined area in the central and southern parts of Sweden. Details of the study design have been reported elsewhere.<sup>8 29 30</sup> For each case a control was randomly selected from the study base by matching age, sex and residential area. For the present study we selected 3176 individuals (829 ACPA-negative cases of RA, 1218 ACPA-positive cases of RA and 1129 controls) for genome-wide genotyping. Informed consent was obtained from all participants and the ethical review board at the Karolinska Institutet approved the study. A portion of the data was included in a previously published GWAS of ACPA-positive RA<sup>13</sup> (see table 1 in online supplement), but here we enlarge the GWAS dataset for ACPA-positive patients with RA and controls and include genome-wide data on ACPA-negative patients for the first time.

NARAC provided genotypes for 889 ACPA-positive patients and 1232 controls. The patients of self-reported white ancestry were recruited as prevalent (69.6%) or incident RA cases from several sites throughout North America. Nearly half of the cases (51.1%) reported a positive family history.<sup>13</sup> Control subjects were selected on the basis of similar self-reported ancestry from 20 000 persons who were part of the New York Cancer Project. Written informed consent was obtained from all subjects who provided blood samples in accordance with protocols approved by the local institutional review boards.

The British RA population comprised 1860 patients and 3000 controls. Recruitment procedures have been described previously, together with frequencies of genetic variations.<sup>28</sup>

Demographic characteristics of patients and controls and the study logistics are shown in table 1 in the online supplement and in figure 1 where corresponding numbers after quality control procedures (see below) are shown. Genotyping, serological analysis and statistical evaluation are shown in the online supplement.

**RESULTS**

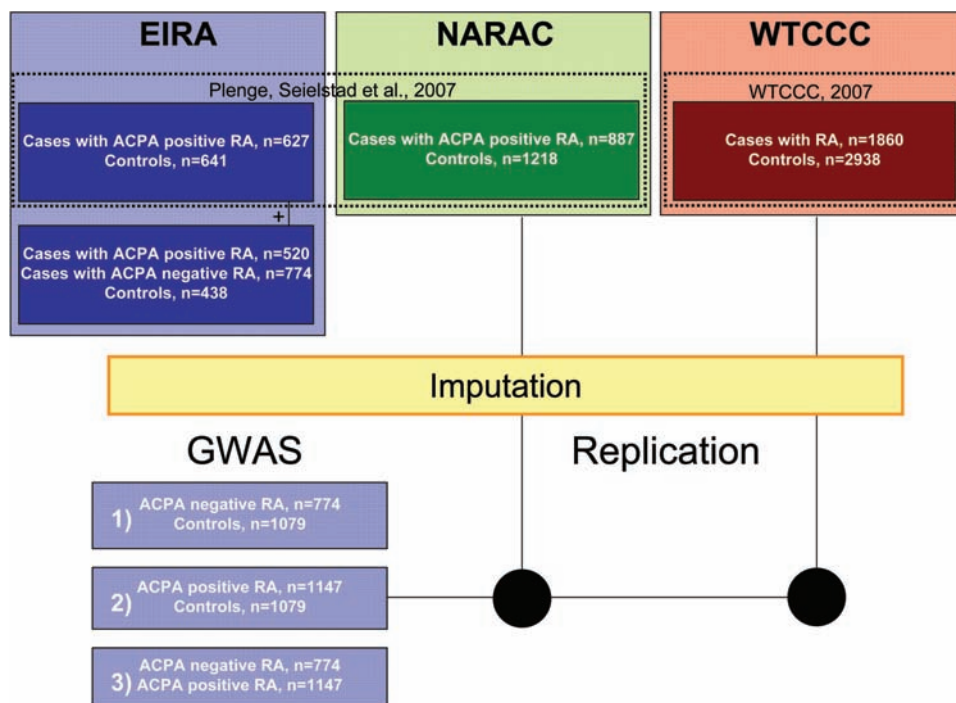
**GWAS of ACPA-negative RA**

We conducted a GWAS for ACPA-negative patients with RA with 774 cases and 1079 controls selected from the EIRA study. Both genotyped and imputed single nucleotide polymorphisms (SNPs) were included in the analysis. No single SNP reached genome-wide significance (figure 2). A Q-Q plot for observed versus expected p values is shown in figure 3A and shows no significant deviation from the expected distribution. As shown in table 1, five SNPs from three genetic loci had p values <10<sup>-5</sup>. One out of five was not associated with ACPA-positive RA while the other four (from two independent loci at chromosome 7) had only nominal association. These results indicate little or no overlap between the two RA subgroups for the five tentative SNPs that may associate with ACPA-negative RA in this GWAS.

