



The Influence of Polygenic Risk Scores on Heritability of Anti-CCP Level in RA

Citation

Cui, J., K. E. Taylor, Y. C. Lee, H. Källberg, M. E. Weinblatt, J. S. Coblyn, L. Klareskog, et al. 2014. "The Influence of Polygenic Risk Scores on Heritability of Anti-CCP Level in RA." *Genes and immunity* 15 (2): 107-114. doi:10.1038/gene.2013.68. <http://dx.doi.org/10.1038/gene.2013.68>.

Published Version

doi:10.1038/gene.2013.68

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:12987317>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Published in final edited form as:

Genes Immun. 2014 March ; 15(2): 107–114. doi:10.1038/gene.2013.68.

The Influence of Polygenic Risk Scores on Heritability of Anti-CCP Level in RA

Jing Cui¹, Kimberly E. Taylor², Yvonne C. Lee¹, Henrik Källberg³, Michael E. Weinblatt¹, Jonathan S. Cobllyn¹, Lars Klareskog³, Lindsey A. Criswell², Peter K. Gregersen¹, Nancy A. Shadick¹, Robert M. Plenge^{1,5,6}, and Elizabeth W. Karlson¹

¹Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA

²Rosalind Russell Medical Research Center for Arthritis, Division of Rheumatology, Department of Medicine, University of California, San Francisco, CA 94117, USA

³Institute of Environmental Medicine, Box 210, Karolinska Institute, 171 77 Stockholm, Sweden

⁴The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, USA

⁵Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA

⁶Medical and Population Genetics Program, Chemical Biology Program, Broad Institute, Cambridge, Massachusetts, 02142, USA

Abstract

Objective—To study genetic factors that influence quantitative anti-cyclic citrullinated peptide (anti-CCP) antibody levels in RA patients.

Methods—We carried out a genome wide association study (GWAS) meta-analysis using 1,975 anti-CCP+ RA patients from 3 large cohorts, the Brigham Rheumatoid Arthritis Sequential Study (BRASS), North American Rheumatoid Arthritis Consortium (NARAC), and the Epidemiological Investigation of RA (EIRA). We also carried out a genome-wide complex trait analysis (GCTA) to estimate the heritability of anti-CCP levels.

Results—GWAS-meta analysis showed that anti-CCP levels were most strongly associated with the human leukocyte antigen (HLA) region with a p-value of 2×10^{-11} for rs1980493. There were 112 SNPs in this region that exceeded the genome-wide significance threshold of 5×10^{-8} , and all were in linkage disequilibrium (LD) with the *HLA-DRB1*03* allele with LD r^2 in the range of 0.25–0.88. Suggestive novel associations outside of the HLA region were also observed for rs8063248 (near the *GP2* gene) with a p-value of 3×10^{-7} . None of the known RA risk alleles (~52 loci) were associated with anti-CCP level. Heritability analysis estimated that 44% of anti-CCP variation was attributable to genetic factors captured by GWAS variants.

Conclusions—Anti-CCP level is a heritable trait. *HLA-DR3* and *GP2* are associated with lower anti-CCP levels.

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding Author: Jing Cui, Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA, Tel: 617 732-5563, Fax: 617 713-3030, jcui3@partners.org.

Keywords

RA; GWAS; anti-CCP; heritability

INTRODUCTION

Anti-cyclic citrullinated peptide (anti-CCP) antibody is an important biomarker for rheumatoid arthritis (RA). It is more sensitive and specific than rheumatoid factor for diagnosing RA. Multiple studies have shown that presence of anti-CCP antibodies and level of anti-CCP in RA correlates with radiographic progression as well as extra-articular manifestations¹⁻⁵. Higher anti-CCP levels are also associated with more active disease, more severe joint damage, worse functional disability and reduced quality of life^{6, 7}, and thus can be considered a proxy for disease severity. Anti-CCP antibody levels are relatively stable over time, have little correlation with disease duration, and change little after treatment⁶, making anti-CCP a good target quantitative trait for genetic studies.

The anti-CCP test is a commercial assay that relies upon reactivity to several modified or synthetic cyclic citrullinated peptides. Research assays have been developed for autoantibodies directed against specific proteins and/or peptides found in the synovium such as vimentin, fibrinogen and enolase⁸⁻¹³ in an effort to identify a specific antigen responsible for inciting RA. Multiplex assays for anti-citrullinated peptide antibodies (ACPA) demonstrate strong correlation between quantitative ACPA count and quantitative CCP levels, suggesting that the CCP assay is a valid proxy for more specific autoantibodies^{12, 13}.

RA heritability estimates suggest that genetic factors explain ~53-68% of RA susceptibility¹⁴. Since anti-CCP level is a measure of RA disease severity, genetic factors may also influence anti-CCP levels. Most previous studies of genetic determinants of anti-CCP have been limited to the human leukocyte antigen (HLA) region^{15, 16} and a small number of RA candidate genes¹⁷. Most studies of the *HLA-DRB1 Shared Epitope (SE)*, the strongest genetic risk factor for RA, have demonstrated consistent findings that the *HLA-SE* is associated with anti-CCP positivity; while some have found that the *HLA-SE* is associated with higher quantitative anti-CCP level specifically^{15, 16}. In contrast to *HLA-SE*, Irigoyen et al reported that *HLA-DRB1*03* was associated with lower anti-CCP levels in a Caucasian population¹⁵, and Bang et al reported that *HLA-DRB1*0901* was associated with lower anti-CCP levels in an Asian population¹⁸. However, the findings related to anti-CCP and other RA risk alleles are not convincing.

Genome wide association studies (GWAS) have been expanding for the past decade, and most RA genetic studies have focused on susceptibility analyses. To date, there are approximately 52+ loci validated as RA risk alleles outside the *HLA-SE* by GWAS and GWAS-meta analysis¹⁹⁻²³. However, very few studies have focused on the genetics of RA disease severity, especially defined by anti-CCP level. Using Affymetrix 100K chip data from 531 patients in the Brigham Rheumatoid Arthritis Sequential Study (BRASS), we previously reported that *HLA-DRB1*03* was associated with lower anti-CCP levels. We also found suggestive associations for SNPs in the HLA region that explained additional anti-CCP variation after adjusting for the *DRB1*03* allele. These results were replicated in the North American Rheumatoid Arthritis Consortium (NARAC)²⁴.

In this study, we extend our GWAS on anti-CCP level to perform a meta-analysis including 3 large cohorts, BRASS, NARAC and the Epidemiological Investigation of RA (EIRA). Our purpose is to identify additional common genetic variants that may influence anti-CCP level to identify novel pathways associated with disease severity in RA. The current study

provides unique information by extending the sample size to improve statistical power and using genotype data from denser and newer versions of the micro-array assay. In addition to GWAS analysis, we also apply two new methods, polygenic analysis and mixed linear modeling analysis, to assess the aggregated effect of common SNPs on anti-CCP level.

