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

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

ABSTRACT
Background Rheumatoid arthritis (RA) can be divided into two major subsets based on the presence or absence of antibodies to citrullinated peptide antigens (ACPAs). 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. However, one SNP close to the locus in chromosome 2 reached a p value of 2 × 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.1–3 Disease outcomes in RA are highly variable, and the presence or absence of antibodies to citrullinated peptide antigens (ACPAs) has proved to be one of the best clinical predictors of the severity of disease course.4 5 In addition, ACPA-positive patients with unspecified arthritis respond differently from ACPA-negative patients with RA to early methotrexate therapy.6

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,7–9 PTPN22 alleles,10 11 as well as a variant in the C5-TRAF1 region,12 13 TNFAIP3,14 15 CD40, CCL21 and many other loci16 17 have been shown to associate with ACPA-positive RA but have not been tested for ACPA-negative RA. By contrast, variations in IRF518 and C-type lectin genes19 appear to be associated with ACPA-negative RA. Association of ACPA-negative disease with HLA-DRB1*04 haplotype was previously suggested20 but has not been replicated in a larger study.21 In at least one study22 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.23 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.8

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.18 24 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 alone11 15 17 19 or grouped both subtypes together.14 24

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...
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^{-5}. 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.13 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...
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.\(^{31}\)

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.\(^{31}\)
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. 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): rs4305917 from chromosome 2 close to the LDHAL3 gene, rs6448119 from chromosome 4 between the KCNIP4 and GPR125 genes, and the PTPN22 SNP (rs2476601) was 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).

**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.
between the two subsets was made. When analysing the pub-
lic PCR-based HLA genotyping. A recent twin study of ACPA-negative RA indicated that this phenotype is
also genetically predetermined. Thus, our present negative
study of ACPA-negative RA indicated that this phenotype
is dominated by susceptibility risk alleles for ACPA-positive RA, and two non-HLA SNPs rs4305317 and rs6448119 seem to be in association primarily with ACPA-negative RA.

The contrasting genetic aetiologies of ACPA-positive and
ACPA-negative RA are most likely evident at the HLA locus,
and the differences between the two subsets, both concerning genes already
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 associa-
tions between the two subsets. Both concerning genetic associa-
tions described as being associated with ACPA-negative RA in previ-
sous genome-wide studies and concerning tentative associations
described as being associated with ACPA-positive RA in previ-
sous studies, both RA subsets and those had a very minor influ-
ence on the genetic risk of RA.

Notably, only very few genetic variations were associated with
the ACPA-positive RA subset in the Swedish EIRA cohort (table 3). As can be seen from this table, most previously detected genetic variations between the two subsets were not significant for ACPA-negative RA.

Concerning the non-HLA genes, several studies have now
been published describing variations associated with either
ACPA-positive RA or with RA where no discrimination
between the two subsets was made. When analysing the pub-
lically published dataset on ACPA-positive and ACPA-negative 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.
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 RA. 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.

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, 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 Mee, Jameelah B S Mohamed, Jason Ong, Eileen Ping and Sigeeta Rajaram for their invaluable laboratory assistance; the participating rheumatologists at each recruiting centre; and all rheumatologists for recruiting patients in the study. We also thank Catherine Keeling, Paul Oxberry and Mark McMichael 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 (A*STAR), 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.
Kurreeman

Replication of putative candidate-gene