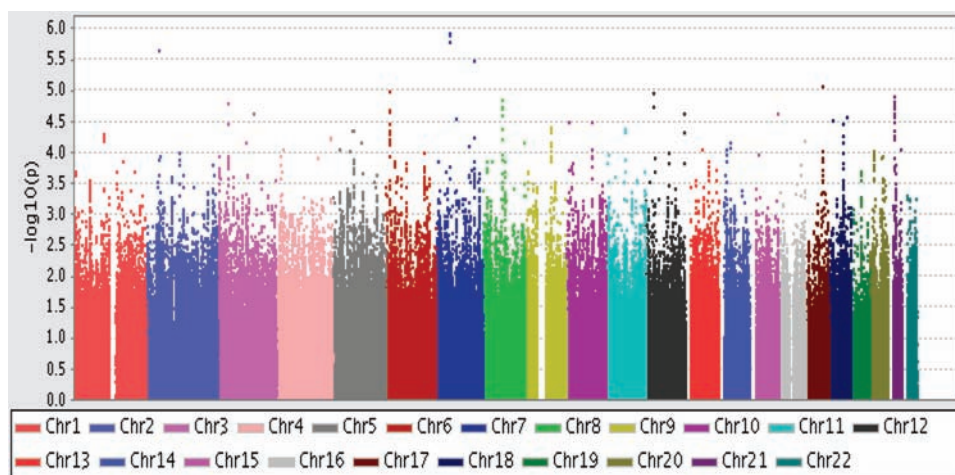
Owing to our inability to identify any additional appropriately-sized case-control studies of ACPA-negative RA, we have so far been unable to replicate our findings for ACPA-negative RA.

**GWAS of ACPA-positive RA**

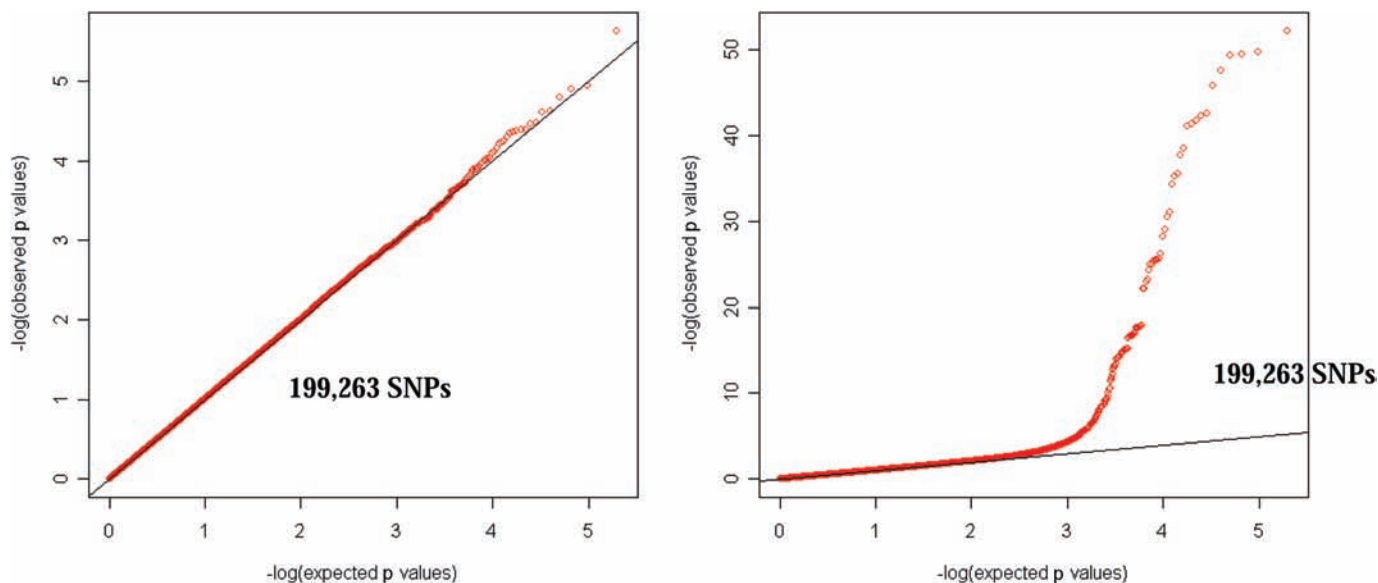
We have previously reported GWAS data on ACPA-positive RA based on a fraction of the EIRA study in combination with NARAC.<sup>13</sup> In addition to the previously used 627 ACPA-positive RA cases and 641 controls from Sweden, we now selected 520 new ACPA-positive RA cases and 438 controls from the EIRA study for extension of the GWAS for ACPA-positive RA. Because all cases and controls were taken from the same study population, we combined all EIRA samples into a single analysis which included both genotyped and imputed SNPs. The diagram of genome-wide association for this analysis is shown in figures 4 and 5 and Q-Q plots are shown in figure 3B. Out of 1 723 056 analysed SNPs, we found



**Figure 1** Work flow for genome-wide analysis (GWAS) of two subgroups of rheumatoid arthritis (RA). Three different Caucasian study populations represented in different colours (blue, green and red: Swedish, US, UK). Selection of data from previously published studies indicated by dashed rectangles with subsequent reference. Difference in sums is due to QC procedures at each stage of analysis. ACPA, antibodies to citrullinated peptide antigens.