RESULTS

All subjects met 1987 ACR classification criteria for RA or were diagnosed by a board-certified rheumatologist. All subjects were anti-CCP positive, assessed by a second generation anti-CCP assay. 474 BRASS subjects, 823 NARAC subjects and 678 EIRA subjects passed all quality controls. The HLA region was the top ranked result with a p value of 2×10^{-11} (Figure 1) from meta-analysis. Table 1 shows the independent top 20 loci after removing the HLA region.

We found no evidence of systematic bias with λ_{GC} equal to 0.997. The Quantile-quantile (Q-Q) plot is shown in Figure 2a. The Q-Q plot removing the HLA region is shown in Figure 2b. A cluster of SNPs outside the HLA region on chromosome 16 were associated with anti-CCP level (Figure 3). Several SNPs, including the top SNP rs8063248, were genotyped in at least one cohort, and effect sizes were consistent among 3 cohorts (small Cochran's Q-value, Supplementary Table 1). The top SNP rs8063248 was 63kb upstream from the *GP2* gene (Glycoprotein 2 /Zymogen granule protein 2), which is an auto-antigen for inflammatory bowel disease²⁵ (IBD, Crohn's disease), $p=3 \times 10^{-7}$.

For the HLA region at (6p22.2 (26.03 Mb) to 6p21.32 (33.59Mb)), the regional association was plotted in Figure 4a. After controlling for the top SNP rs1980493, the association signal was no longer significant (Figure 4b), indicating that HLA associations with anti-CCP level are due to LD with the top SNP, or LD with causal variants in LD with the top SNP.

Using high resolution HLA typing in the NARAC samples, we explored LD between the individual HLA alleles, *HLA-DRB1*0401*, *HLA-DRB1*0404* *HLA-DRB1*03* etc., and top SNPs in the HLA region that were associated with anti-CCP level. Most of the top HLA SNPs were in LD with *HLA-DRB1*03* with the highest $r^2 = 0.88$ (Table 2). *DRB1*03* was associated with low anti-CCP level in NARAC samples with $p=0.0001$, in BRASS samples with $p=0.0003$, in EIRA (imputed) samples with $p=0.0003$, with overall $p=3 \times 10^{-11}$. Table 3 shows the association between *HLA-DRB1* subtype and anti-CCP level. Association between amino acid sites and anti-CCP level also supports that *HLA-DRB1*03* is associated with anti-CCP level (Supplementary Table 2).

No established RA risk alleles²⁶ were associated with anti-CCP level from the meta-analysis after Bonferroni correction (Supplementary Table 3).

For the polygenic analysis, when we used EIRA as the discovery set, the polygenic risk score was significantly associated with anti-CCP level in BRASS. The highest variance explained was 1.7%, using a SNP set with a threshold of $p < 0.05$. The association between the polygenic risk score and anti-CCP level was not significant in NARAC (Figure 5a). We found similar results using BRASS as a discovery set, namely that the polygenic risk score was associated with anti-CCP level in EIRA but not in NARAC (Figure 5b). Using NARAC as the discovery set, the polygenic risk score was not associated with anti-CCP level in either EIRA or BRASS (data not shown). GRS composed using RA risk alleles was not associated with anti-CCP level.

In BRASS, EIRA and NARAC (the largest subset $N=429$ was used), genetic information summarized from the whole genome contributed to 65%, 37% and 41%, respectively, of total anti-CCP variation. The heritability estimates (Table 4) were not significant in BRASS,

EIRA and NARAC ($p=0.12$, 0.15 and 0.2 respectively) due to the large standard error caused by small sample sizes. When we combined all 3 cohorts, heritability was estimated as 44% with a significant of p -value of 0.04 .

DISCUSSION

We carried out a meta-analysis of GWAS for anti-CCP level from 3 cohorts comprising 1,975 subjects. We found that Anti-CCP level is a heritable trait. In addition to the HLA region, suggestive genetic effects outside the HLA region influenced anti-CCP level.

This study expands upon our previous 100K GWAS on anti-CCP titer for 531 BRASS subjects, followed by replication in NARAC only for the top 203 SNP findings. In the current report, we applied a different study design that used more homogeneous samples by restricting for anti-CCP+ cases, and excluding anti-CCP negative disease, which may have a different genetic architecture. We used a newer genotype platform with denser SNPs to achieve better power to identify genetic variants. We used imputation as a bridge to make BRASS, NARAC and EIRA samples comparable for the same set of SNPs. Since anti-CCP is not normally distributed, and measures from different assays are difficult to compare, we chose to use the inverse normal transformation method to rank the anti-CCP level. This normalization makes the estimated SNP effect comparable among cohorts, assuming that each cohort has similar anti-CCP level distributions, so that we can carry out a meta-analysis to combine GWAS results for all 3 cohorts. We also need to interpret the association with transformed anti-CCP level carefully. We would use the magnitude and the sign of beta to interpret the association, negative beta means the genetic variant is associated with lower CCP level, while the larger beta means a larger effect size.

Common HLA-SE risk alleles (*DRB1*0101*, *DRB1*0401*, *DRB1*0404*) were not strongly associated with anti-CCP level ($p \sim 0.01$ in NARAC, $p \sim 0.002$ for all 3 cohorts using imputed HLA type); however *DRB1*03* was strongly associated with lower anti-CCP levels ($p=0.0001$ in NARAC, $p=3 \times 10^{-11}$ for all 3 cohorts), and explained 2.5% of the anti-CCP variance. The amino acid analysis provided additional support for the association between *DRB1*03* (combination of 11S-13S-71K-74R²⁷) and anti-CCP level. Specifically, the amino acid analysis identified *HLA-DRB1* positions 11(Ser, S), 13(Ser, S), 71(Lys, K), 74(Arg, R) as the sites most significantly associated with anti-CCP level. The overall p value for these four amino acid sites are all significant (Supplementary table 2). This observation is consistent with a recent study of $\sim 20,000$ subjects, which reported that the *DRB1*03* allele has a protective effect on RA, with a hazard ratio of 0.63^{27} and with prior studies of *DRB1*03* and anti-CCP level^{15, 24}. LD between the top SNPs and the *DRB1*03* allele are ~ 0.88 in NARAC and ~ 0.57 in BRASS. A regional plot of the results in the HLA region, before and after adjusting for the top SNP, suggests that association signals in the HLA region are due to one variant. However, due to the complexity of LD structure in the HLA region, it is challenging to identify the specific causal variant(s). Further strategies such as functional experiments are needed to discover the true causal gene/variant(s) in the HLA region.

