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 (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 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 $2 \times 10^5$ 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 (ACPA) has proved to be one of the best clinical predictors of the severity of disease course.4 5 In addition, ACPA-positive patients with unspecific 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*03 haplotype was previously suggested20 21 but has not been replicated in a larger study.22 In at least one study23 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.24 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 25 26 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 alone15 17 27 or grouped both subtypes together.14 28

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. 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=9.77\times10^{-9}).\ 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]

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

<table>
<thead>
<tr>
<th>Reference sequence</th>
<th>Chromosome</th>
<th>Closest gene</th>
<th>Minor allele</th>
<th>Allelic counts for cases</th>
<th>Allelic counts for controls</th>
<th>OR (95% CI)</th>
<th>p Value†</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4305317</td>
<td>2</td>
<td>RPS12P4</td>
<td>T</td>
<td>384/1132</td>
<td>691/1403</td>
<td>0.69 (0.59 to 0.80)</td>
<td>2.32E-06</td>
</tr>
<tr>
<td>rs6972219</td>
<td>7</td>
<td>IGFBP1</td>
<td>T</td>
<td>45/1299</td>
<td>20/1960</td>
<td>3.39 (2.00 to 5.78)</td>
<td>1.77E-06</td>
</tr>
<tr>
<td>rs2173035</td>
<td>7</td>
<td>IGFBP1</td>
<td>A</td>
<td>47/1303</td>
<td>21/1957</td>
<td>3.36 (2.00 to 5.65)</td>
<td>1.27E-06</td>
</tr>
<tr>
<td>rs6956740</td>
<td>7</td>
<td>IGFBP1</td>
<td>T</td>
<td>48/1294</td>
<td>21/1973</td>
<td>3.49 (2.08 to 5.85)</td>
<td>5.12E-07</td>
</tr>
<tr>
<td>rs3807306</td>
<td>7</td>
<td>IRF5</td>
<td>T</td>
<td>785/555</td>
<td>974/992</td>
<td>1.44 (1.25 to 1.68)</td>
<td>3.15E-07</td>
</tr>
</tbody>
</table>

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.
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...
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. 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-typing.  

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 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-negative 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. Thus, our present negative
rs4810485

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

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Gene/loci</th>
<th>Minor allele*</th>
<th>EIRA ACPA-positive MAF cases</th>
<th>EIRA ACPA-negative MAF cases</th>
<th>NARAC MAF cases</th>
<th>WTCCC MAF cases</th>
<th>OR (95%CI)†</th>
<th>OR (95%CI)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3890745</td>
<td>1p36</td>
<td>MMEL1-TNFRSF14</td>
<td>C</td>
<td>0.28</td>
<td>0.31</td>
<td>0.29</td>
<td>0.31</td>
<td>0.29</td>
<td>0.32</td>
</tr>
<tr>
<td>rs2476601</td>
<td>1p13</td>
<td>PTPN22</td>
<td>A</td>
<td>0.18</td>
<td>0.12</td>
<td>0.14</td>
<td>0.12</td>
<td>0.16</td>
<td>0.08</td>
</tr>
<tr>
<td>rs7574865</td>
<td>2q32</td>
<td>STAT4</td>
<td>T</td>
<td>0.24</td>
<td>0.22</td>
<td>0.26</td>
<td>0.22</td>
<td>0.25</td>
<td>0.23</td>
</tr>
<tr>
<td>rs3087243</td>
<td>2q33</td>
<td>CTLA4</td>
<td>A</td>
<td>0.34</td>
<td>0.39</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>0.42</td>
</tr>
<tr>
<td>rs2064476</td>
<td>6p21</td>
<td>HLA-DRB1</td>
<td>G</td>
<td>0.20</td>
<td>0.26</td>
<td>0.27</td>
<td>0.26</td>
<td>0.22</td>
<td>0.30</td>
</tr>
<tr>
<td>rs10499194</td>
<td>6q23</td>
<td>OLIG3-TNFAP1A</td>
<td>T</td>
<td>0.19</td>
<td>0.21</td>
<td>0.20</td>
<td>0.21</td>
<td>0.25</td>
<td>0.30</td>
</tr>
<tr>
<td>rs6920220</td>
<td>6q23</td>
<td>OLIG3-TNFAP1A</td>
<td>A</td>
<td>0.25</td>
<td>0.23</td>
<td>0.24</td>
<td>0.23</td>
<td>0.24</td>
<td>0.19</td>
</tr>
<tr>
<td>rs3807306</td>
<td>7q32</td>
<td>IRF5</td>
<td>G</td>
<td>0.45</td>
<td>0.49</td>
<td>0.41</td>
<td>0.49</td>
<td>0.48</td>
<td>0.50</td>
</tr>
<tr>
<td>rs2812378</td>
<td>9p13</td>
<td>CCL21</td>
<td>G</td>
<td>0.35</td>
<td>0.32</td>
<td>0.36</td>
<td>0.32</td>
<td>0.35</td>
<td>0.33</td>
</tr>
<tr>
<td>rs3761847</td>
<td>9q33</td>
<td>TRAF1-C5</td>
<td>G</td>
<td>0.51</td>
<td>0.46</td>
<td>0.48</td>
<td>0.46</td>
<td>0.47</td>
<td>0.38</td>
</tr>
<tr>
<td>rs4750316</td>
<td>10p15</td>
<td>PRKCQ</td>
<td>C</td>
<td>0.17</td>
<td>0.21</td>
<td>0.19</td>
<td>0.21</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>rs1678542</td>
<td>12q13</td>
<td>KIF5A-PIP4K2C</td>
<td>G</td>
<td>0.40</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.36</td>
<td>0.37</td>
</tr>
<tr>
<td>rs1328674</td>
<td>13q14-q21</td>
<td>HTR2A</td>
<td>T</td>
<td>0.04</td>
<td>0.02</td>
<td>0.04</td>
<td>0.02</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>rs4810485</td>
<td>20q13</td>
<td>CD40</td>
<td>T</td>
<td>0.20</td>
<td>0.24</td>
<td>0.24</td>
<td>0.24</td>
<td>0.21</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Statistically significant odds ratios shown in bold.

*Mantel–Haenszel OR for meta-analysis of EIRA ACPA-positive RA/NARAC/WTCCC.
†Mantel–Haenszel OR for EIRA ACPA-positive RA.
‡GWAS 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.13 20 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, RS12P4 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 rs4505317, close to RS12P4 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.17 27 34 35 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,30 whereas the other subset of the same criterion-based disease may share genetic linkages and pathogenetic 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 which 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 S B 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 Jomhøj 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 (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.

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

REFERENCES