**Figure 2** Probability plot for association with ACPA-negative rheumatoid arthritis (774 cases) versus 1079 controls,  $\lambda_{GC} = 1.0132$  based on 1 723 056 single nucleotide polymorphisms. ACPA, antibodies to citrullinated peptide antigens.



**Figure 3** Quintile-quintile probability plots for observed versus expected distribution of p values for  $\chi^2$  statistics for (A) ACPA-negative RA versus controls and (B) ACPA-positive RA versus controls. ACPA, antibodies to citrullinated peptide antigens; RA, rheumatoid arthritis; SNP, single nucleotide polymorphism.

**Table 1** Characteristics of the top associated SNPs from GWAS of ACPA-negative RA versus controls

Reference sequence	Chromosome	Closest gene	Minor allele*	ACPA-negative RA versus controls				ACPA-positive RA versus controls		
				Allelic counts for cases	Allelic counts for controls	OR (95% CI)	p Value†	Allelic counts for cases	Allelic counts for controls	p Value†
rs4305317	2	RPS12P4	T	384/1132	691/1403	0.69 (0.59 to 0.80)	2.32E-06	777/1499	691/1403	0.453
rs6972219	7	IGFBP1	T	45/1299	20/1960	3.39 (2.00 to 5.78)	1.77E-06	42/2206	20/1960	0.0205
rs2173035	7	IGFBP1	A	47/1303	21/1957	3.36 (2.00 to 5.65)	1.27E-06	50/2198	21/1957	0.0033
rs6956740	7	IGFBP1	T	48/1294	21/1973	3.49 (2.08 to 5.85)	5.12E-07	49/2217	21/1973	0.0045
rs3807306	7	IRF5	T	785/555	974/992	1.44 (1.25 to 1.66)	3.15E-07	1199/1051	974/992	0.0152

All frequencies were evaluated from the real genotyping data.

\*Same direction for association in RA subgroups.

†Cochran–Armitage trend test.

ACPA, antibodies to citrullinated peptide antigens; GWAS, genome-wide association studies; RA, rheumatoid arthritis; SNP, single nucleotide polymorphism.

719 SNPs which passed a genome-wide significance threshold (see table 2 in online supplement). Of note, 718 of these SNPs were located within the HLA locus at chromosome 6 with physical positions between 31 279 236 and 33 164 413 (1 885 177 bp). A single non-HLA SNP rs2476601 was from *PTPN22* gene at chromosome 1 with OR 1.66 (95% CI 1.40 to

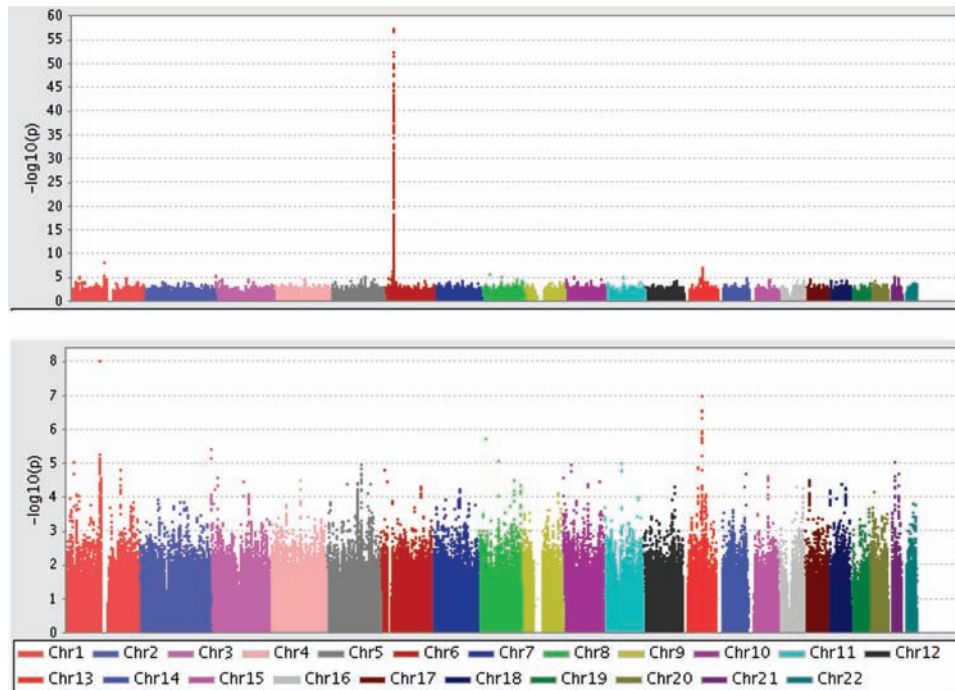
1.96,  $p=9.77E-09$ ). After relaxing the threshold for significance up to  $10^{-6}$ , an additional 196 SNPs were found to be significant from the HLA locus and seven non-HLA SNPs from the cluster of a ‘gene desert’ at chromosome 13 (see table 2 in online supplement). These seven SNPs are all in a recombination block according to HapMap data.<sup>31</sup>