In addition to the HLA region, the regional plot shows that a number of SNPs displayed suggestive associations with anti-CCP level on Chromosome 16 (Figure 3). Given the sample size, our study has limited power to detect moderate effect sizes to achieve genome wide significance. For example, our study had only 29% power to identify associations at the 5×10^{-8} significance level for the top chromosome 16 signal. We would have needed a sample size of 3131 to achieve 80% power. Also, GWAS assesses only common variants with minor allele frequency $\geq 5\%$. However, several criteria were applied to increase the likelihood of a true positive association. Specifically, we required that: 1) some of these

SNPs were genotyped for at least one of the three cohorts, 2) the imputed SNPs had high accuracy scores (info score >0.8), and 3) association effect sizes were consistent among the three cohorts according to Cochran's heterogeneity tests (Supplementary table 1).

The top finding outside HLA was on Chromosome 16, 3' to the *GP2* gene. *GP2* plays an important role in bacterial and viral infections by recognizing antigen, thereby triggering antigen-specific immune responses²⁸. In inflammatory bowel disease, antibodies to *GP2* have been developed as diagnostic tools²⁵. Since *GP2* is immunogenic, it also may play a role in RA pathogenesis. Further studies of the role of *GP2* in RA are necessary.

Literature searches focused on the remaining top SNPs revealed that *ELMO1* is associated with diabetic nephropathy²⁹, whereas *ITGA4* is a candidate gene for celiac disease³⁰. *CD133+* cells are endothelial progenitor cells that are mobilized in vasculogenesis in the inflammatory articular process that occurs in RA³¹. *PIP5K1B* is a candidate gene that may have a critical role in Sézary syndrome, a rare form of cutaneous T-cell lymphoma³². Although the relationship between these genes and anti-CCP is unknown, existing literature linking these genes and immunologic traits supports a role for these genes in RA.

Most prior genetic studies for anti-CCP were focused on anti-CCP status (+/-), showing an association between some RA risk alleles and anti-CCP status^{33, 34}. Besides the reported association of *HLA-DRB1*03* with lower anti-CCP level¹⁵, one paper studied genetic associations with anti-CCP level and reported that one *PADI4* haplotype, carrying a *PADI4* risk allele, was associated with higher anti-CCP level in new-onset RA¹⁷. We examined 52 recently validated RA risk alleles²⁶ in the present study and found no evidence for association with anti-CCP level, suggesting that anti-CCP genetic architecture may differ from disease susceptibility architecture.

Our study does have some limitations. First, our analysis was limited to a study of the CCP test, a commercial assay that tests reactivity to several modified or synthetic cyclic citrullinated peptides³⁵, as opposed to ACPA assays, which test reactivity to individual peptides (e.g., vimentin, fibrinogen, enolase). Serum anti-CCP antibody levels are, however, highly correlated with total serum ACPA responses^{9, 13, 36, 37}.

Second, although anti-CCP levels were measured in a single laboratory for BRASS and EIRA, anti-CCP was measured in three laboratories for NARAC. To examine the potential effects of differing techniques between these laboratories, we performed sensitivity analyses treating each NARAC laboratory subset as a separate cohort and transformed anti-CCP level separately for each subset. GWAS results from the meta-analysis combining data from BRASS, EIRA and the three NARAC sub-datasets were similar to results from the primary analysis, which utilized combined NARAC data (Supplemental figure 1). Because of small sample sizes, we did not use the NARAC subsets as discovery sets for the polygenic model. Because the heritability estimate is sensitive to the homogeneity of phenotype, heritability was estimated only for the largest NARAC sub-datasets (N=429), not the smaller two sub-datasets (N=154, 245 respectively).

In summary, we estimated that heritability for anti-CCP is approximately 44%. Given the modest sample size, the heritability estimates was not highly significant. However, the estimate was similar across cohorts, supporting the veracity of our findings. Independent samples will be needed to confirm our findings. Using the GCTA approach, we estimated that all SNPs from the MHC region explain ~3% of the anti-CCP variance, whereas the polygenic risk score explains an additional ~2% of the variance. Thus, much of the heritability remains unexplained. The challenge is to identify rare variants with larger effect

sizes through next generation sequencing and to find additional common variants with low to moderate effect size using cohorts with larger sample sizes.

Conclusion

We confirmed that HLA is associated with anti-CCP level in our quantitative trait analysis. Specifically, *HLA-DR3* and *GP2* are associated with lower anti-CCP levels. Further studies of the functional role of *HLA-DR3* and *GP2* may provide insights into the mechanisms of disease severity and ultimately lead to targeted therapy if SNPs can predict patients who need more or less intense therapy. Although SNPs outside the HLA region also show modest associations with anti-CCP level, these associations do not reach genome wide levels of significance. Anti-CCP level is a heritable trait in which genetic factors explain ~44% variance based on heritability analysis, where *HLA-DR3* explains ~2.5% of the variance. Based on polygenic results, the additive genetic effect of common alleles outside the HLA region only explains ~2% of the variance. Therefore, much of the heritability is ‘hidden’ and other techniques such as next generation sequencing should be pursued.

METHODS

Subjects

BRASS is a prospective, observational, hospital-based RA cohort established to identify biomarkers and genetic predictors of drug response and disease activity³⁸. Patients were eligible if they had a diagnosis of RA and were older than 18 years. All study protocols were approved by the Brigham and Women’s Hospital Institutional Review Board and informed consent was obtained from all subjects. The cohort is predominantly female (82.3%) with a mean age of 58.0 (\pm 13.6) years and average disease duration of 16.1 (\pm 12.5) years.

NARAC samples were collected by 4 different study designs. “Family collection” samples were from multiplex families (primarily affected sibling pairs) in which at least one sibling had documented joint erosions. The other collections included samples from three prospective observational cohorts, the National Data Bank of Rheumatic Diseases, the National Inception Cohort of Rheumatoid Arthritis, and the Study of New Onset Rheumatoid Arthritis. A more detailed description can be found elsewhere³³. 74.1% of the cohort were female, with a mean age of 56.0 (\pm 12.3) years and average disease duration of 10.6 (11.1) years.

EIRA is a population based case control study carried out in Sweden, comprising residents of south and central Sweden between the ages of 18 and 70 years in 1996 to 2005³⁹. An incident case is defined as a person in the collection who for the first time receives a diagnosis of RA (newly diagnosed cases) according to the 1987 ACR criteria. 72.3% of EIRA cases were females, with a mean age of 50.8(\pm 12.8) years, and all are new onset cases (<2 years from the onset of symptoms, with 85% within 1 year after the first onset of symptoms).