We subsequently extended the observations on patients with ACPA-positive RA in EIRA with a replication in NARAC using patients with ACPA-positive RA as cases and healthy individuals from a New York cancer surveillance study as controls. Many of the SNPs in the HLA region and one PTPN22 SNP (rs2476601) from EIRA were well replicated in NARAC (see table 3 in online supplement).

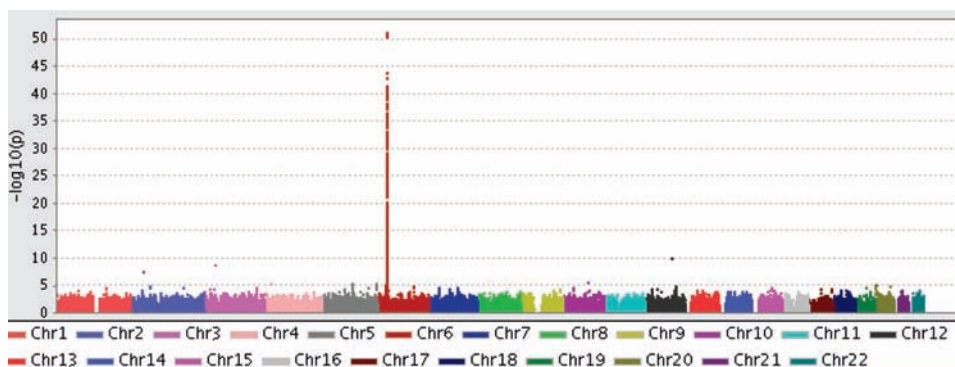
We also made an extra validation against the WTCCC study where, however, we were not able to separate patients based on ACPA status. It is known, however, that the large majority of the WTCCC RA cohort is rheumatoid factor (RF)-positive and, owing to a high correlation between RF and ACPA status, it most likely dominated by ACPA-positive RA cases.<sup>28</sup> Many of the SNPs in the HLA region and one PTPN22 SNP (rs2476601) were well replicated also in WTCCC (table 3 in online supplement) with an overall OR for rs2476601 in the three studies of 1.74 (95% CI 1.59 to 1.90,  $p=2.73045E-36$ , Mantel-Haenszel  $\chi^2$  test for 17 520 chromosomes).

### Contrasts between ACPA-positive and ACPA-negative RA

To formally test the hypothesis of a contrast between the two disease subgroups, we used the EIRA study to perform a direct comparison between ACPA-positive and ACPA-negative RA using the full GWAS data sets for these two populations of RA patients. The threshold for genome-wide significance was estimated as  $2.9 \times 10^{-8}$  (after Bonferroni correction for 1 723 056 tests). After corrections for multiple testing we found significant differences only in the HLA region of chromosome 6p with 814 SNPs spanning between 31 278 893 and 33 164 413 (see table 4 in online supplement). When we increased the threshold to  $10^{-5}$  we identified an additional 352 SNPs in the HLA region targeted to physical positions in the extended HLA locus between 30 155 944 and 33 886 942 (3 730 998 bp) and three additional non-HLA SNPs (table 2): rs4305317 from chromosome 2 close to the *LDHAL3* (lactate dehydrogenase A-like 3) gene, rs6448119 from chromosome 4 between the *KCNIP4* (Kv channel interacting protein 4) and the *GPR125* (G protein-coupled receptor



**Figure 4** Probability plot for association with ACPA-positive rheumatoid arthritis (1147 cases) versus 1079 controls,  $\lambda_{GC} = 1.0263$  based on 1 723 056 single nucleotide polymorphisms (SNPs). Upper figure represents data for all SNPs including MHC locus and lower figure is without data for SNPs from MHC locus. ACPA, antibodies to citrullinated peptide antigens.



**Figure 5** Probability plot for association for ACPA-positive RA (1147 cases) versus ACPA-negative RA (774 cases).  $\lambda_{GC} = 1.0029$  based on 1 723 056 single nucleotide polymorphisms. ACPA, antibodies to citrullinated peptide antigens; RA, rheumatoid arthritis.

125) genes, and rs2961663 from chromosome 5 in the *GMCL1L* (germ cell-less homolog 1) gene.

The described differences in the HLA region between ACPA-positive and ACPA-negative disease are well in line with a previously published dataset on ACPA-positive and ACPA-negative patients with RA compared with controls which was focused only on the HLA using another set of SNPs together with classical PCR-based HLA-typing.<sup>7 32</sup>

Since the difference in allelic frequency in this analysis could be linked to the effect from any of the RA subgroups although it is dominated by susceptibility risk alleles for ACPA-positive RA, it is interesting to note that the two non-HLA SNPs rs4305317 and rs6448119 seem to be in association primarily with ACPA-negative RA.

**DISCUSSION**

We report the first GWAS in one of the major subsets of RA defined by absence of ACPA reactivity, and the first comparison of GWAS results between ACPA-positive and ACPA-negative RA. Overall, we found significant differences in genetic associations between the two subsets, both concerning genes already described as being associated with ACPA-positive RA in previous genome-wide studies and concerning tentative associations that are preferentially seen for the ACPA-negative subset of RA. Notably, only very few genetic variations were associated with both RA subsets and those had a very minor influence on the genetic risk of RA.

The contrasting genetic aetiologies of ACPA-positive and ACPA-negative RA are most starkly evident at the HLA locus, and these data are in line with previous studies that were specifically focused on the HLA region and used other SNPs and smaller numbers than in the present study.<sup>7 32</sup> These data therefore show that the functional conclusions concerning molecular pathogenesis of arthritis that have since long been deduced from the association with HLA genes, and particularly with variations in HLA class II genes, are valid only for the ACPA-positive subset of the disease; the implications of this discrepancy have also been discussed previously on the basis of data from classical PCR-based HLA genotyping.<sup>7 9 25 32</sup>

Concerning the non-HLA genes, several studies have now been published describing variations associated with either ACPA-positive RA<sup>13 14 17 27</sup> or with RA where no discrimination between the two subsets was made.<sup>28</sup> When analysing the published genetic variations that showed genome-wide or suggestive associations with ACPA-positive disease or with the entire RA population and also with the ACPA-negative RA subset in the EIRA study, the genetic variants in association with ACPA-positive RA were most often not significant for ACPA-negative RA and vice versa.

We performed an analysis of 14 well-established variations for association with ACPA-positive RA in combined material from Sweden, the USA and the UK and for ACPA-negative RA in the Swedish EIRA cohort (table 3). As can be seen from this analysis, most previously detected genetic variations (9/14) were associated only with ACPA-positive RA, three appeared to be associated with both subgroups, one appeared to be specific for ACPA-negative disease and one appeared to provide opposite results in the two subgroups. Since most efforts in the genetics of RA have been devoted to ACPA-positive RA, this difference may be due to selection rather than to a true absence of association with ACPA-negative RA. A recent twin study of ACPA-negative RA indicated that this phenotype is also genetically predetermined.<sup>33</sup> Thus, our present negative

**Table 2** Characteristics of the top non-HLA SNPs from GWAS of ACPA-positive versus ACPA-negative RA for EIRA in comparison with NARAC and WTCCC studies

Reference sequence	Chromosome	EIRA ACPA-positive versus ACPA-negative		EIRA ACPA positive RA		EIRA ACPA negative RA		NARAC		WTCCC		Genotyping status*							
		ACPA positive	ACPA negative	ACPA positive	Allelic frequency in cases	Allelic frequency in controls	Allelic frequency in cases	Allelic frequency in controls	Allelic frequency in cases	Allelic frequency in controls	Allelic frequency in cases	Allelic frequency in controls	p Value	EIRA	NARAC	WTCCC			
rs4305317	2	777/1499	384/1132	3.86E-08	0.34	0.33	0.453	0.25	0.33	2.32E-06	0.46	0.46	0.7962	0.48	0.45	0.008084	gen	imp	imp
rs6448119	4	283/2007	116/1420	7.27E-06	0.12	0.12	0.5	0.08	0.12	8.98E-05	0.03	0.03	0.8406	0.03	0.02	0.5038	gen/imp	imp	imp
rs2961663	5	237/2051	238/1308	4.37E-06	0.10	0.13	0.004487	0.15	0.13	0.05238	0.10	0.10	0.9345	0.09	0.10	0.127	gen	gen	imp

\*Genotyping status referred as directly genotyped (gen<sup>†</sup>) or imputed (imp).

<sup>†</sup>Cochran–Armitage trend test.

<sup>‡</sup>Genotyping frequency for ACPA-negative RA group presented for opposite allele due to different direction of association.

ACPA, antibodies to citrullinated peptide antigens; EIRA, Swedish Epidemiological Investigation of Rheumatoid Arthritis; GWAS, genome-wide association studies; NARAC, North American RA Consortium; RA, rheumatoid arthritis; SNP, single nucleotide polymorphism; WTCCC, Wellcome Trust Case-Control Consortium.