Outcome

Anti-CCP levels were measured using second generation ELISA assays (NARAC and BRASS: Inova Diagnostics; EIRA: Immunoscan-RA Mark2 ELISA). Anti-CCP antibody levels were transformed to a normal distribution by taking the inverse normal of the rank in BRASS, NARAC and EIRA separately. (Supplementary Figure 2 shows the anti-CCP distribution before and after transformation).

Genotyping

BRASS samples were genotyped using the Affymetrix 6.0 chip (900K). NARAC samples were genotyped using the Illumina 550K chip. EIRA samples were genotyped using the Illumina 317K chip. A quality control (QC) process was performed in these 3 datasets separately, where QC steps included filtering SNPs with >1% missing data and individuals with >5% missing data, followed by filtering SNPs with minor allele frequency (MAF) <1% and a Chi-Square test of Hardy Weinberg equilibrium $P_{HWE} < 10^{-5}$. We then used individual pair-wise identity-by-state estimates to remove related and potentially contaminated samples. Data processing and QC were performed in PLINK⁴⁰. To address population stratification and remove outliers, Principle Components Analysis (PCA) was done on the combined dataset, using more than 30K high quality, commonly genotyped SNPs across three cohorts, and genetic outliers were removed using the default setting in EIGENSTRAT⁴¹ (Supplementary Figure 3 shows the PCA result after the outliers were removed).

Imputation

Imputation was conducted on genotype data for each data set separately from IMPUTE software, using haplotype-phased HapMap Phase 2 European CEU founders as the reference panel, allowed us to merge data across different genotype platforms⁴². Imputation yielded posterior genotype probabilities as well as imputation quality scores at SNPs not genotyped with a minor allele frequency 1% in the HapMap CEU. We removed SNPs with imputation quality scores <0.5 from our final analysis. Imputation for HLA type was also conducted in SNP2HLA software⁴³, using the Type 1 Diabetes Genetics Consortium collection as high quality HLA reference panel (N=>5,000)⁴³.

GWAS

Using additive models, association analyses were performed for more than 2 million genotyped/imputed SNPs with transformed anti-CCP values in each dataset. To adjust for residual population stratification effects, the first three eigenvectors from the PCA were controlled for in each model. All association analyses were done using SNPTEST⁴². Inverse variance weighted meta-analysis was utilized to combine GWAS results for all 3 cohorts to identify loci associated with anti-CCP level (N=1,975). We also conducted Cochran's Q tests for heterogeneity across collections using the β coefficients. We did not consider SNPs if they showed strong evidence of heterogeneity across the collections (Cochran's Q $P < 0.001$). Quantile-quantile (Q-Q) plots and genomic control inflation factor- λ_{GC} were employed to illustrate any unmeasured systematic bias. We studied 52 non-HLA risk alleles for RA for association with CCP level based on a recently published meta-analysis²⁶.

Polygenic modeling

Since genome-wide significance criteria are stringent, most genetic variations will not achieve those thresholds. Polygenic modeling is a method to evaluate the aggregate effects of many variants with small effects. Information from different SNPs are combined into a 'risk' score to test whether that score has any association with anti-CCP level⁴⁴. BRASS, NARAC and EIRA were each used as discovery sets using the same GWAS approach, adjusted for the PCA. Linkage disequilibrium (LD) pruning (LD $r^2 > 0.1$) was done with respect to the association p-value to obtain a set of independent SNPs, which were maximally associated with anti-CCP level. The PRS then was computed in each of the other two test sets. SNP sets were selected for nine different threshold p-values ($p < 10^{-5}$, 10^{-4} , 10^{-3} , 0.01, 0.05, 0.1, 0.2, 0.3, 0.4). For each SNP set, beta-weighted allele counts were summed into a polygenic risk score. The polygenic risk score was calculated for each validation set individual i as

$$PRS_i = \sum_{j \in SNPs} \hat{\beta}_j d_{ij}$$

where $\hat{\beta}_j > 0$ is the discovery GWAS β -coefficient with respect to anti-CCP level, and d_{ij} is the individual i 's posterior probability dosage (from zero to two) of that allele. The resulting PRS was tested for association with anti-CCP level using linear regression modeling, adjusting for 3 eigenvectors. P-values for the association between polygenic risk score and anti-CCP level, the numbers of SNPs and estimated variance explained were reported. Specific PRS (genetic risk score, GRS) was also composed using RA risk alleles. Association between GRS and anti-CCP level was conducted.

Mixed linear model

Heritability was directly estimated from the BRASS, NARAC and EIRA GWAS datasets. Genome genotype data and imputed SNP data were summarized into a Kinship matrix described by Yang et al.⁴⁵, where A_{ij} is the relationship estimate for subject i and j for i, j . Genetic variance then was estimated as the random effect coefficient of the kinship matrix in a MIXED model using the genome-wide complex trait analysis (GCTA) software⁴⁵. The heritability estimate variance was calculated from the inverse of the variance-covariance matrix, and significance was assessed by the Wald test.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Syversen SW, Gaarder PI, Goll GL, Odegard S, Haavardsholm EA, Mowinckel P, et al. High anti-cyclic citrullinated peptide levels and an algorithm of four variables predict radiographic progression in patients with rheumatoid arthritis: results from a 10-year longitudinal study. *Annals of the rheumatic diseases*. 2008; 67(2):212–7. [PubMed: 17526555]
2. Riedemann JP, Munoz S, Kavanaugh A. The use of second generation anti-CCP antibody (anti-CCP2) testing in rheumatoid arthritis—a systematic review. *Clinical and experimental rheumatology*. 2005; 23(5 Suppl 39):S69–76. [PubMed: 16273788]
3. Syversen SW, Goll GL, van der Heijde D, Landewe R, Lie BA, Odegard S, et al. Prediction of radiographic progression in rheumatoid arthritis and the role of antibodies against mutated citrullinated vimentin: results from a 10-year prospective study. *Annals of the rheumatic diseases*. 2010; 69(2):345–51. [PubMed: 19648126]
4. Im CH, Kang EH, Ryu HJ, Lee JH, Lee EY, Lee YJ, et al. Anti-cyclic citrullinated peptide antibody is associated with radiographic erosion in rheumatoid arthritis independently of shared epitope status. *Rheumatology international*. 2009; 29(3):251–6. [PubMed: 18751982]
5. Turesson C, Jacobsson LT, Sturfelt G, Matteson EL, Mathsson L, Ronnelid J. Rheumatoid factor and antibodies to cyclic citrullinated peptides are associated with severe extra-articular manifestations in rheumatoid arthritis. *Annals of the rheumatic diseases*. 2007; 66(1):59–64. [PubMed: 16901955]
6. Mjaavatten MD, van der Heijde D, Uhlig T, Haugen AJ, Nygaard H, Sidenvall G, et al. The likelihood of persistent arthritis increases with the level of anti-citrullinated peptide antibody and immunoglobulin M rheumatoid factor: a longitudinal study of 376 patients with very early undifferentiated arthritis. *Arthritis research and therapy*. 2010; 12(3):R76. [PubMed: 20444271]
7. Ibn Yacoub Y, Amine B, Laataris A, Hajjaj-Hassouni N. Rheumatoid factor and antibodies against citrullinated peptides in Moroccan patients with rheumatoid arthritis: association with disease parameters and quality of life. *Clinical rheumatology*. 2012; 31(2):329–34. [PubMed: 21814754]

8. Montes A, Perez-Pampin E, Calaza M, Gomez-Reino JJ, Gonzalez A. Association of anti-citrullinated vimentin and anti-citrullinated alpha-enolase antibodies with subsets of rheumatoid arthritis. *Arthritis and rheumatism*. 2012; 64(10):3102–10. [PubMed: 22674012]
9. Lundberg K, Bengtsson C, Kharlamova N, Reed E, Jiang X, Kallberg H, et al. Genetic and environmental determinants for disease risk in subsets of rheumatoid arthritis defined by the anticitrullinated protein/peptide antibody fine specificity profile. *Annals of the rheumatic diseases*. 2012; 72(5):652–8. [PubMed: 22661643]
10. Snir O, Widhe M, von Spee C, Lindberg J, Padyukov L, Lundberg K, et al. Multiple antibody reactivities to citrullinated antigens in sera from patients with rheumatoid arthritis: association with HLA-DRB1 alleles. *Annals of the rheumatic diseases*. 2009; 68(5):736–43. [PubMed: 18635594]
11. van de Stadt LA, de Koning MH, van de Stadt RJ, Wolbink G, Dijkmans BA, Hamann D, et al. Development of the anti-citrullinated protein antibody repertoire prior to the onset of rheumatoid arthritis. *Arthritis and rheumatism*. 2011; 63(11):3226–33. [PubMed: 21792832]
12. Sokolove J, Bromberg R, Deane K, Lahey L, Derber L, Chandra P, et al. Autoantibody epitope spreading in the pre-clinical phase predicts progression to rheumatoid arthritis. *PLoS One*. 2012; 7(5):e35296. [PubMed: 22662108]
13. Arkema EVGB, Robinson WH, Cramb C, Sokolove J, Cui J, Malspeis S, Karlson EW, Costenbader KH. Anti-citrullinated peptide autoantibodies to rheumatoid synovium epitopes in women and risk of future rheumatoid arthritis. *Arthritis research and therapy*. 2013; 15(5):R159. [PubMed: 24286474]
14. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature genetics*. 2012; 44(5):483–9. [PubMed: 22446960]
15. Irigoyen P, Lee AT, Wener MH, Li W, Kern M, Batliwalla F, et al. Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles. *Arthritis and rheumatism*. 2005; 52(12):3813–8. [PubMed: 16320316]
16. van Gaalen FA, van Aken J, Huizinga TW, Schreuder GM, Breedveld FC, Zanelli E, et al. Association between HLA class II genes and autoantibodies to cyclic citrullinated peptides (CCPs) influences the severity of rheumatoid arthritis. *Arthritis and rheumatism*. 2004; 50(7):2113–21. [PubMed: 15248208]
17. Cha S, Choi CB, Han TU, Kang CP, Kang C, Bae SC. Association of anti-cyclic citrullinated peptide antibody levels with PADI4 haplotypes in early rheumatoid arthritis and with shared epitope alleles in very late rheumatoid arthritis. *Arthritis and rheumatism*. 2007; 56(5):1454–63. [PubMed: 17469103]
18. Bang SY, Lee HS, Lee KW, Bae SC. Interaction of HLA-DRB1*09:01 and *04:05 with smoking suggests distinctive mechanisms of rheumatoid arthritis susceptibility beyond the shared epitope. *The Journal of rheumatology*. 40(7):1054–62. [PubMed: 23637323]
19. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature genetics*. 2010
20. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burtt NP, et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nature genetics*. 2008; 40(10):1216–23. [PubMed: 18794853]
21. Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, Guiducci C, et al. Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. *Nature genetics*. 2009; 41(12):1313–8. [PubMed: 19898481]
22. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, Suzuki A, et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nature genetics*. 2012; 44(5): 511–6. [PubMed: 22446963]
23. Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS genetics*. 2011; 7(2):e1002004. [PubMed: 21383967]