**Table 3** Frequencies of previously detected genetic variants for RA: comparison of meta-analysis of ACPA-positive RA patients from three populations with ACPA-negative RA patients

SNP	Chromosome	Gene/loci	Minor allele*	EIRA ACPA-positive		EIRA ACPA-negative		NARAC		WTCCC		ACPA-positive RA	ACPA-negative RA
				MAF cases	MAF controls	MAF cases	MAF controls	MAF cases	MAF controls	MAF cases	MAF controls	OR (95%CI)*	OR (95%CI)†
rs3890745	1p36	<i>MMEL1-TNFRSF14</i>	C	0.28	0.31	0.29	0.31	0.29	0.32	0.29	0.34	<b>0.85 (0.80 to 0.91)</b>	0.94 (0.82 to 1.09)
rs2476601	1p13	<i>PTPN22</i>	A	0.18	0.12	0.14	0.12	0.16	0.08	0.15	0.09	<b>1.74 (1.59 to 1.90)</b>	1.17 (0.96 to 1.42)
rs7574865	2q32	<i>STAT4</i>	T	0.24	0.22	0.26	0.22	0.25	0.23	0.23	0.21	<b>1.12 (1.05 to 1.21)</b>	<b>1.23 (1.05 to 1.43)</b>
rs3087243	2q33	<i>CTLA4</i>	A	0.34	0.39	0.38	0.39	0.38	0.42	0.43	0.45	<b>0.88 (0.83 to 0.93)</b>	0.97 (0.84 to 1.11)
rs2064476	6p21	<i>HLA-DPB1</i>	G	0.20	0.26	0.27	0.26	0.22	0.30	0.25	0.31	<b>0.71 (0.66 to 0.76)</b>	1.02 (0.88 to 1.18)
rs10499194	6q23	<i>OLIG3, TNFAIP3</i>	T	0.19	0.21	0.20	0.21	0.25	0.30	0.25	0.27	<b>0.86 (0.80 to 0.92)</b>	<i>1.28 (1.10 to 1.50)</i>
rs6920220	6q23	<i>OLIG3, TNFAIP3</i>	A	0.25	0.23	0.24	0.23	0.24	0.19	0.26	0.22	<b>1.22 (1.14 to 1.31)</b>	1.08 (0.93 to 1.26)
rs3807306	7q32	<i>IRF5</i>	G	0.45	0.49	0.41	0.49	0.48	0.50	0.46	0.50	<b>0.87 (0.81 to 0.92)</b>	<b>0.72 (0.63 to 0.82)</b>
rs2812378	9p13	<i>CCL21</i>	G	0.35	0.32	0.36	0.32	0.35	0.33	0.37	0.34	<b>1.14 (1.07 to 1.21)</b>	<b>1.21 (1.05 to 1.39)</b>
rs3761847	9q33	<i>TRAF1-C5</i>	G	0.51	0.46	0.48	0.46	0.47	0.38	0.43	0.43	<b>1.32 (1.21 to 1.44)‡</b>	1.08 (0.95 to 1.23)
rs4750316	10p15	<i>PRKCC</i>	C	0.17	0.21	0.19	0.21	0.18	0.19	0.16	0.20	<b>0.83 (0.77 to 0.90)</b>	0.87 (0.74 to 1.03)
rs1678542	12q13	<i>KIF5A-PIP4K2C</i>	G	0.40	0.42	0.42	0.42	0.36	0.37	0.33	0.37	<b>0.88 (0.82 to 0.93)</b>	0.99 (0.87 to 1.14)
rs1328674	13q14-q21	<i>HTR2A</i>	T	0.04	0.02	0.04	0.02	0.04	0.04	0.04	0.04	1.16 (0.99 to 1.35)	<b>1.83 (1.25 to 2.68)</b>
rs4810485	20q13	<i>CD40</i>	T	0.20	0.24	0.24	0.24	0.21	0.27	0.23	0.24	<b>0.77 (0.69 to 0.85)‡</b>	0.98 (0.84 to 1.14)

Statistically significant odds ratios shown in bold.

\*Mantel-Haenszel OR from meta-analysis of EIRA ACPA-positive RA/NARAC/WTCCC.

†Mantel-Haenszel OR for EIRA ACPA-negative RA.

‡WTCCC excluded due to heterogeneity between ACPA positive cohorts.

ACPA, antibodies to citrullinated peptide antigens; EIRA, Swedish Epidemiological Investigation of Rheumatoid Arthritis; MAF, minor allele frequency; NARAC, North American RA Consortium; RA, rheumatoid arthritis; SNP, single nucleotide polymorphism; WTCCC, Wellcome Trust Case-Control Consortium.

results have to be interpreted with caution given the limited power of our investigations. Nevertheless, we can be confident that the major discrepancies described by GWAS—that is, the differential associations in the HLA region—are indeed true differences between the ACPA-positive and ACPA-negative RA subsets.

Concerning the non-HLA genes associated with the different subsets of RA, we performed additional genotyping to confirm the accuracy of imputation of SNPs using real genotyping data, when possible, to evaluate differences between the two RA subsets in order to decrease the risk of false positive findings due to use of imputations. Using this methodology we were not able to identify any strong genetic risk factors for ACPA-negative RA, while a number of SNPs from the HLA region as well as non-HLA SNPs were in association with ACPA-positive disease, confirming previously published data for ACPA-positive RA.<sup>13 28</sup> Thus, although we cannot totally exclude overlapping genetic risk factors for the two RA subgroups, it is unlikely that this overlap is very big. We can also confirm the major differences between ACPA-positive and ACPA-negative RA concerning linkage to the HLA region.

Deciphering the pathogenesis of ACPA-negative RA remains a major challenge for genetic studies. In this study we found suggestive evidence—but not genome-wide significance-based evidence—for two new candidate loci in ACPA-negative disease (ie, RPS12P4 and IGFBP1). We also found suggestive evidence for the previously described IRF5 locus as susceptibility genes for ACPA-negative RA. Thus, most of the difference in genetic factors between ACPA-negative and ACPA-positive RA is seen in the HLA region, close to the HLA-DRB1 locus. The non-HLA variant rs4305317, close to RPS12P4 at chromosome 2, is the best candidate for association with ACPA-negative RA but not with ACPA-positive RA. It should be emphasised, however, that this association as well as other data related to associations with ACPA-negative RA need independent replication owing to the limited size of the present study.

GWAS provide new potentials to determine genetic variations and molecular pathways that are shared between

different inflammatory diseases. The associations of several diseases and disease subsets with PTPN22, CD40, TRAF1-C5 and STAT4 are typical examples of this sharing.<sup>17 27 34 35</sup> The present study illustrates the other complementary perspective—namely, the potential also to use GWAS to subdivide criterion-based diseases such as RA into new entities. As exemplified here, one subset of a disease may then share certain risk genes and possibly pathogenic pathways with some other autoimmune entities, as is the case for PTPN22 and ACPA-positive RA and type I diabetes,<sup>36</sup> whereas the other subset of the same criterion-based disease may share genetic linkages and pathogenic pathways with still other autoimmune entities. We can expect refined classifications of criterion-based diseases such as RA to lead to the use of genetics to provide an indication of what molecular pathways are involved in the pathogenesis of different subsets of today's criterion-based diseases. This knowledge should, in turn, be indispensable when trying to find treatments to target these pathways in patients in these different and distinct subsets of chronic autoimmune diseases.

**Acknowledgements** The authors acknowledge the help of Peter K Gregersen and Jane Worthington in access to NARAC and WTCCC data. They also thank Robert M Plenge for careful reading of the manuscript and critical discussions; Kian Mun Chan, Boon Yeong Goh, Wee Yang Meah, Jameelah B S Mohamed, Jason Ong, Eileen Ping and Sigeeta Rajaram for their invaluable laboratory assistance; the participating patients with RA and controls and all rheumatologists for recruiting patients in the EIRA study; and Marie-Louise Serra, Camilla Bengtsson, Eva Jemseby and Lena Nise for their invaluable contributions to the collection of data and maintenance of the database.

**Competing interests** None.

**Funding** Supported by grants from the Swedish Research Council, the Swedish Council for Working Life and Social Research, King Gustaf V's 80-year Foundation, the Swedish Rheumatism Foundation, the insurer, AFA and the EU-supported projects AutoCure and Masterswitch. The Agency for Science Technology and Research (ASTAR), Singapore supported the genotyping and data analysis. The funding agencies had no influence on study design, evaluation of results and decision to publish the data.