24. Cui J, Taylor KE, Destefano AL, Criswell LA, Izmailova ES, Parker A, et al. Genome-wide association study of determinants of anti-cyclic citrullinated peptide antibody titer in adults with rheumatoid arthritis. *Molecular medicine*. 2009; 15(5-6):136–43. [PubMed: 19287509]
25. Roggenbuck D, Reinhold D, Wex T, Goihl A, von Arnim U, Malfertheiner P, et al. Autoantibodies to GP2, the major zymogen granule membrane glycoprotein, are new markers in Crohn's disease. *Clinica chimica acta; international journal of clinical chemistry*. 2011; 412(9-10):718–24.
26. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature genetics*. 2012; 44(12):1336–40. [PubMed: 23143596]
27. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature genetics*. 2012; 44(3):291–6. [PubMed: 22286218]
28. Dias JM, Kuehne AI, Abelson DM, Bale S, Wong AC, Halfmann P, et al. A shared structural solution for neutralizing ebolaviruses. *Nature structural and molecular biology*. 2011; 18(12):1424–7.
29. Mooyaart AL, Valk EJ, van Es LA, Bruijn JA, de Heer E, Freedman BI, et al. Genetic associations in diabetic nephropathy: a meta-analysis. *Diabetologia*. 2011; 54(3):544–53. [PubMed: 21127830]
30. Garner CP, Murray JA, Ding YC, Tien Z, van Heel DA, Neuhausen SL. Replication of celiac disease UK genome-wide association study results in a US population. *Human molecular genetics*. 2009; 18(21):4219–25. [PubMed: 19648293]
31. Jodon de Villeroche V, Avouac J, Ponceau A, Ruiz B, Kahan A, Boileau C, et al. Enhanced late-outgrowth circulating endothelial progenitor cell levels in rheumatoid arthritis and correlation with disease activity. *Arthritis research and therapy*. 2010; 12(1):R27. [PubMed: 20158894]
32. Caprini E, Cristofaletti C, Arcelli D, Fadda P, Citterich MH, Sampogna F, et al. Identification of key regions and genes important in the pathogenesis of Sezary syndrome by combining genomic and expression microarrays. *Cancer research*. 2009; 69(21):8438–46. [PubMed: 19843862]
33. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study. *N Engl J Med*. 2007; 357(12):1199–209. [PubMed: 17804836]
34. Karlson EW, Chibnik LB, Cui J, Plenge RM, Glass RJ, Maher NE, et al. Associations between human leukocyte antigen, PTPN22, CTLA4 genotypes and rheumatoid arthritis phenotypes of autoantibody status, age at diagnosis and erosions in a large cohort study. *Annals of the rheumatic diseases*. 2008; 67(3):358–63. [PubMed: 17666451]
35. Whiting PF, Smidt N, Sterne JA, Harbord R, Burton A, Burke M, et al. Systematic review: accuracy of anti-citrullinated Peptide antibodies for diagnosing rheumatoid arthritis. *Ann Intern Med*. 2010; 152(7):456–64. W155–66. [PubMed: 20368651]
36. Ioan-Facsinay A, el-Bannoudi H, Scherer HU, van der Woude D, Menard HA, Lora M, et al. Anti-cyclic citrullinated peptide antibodies are a collection of anti-citrullinated protein antibodies and contain overlapping and non-overlapping reactivities. *Annals of the rheumatic diseases*. 2011; 70(1):188–93. [PubMed: 20736390]
37. Sokolove JBR, Deane KD, et al. Autoantibody epitope spreading in the pre-clinical phase predicts progression to rheumatoid arthritis. *PLoS One*. 2012; 7(5):e35296. [PubMed: 22662108]
38. Shadick NA, Heller JE, Weinblatt ME, Maher NE, Cui J, Ginsburg G, et al. Opposing effects of the D70 mutation and the shared epitope in HLA-DR4 on disease activity and certain disease phenotypes in rheumatoid arthritis. *Annals of the rheumatic diseases*. 2007; 66(11):1497–502. [PubMed: 17491100]
39. Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. *Arthritis and rheumatism*. 2004; 50(10):3085–92. [PubMed: 15476204]
40. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007; 81(3):559–75. [PubMed: 17701901]

41. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*. 2006; 38(8):904–9. [PubMed: 16862161]
42. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet*. 2007; 39(7):906–13. [PubMed: 17572673]
43. Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One*. 2013; 8(6):e64683. [PubMed: 23762245]
44. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460(7256):748–52. [PubMed: 19571811]
45. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*. 2010; 42(7):565–9. [PubMed: 20562875]

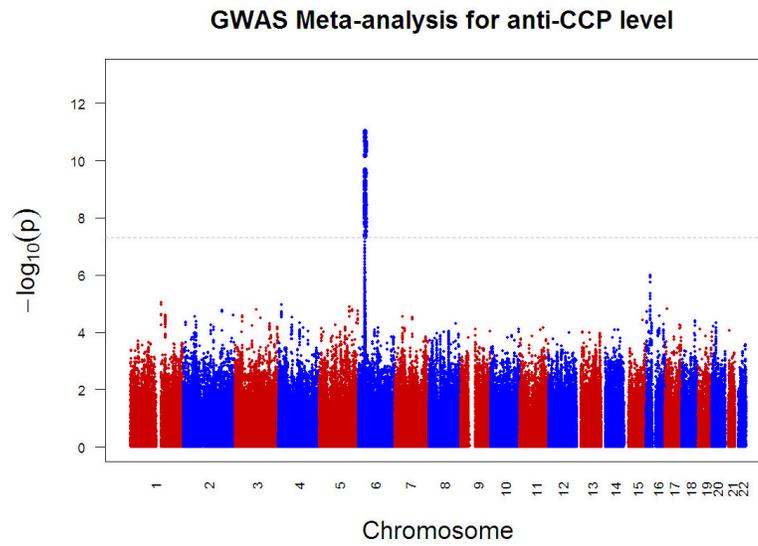


Figure 1. Genome-wide association analysis results for anti-CCP level.
Shown are strengths of association ($-\log_{10} P$ -value) for each SNP versus position along chromosomes 1 to 22.

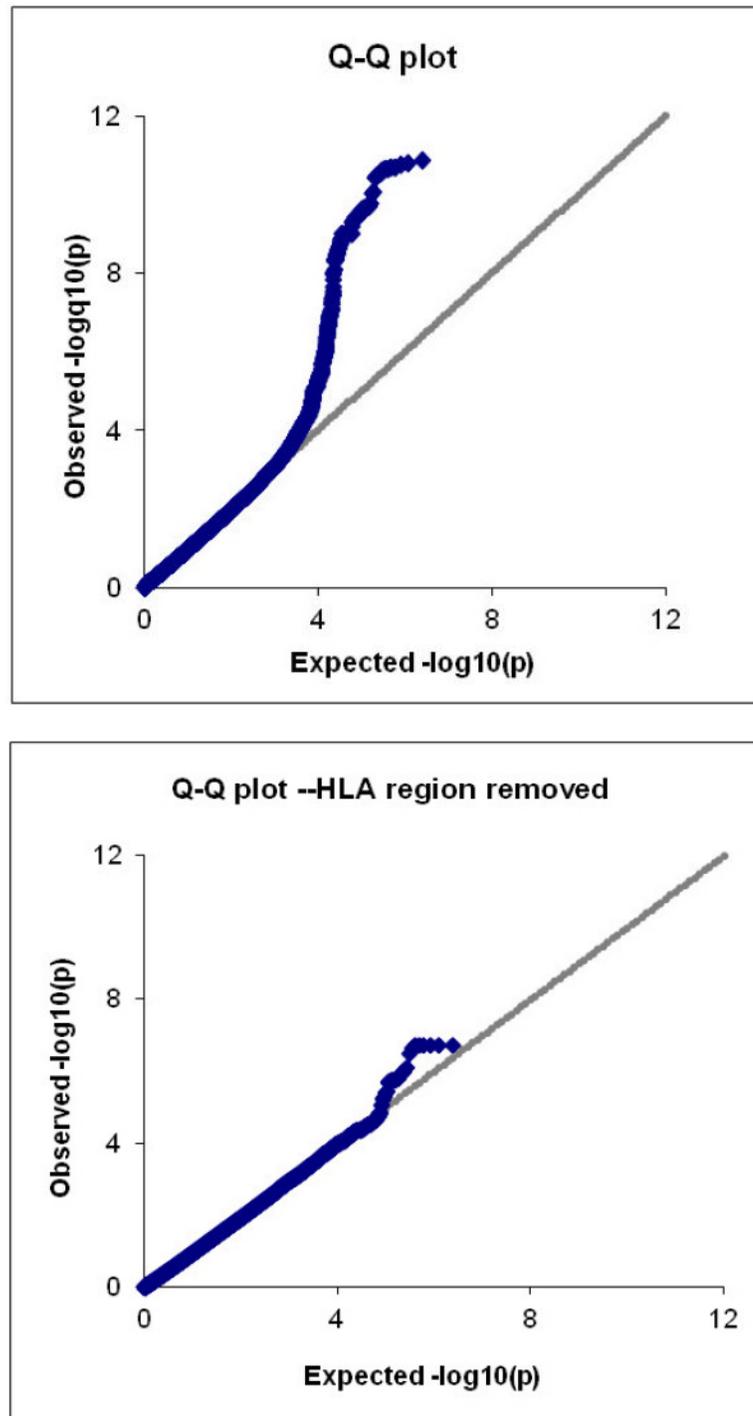


Figure 2. Quantile-quantile (QQ) plots.

Shown are comparisons between observed $-\log_{10}(p)$ and expected $-\log_{10}(p)$. (a) QQ plot for all SNPs; (b) QQ plot after HLA region removed.

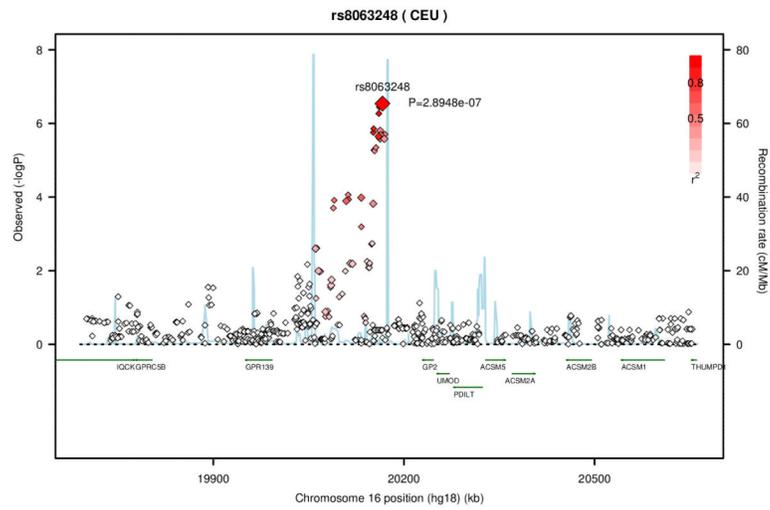


Figure 3. Regional association plot.

Showing strengths of association ($-\log_{10} P$ -value) versus position around rs8063248 along chromosome 16.

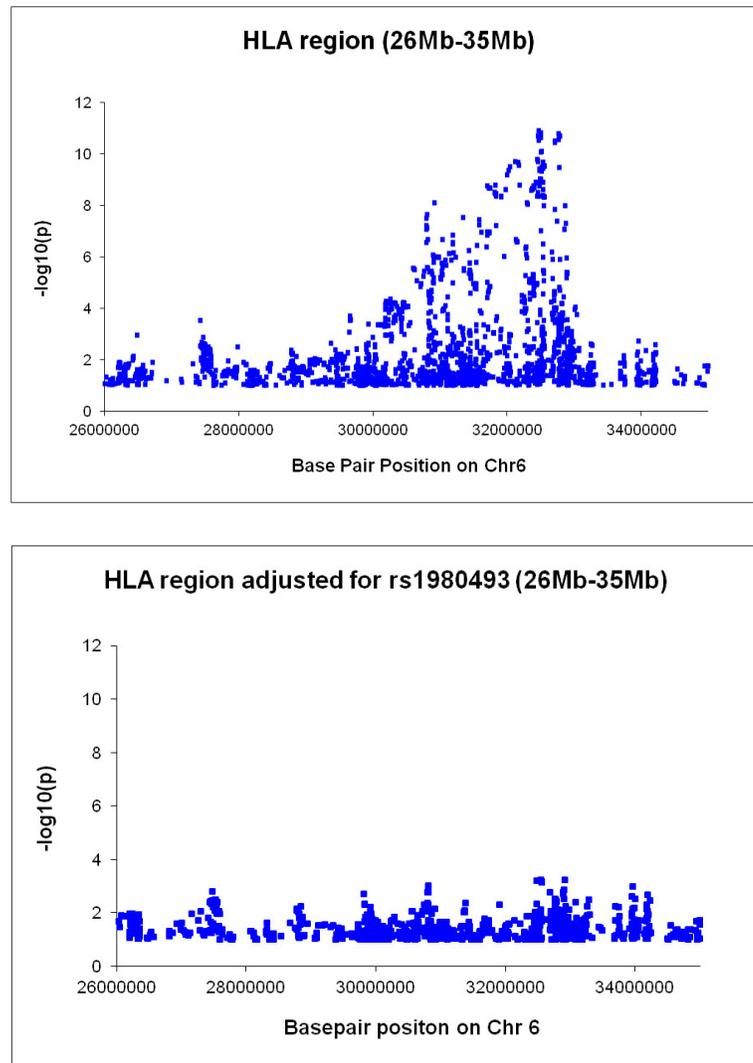


Figure 4. Association results for HLA region. (a) Strengths of association ($-\log_{10} P$ -value) versus position along HLA region; (b) Association ($-\log_{10} P$ -value) versus position along HLA region adjusted for the top SNP rs1980493.