**Ethics approval** This study was conducted with the approval of the Regional Ethical Review Board in Stockholm.

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

## REFERENCES

1. **Gregersen PK**, Olsson LM. Recent advances in the genetics of autoimmune disease. *Annu Rev Immunol* 2009;**27**:363–91.
2. **Klareskog L**, Padyukov L, Rönnelid J, *et al*. Genes, environment and immunity in the development of rheumatoid arthritis. *Curr Opin Immunol* 2006;**18**:650–5.
3. **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.
4. **Kastbom A**, Strandberg G, Lindroos A, *et al*. Anti-CCP antibody test predicts the disease course during 3 years in early rheumatoid arthritis (the Swedish TIRA project). *Ann Rheum Dis* 2004;**63**:1085–9.
5. **van der Helm-van Mil AH**, Verpoort KN, Breedveld FC, *et al*. Antibodies to citrullinated proteins and differences in clinical progression of rheumatoid arthritis. *Arthritis Res Ther* 2005;**7**:R949–58.
6. **van Dongen H**, van Aken J, Lard LR, *et al*. Efficacy of methotrexate treatment in patients with probable rheumatoid arthritis: a double-blind, randomized, placebo-controlled trial. *Arthritis Rheum* 2007;**56**:1424–32.
7. **Ding B**, Padyukov L, Lundström E, *et al*. Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region. *Arthritis Rheum* 2009;**60**:30–8.
8. **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.
9. **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.
10. **Begovic 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.
11. **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.
12. **Kurreeman FA**, Padyukov L, Marques RB, *et al*. A candidate gene approach identifies the TRAF1/C5 region as a risk factor for rheumatoid arthritis. *PLoS Med* 2007;**4**:e278.
13. **Plenge RM**, Seielstad M, Padyukov L, *et al*. TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. *N Engl J Med* 2007;**357**:1199–209.
14. **Plenge RM**, Cotsapas C, Davies L, *et al*. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007;**39**:1477–82.
15. **Thomson W**, Barton A, Ke X, *et al*. Wellcome Trust Case Control Consortium; YEAR Consortium. Rheumatoid arthritis association at 6q23. *Nat Genet* 2007;**39**:1431–3.
16. **Barton A**, Thomson W, Ke X, *et al*. Wellcome Trust Case Control Consortium; YEAR Consortium; BIRAC Consortium. Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat Genet* 2008;**40**:1156–9.
17. **Raychaudhuri S**, Remmers EF, Lee AT, *et al*. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat Genet* 2008;**40**:1216–23.
18. **Sigurdsson S**, Padyukov L, Kurreeman FA, *et al*. Association of a haplotype in the promoter region of the interferon regulatory factor 5 gene with rheumatoid arthritis. *Arthritis Rheum* 2007;**56**:2202–10.
19. **Lorentzen JC**, Flornæs L, Eklöv C, *et al*. Association of arthritis with a gene complex encoding C-type lectin-like receptors. *Arthritis Rheum* 2007;**56**:2620–32.
20. **Irigoyen P**, Lee AT, Wener MH, *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.
21. **Verpoort KN**, van Gaalen FA, van der Helm-van Mil AH, *et al*. Association of HLA-DR3 with anti-cyclic citrullinated peptide antibody-negative rheumatoid arthritis. *Arthritis Rheum* 2005;**52**:3058–62.
22. **Lundström E**, Källberg H, Smolnikova M, *et al*. Opposing effects of HLA-DRB1\*13 alleles on the risk of developing anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum* 2009;**60**:924–30.
23. **Pierer M**, Kaltenhäuser S, Arnold S, *et al*. Association of PTPN22 1858 single-nucleotide polymorphism with rheumatoid arthritis in a German cohort: higher frequency of the risk allele in male compared to female patients. *Arthritis Res Ther* 2006;**8**:R75.
24. **Lee YH**, Woo JH, Choi SJ, *et al*. Association between the rs7574865 polymorphism of STAT4 and rheumatoid arthritis: a meta-analysis. *Rheumatol Int* 2010;**30**:661–6.
25. **Klareskog L**, Rönnelid J, Lundberg K, *et al*. Immunity to citrullinated proteins in rheumatoid arthritis. *Annu Rev Immunol* 2008;**26**:651–75.
26. **Klareskog L**, Catrina AI, Paget S. Rheumatoid arthritis. *Lancet* 2009;**373**:659–72.
27. **Remmers EF**, Plenge RM, Lee AT, *et al*. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N Engl J Med* 2007;**357**:977–86.
28. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661–78.
29. **Padyukov L**, Silva C, Stolt P, *et al*. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis Rheum* 2004;**50**:3085–92.
30. **Stolt P**, Bengtsson C, Nordmark B, *et al*. EIRA Study Group. 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.
31. The International HapMap Consortium. The International HapMap Project. *Nature* 2003;**426**:789–96.
32. **Lee HS**, Lee AT, Criswell LA, *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.
33. **van der Woude D**, Houwing-Duistermaat JJ, Toes RE, *et al*. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum* 2009;**60**:916–23.
34. **Carlton VE**, Hu X, Chokkalingam AP, *et al*. PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid arthritis. *Am J Hum Genet* 2005;**77**:567–81.
35. **Kurreeman FS**, Alizadeh BZ, Goulielmos GN, *et al*. The TRAF1-C5 region on chromosome 9q33 is associated with multiple autoimmune diseases. *Ann Rheum Dis* 2010;**69**:696–9.
36. **Criswell LA**, Pfeiffer KA, Lum RF, *et al*. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am J Hum Genet* 2005;**76**:561–71.