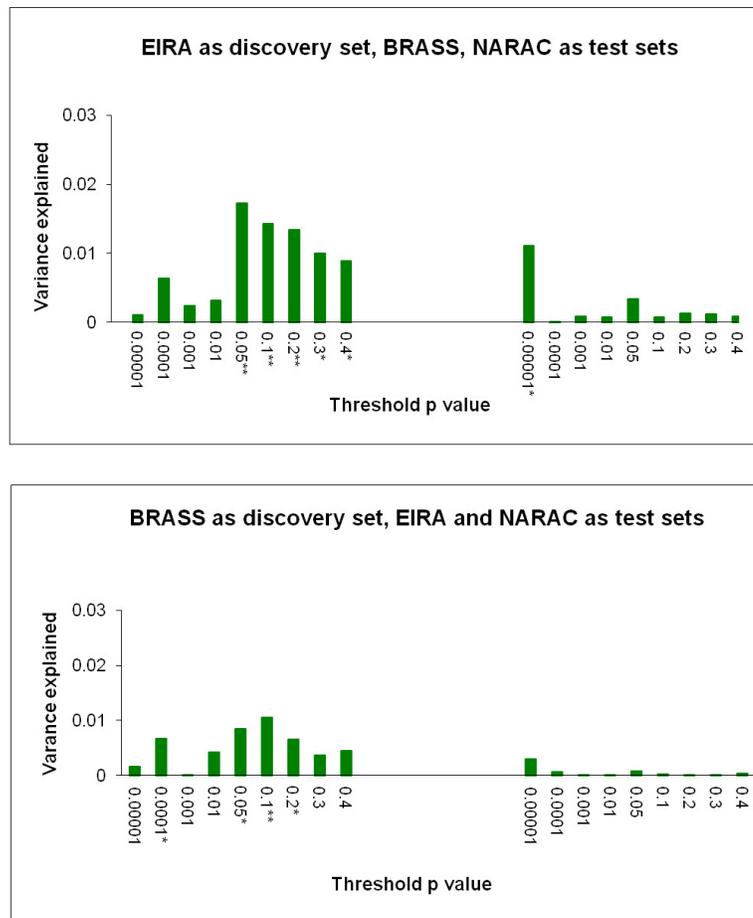


Figure 5. Association of polygenic risk scores with anti-CCP levels in the independent test datasets.

Variance explained by the PRS in the target samples for SNP set derived from discovery dataset for nine significance thresholds ($p < 10^{-5}$, 10^{-4} , 10^{-3} , 0.01, 0.05, 0.1, 0.2, 0.3, 0.4). The y axis indicates R^2 . The significance level was indicated next to the threshold value. (a) EIRA as discovery set, BRASS, NARAC as test sets; (b) BRASS as discovery set, EIRA, NARAC as test sets. *: P -value 0.05; **: P -value 0.01.

Table 1
Top 20 loci associated with anti-CCP level outside the MHC region.

Chromosome	SNP	gene	bp	ma ^a	maf ^b	beta	p ^c	q ^d	Genotyped ^e
16	rs8063248	<i>GP2</i>	20166228	A	0.08	-0.29	2.9E-07	0.61	1
7	rs11766683	<i>ELMO1</i>	36801906	C	0.48	0.14	5.6E-06	0.44	0
2	rs155124	<i>ITGA4</i>	182016183	G	0.04	-0.42	1.0E-05	0.50	0
4	rs10489063	<i>CD133</i>	15926529	A	0.08	-0.25	1.2E-05	0.80	0
9	rs12345359	<i>PIP5K1B</i>	70777493	A	0.01	0.65	1.2E-05	0.92	0
17	rs8067785	<i>ZNF18</i>	11871706	G	0.15	-0.19	1.4E-05	0.46	1
2	rs4669784	<i>NTSR2</i>	11955419	C	0.38	0.15	1.8E-05	0.81	0
4	rs13143069		63108288	G	0.47	0.14	2.2E-05	0.54	0
3	rs4527407	<i>SLITRK3</i>	166693890	C	0.40	-0.14	2.5E-05	0.95	0
14	rs4903534	<i>C14orf4</i>	76551456	A	0.47	-0.13	2.8E-05	0.48	1
10	rs1147594	<i>SLIT1</i>	98955091	A	0.01	-0.57	3.0E-05	0.32	0
10	rs1907336		77824762	G	0.16	0.18	3.1E-05	0.89	1
2	rs1530875	<i>B3GNT7</i>	231907722	C	0.14	0.24	3.2E-05	0.41	0
5	rs2024141	<i>CENTD3</i>	140745184	A	0.08	0.25	3.2E-05	0.47	0
3	rs12485750	<i>OSBPL10</i>	31392045	A	0.09	-0.24	3.4E-05	0.30	0
18	rs7227796		60653976	A	0.12	0.20	3.5E-05	0.80	0
1	rs662997	<i>BCL9</i>	145488076	A	0.03	0.39	3.7E-05	0.79	0
2	rs12615247		59040711	A	0.28	0.14	3.8E-05	0.58	1
11	rs1469170	<i>ARHGAP20</i>	110246224	A	0.30	-0.14	4.0E-05	0.88	1
9	rs12342611		109800925	C	0.48	0.13	4.3E-05	0.43	1

^a ma: minor allele

^b maf: minor allele frequency

^c p value is adjusted for 3 eigenvectors

^d Cochran's Q tests for heterogeneity

^e 1: genotyped in any of the three datasets

Table 2
Linkage disequilibrium between top MHC region SNPs and HLA-DRB1 alleles.

rs	association with anti-CCP level p [*]	drb0101	drb0401	drb0404	drb03
rs2856674	2.39E-11	0.0001	0.02	0.006	0.88
rs9275576	2.86E-11	0.0001	0.02	0.006	0.86
rs2647044	3.19E-11	0.0001	0.02	0.006	0.86
rs1794520	4.20E-11	0.0001	0.01	0.006	0.85
rs3135353	3.58E-11	0.0004	0.02	0.008	0.71
rs3129861	2.69E-11	0.0009	0.02	0.008	0.69
rs2227138	3.66E-11	0.0004	0.02	0.008	0.69
rs1980493	2.23E-11	0.0009	0.02	0.01	0.64
rs3117097	3.23E-11	0.0009	0.02	0.01	0.64
rs3129953	4.71E-11	0.0009	0.02	0.01	0.64

* p-value adjusted for 3 eigenvectors

Table 3
Association between *HLA-DRB1* subtype and anti-CCP level.

HLA DRB1 type	frequency	beta	se	p*	q**
DRB1_03	0.08	-0.40	0.06	3.3E-11	0.59
DRB1_11	0.04	0.22	0.08	0.00976	0.31
DRB1_15	0.09	0.13	0.05	0.014	0.33
DRB1_01	0.15	0.10	0.05	0.026	0.11
DRB1_0101	0.13	0.13	0.05	0.010	0.13
DRB1_08	0.02	-0.24	0.12	0.052	0.89
DRB1_10	0.02	0.17	0.11	0.11	0.62
DRB1_07	0.05	0.09	0.07	0.20	0.63
DRB1_13	0.04	-0.10	0.08	0.24	0.01
DRB1_04	0.46	-0.02	0.03	0.56	0.42
DRB1_0401	0.31	0.04	0.04	0.26	0.57
DRB1_0404	0.10	-0.17	0.06	0.002	0.13
DRB1_0408	0.02	0.26	0.12	0.03	0.15
DRB1_09	0.01	-0.07	0.13	0.61	0.10

* p value is adjusted for 3 eigenvectors

** Cochran's Q tests for heterogeneity

Table 4
Heritability estimates for quantitative anti-CCP level.

Cohort	h^2 *	se	p
BRASS	0.65	0.54	0.12
EIRA	0.37	0.36	0.15
NARAC	0.41	0.58	0.2
BRASS, NARAC, EIRA	0.44	0.27	0.04

* h^2 : heritability (0- 1)