Genome-wide association study in a Korean population identifies six novel susceptibility loci for rheumatoid arthritis


Objective Genome-wide association studies (GWAS) in rheumatoid arthritis (RA) have discovered over 100 RA loci, explaining patient-relevant RA pathogenesis but showing a large fraction of missing heritability. As a continuous effort, we conducted GWAS in a large Korean RA case-control population.

Methods We newly generated genome-wide variant data in two independent Korean cohorts comprising 4068 RA cases and 36487 controls, followed by a whole-genome imputation and a meta-analysis of the disease association results in the two cohorts. By integrating publicly available omics data with the GWAS results, a series of bioinformatic analyses were conducted to prioritise the RA-risk genes in RA loci and to dissect biological mechanisms underlying disease associations.

Results We identified six new RA-risk loci (SLAMF6, CCL13, SWAP70, NFKBIA, ZFP36L1 and LINC00158) that reached the genome-wide significance threshold (p=5×10^{-8}) and consistent disease effect sizes in the two cohorts. A total of 122 genes were prioritised from the 6 novel and 13 replicated RA loci based on physical distance, regulatory variants and chromatin interaction. Bioinformatics analyses highlighted potentially RA-relevant tissues (including immune tissues, lung and small intestine) with tissue-specific expression of RA-associated genes and suggested the immune-mediated gene sets (such as CD40 pathway, IL-21-mediated pathway and citrullination) and the risk-allele sharing with other diseases.

Conclusion This study identified six new RA-associated loci that contributed to better understanding of the genetic aetiology and biology in RA.

INTRODUCTION
Rheumatoid arthritis (RA) is a complex auto-immune disorder characterised by chronically inflamed joints and autoantibody production. A combination between genetic background and environmental triggers confers an increased risk for RA. The overall heritability of RA has been estimated to be about 50%–65% in twin studies. Genome-wide association studies (GWAS) have discovered over 100 RA-associated genetic variants in multiple ancestries. Thus, these loci collectively explained only a small portion (about 15%) of the phenotypic variance for RA. Thus, continuous effort in identifying additional RA variants is necessary to better understand the disease aetiology. Despite the substantial missing fraction of heritability, the reported RA loci have provided insights into the pathogenesis of RA by the so-called post-GWAS analysis that integrates GWAS data with multiple biological resources.

Here, we performed GWAS to identify novel loci exceeding the genome-wide significance threshold (p=5×10^{-8}) in Korean cohorts comprising 4068 RA cases and 36487 controls, followed by
subsequent post-GWAS analyses for prioritising RA-relevant genes and identifying variant-highlighted biology.

**METHODS**

**GWAS participants**

A total of 4068 RA cases and 36487 healthy controls from two independent case–control cohorts were analysed (3177 RA cases and 32820 controls in cohort #1; 891 RA cases and 3667 controls in cohort #2). All the cases were recruited from eight participating university hospitals in Korea and diagnosed through the 1987 revised American College of Rheumatology RA classification criteria.6 Anti-citrullinated protein antibodies (ACPs) were positive in 83.2%, negative in 14.6% and not examined in 2.2% of participants. The controls were recruited through the KoGES and Hanyang University Hospital for Rheumatic Diseases. The genomic DNAs from the KoGES samples were stored in the National Biobank of Korea. All participants provided written informed consent for the study and the Institutional Review Board of Hanyang University approved this study.

**Genotyping and whole genome imputation**

Cohort #1 was newly genotyped with a customised genotyping array, Korea Biobank Array (KoreanChip). Genotyping of the RA cases of cohort #2 was performed on an Illumina HumanOmni-2.5Exome-8 BeadChip, while the control genotype data were produced using Illumina Human Omni1-Quad BeadChip. The overlapping variants between cases and controls in cohort #2 were merged and used in the downstream analyses. The genotyping data for each cohort were filtered based on the general criteria to retain good-quality genetic data (~465 K variants in set #1 of 3177 cases and 32820 controls; ~559 K variants in set #2 of 891 cases and 3667 controls) that showed a high call rate per individual and variant (≥0.99), Hardy-Weinberg equilibrium (P_HWE ≥5×10^{-6}), no excessive heterozygosity, no difference in call rates per variant between cases and controls (p≥5×10^{-6}), no cryptic first-degree relatedness among individuals, homogeneously genetic background among individuals and minor allele frequency (MAF) ≥0.05.

Imputation for autosomal variants was performed by Eagle28 and IMPUTE49 using the reference panel constructed from the 1000 Genomes Project (1KGP) phase 310 reference panel and whole genome sequencing data for 397 Koreans (Korean Personal Genome Project).11 Each imputation chunk included 10 000 variants with a 2 Mb buffer region. Imputation for the non-pseudo-autosomal region of the X chromosome was performed by Shapeit12 and Minimac313 with the 1KGP phase 310 reference panel, making imputation chunks of 5 Mb regions with 1 Mb buffer. The numbers of the imputed variants with good quality (showing imputation quality index ≥0.3, MAF ≥0.5% and p_{HWE} in control ≥1×10^{-6}) were 10 934 705 in cohort #1 and 10 868 813 in cohort #2. These variants were analysed in the subsequent analyses.

**Association analysis**

The genetic association between RA and each autosomal variant was tested by logistic regression adjusting for gender and the top four principal components using EPACTS.14 The association analysis for the X chromosome variants was performed separately for females and males by the same model excluding the gender variable. An inverse-variance-weighted fixed-effects meta-analysis for the genomic-control association results in the two independent cohorts were conducted using METAL.15 To investigate statistically independent association signals within each RA locus, we performed a conditional analysis by GCTA-COJO16 using the association summary statistics based on the reference linkage disequilibrium (LD) calculated in cohort #1.

**Gene set and tissue-specific expression analysis**

Gene-level association p values for 19 840 protein-coding genes were calculated by MAGMA17 using a variant-wide mean model based on variant-level association summary statistics within and around genes. Statistical significance for the association of the gene-level Z scores (converted from gene-level p values) with MSigDB18 gene sets including curated gene sets and gene ontology terms and tissue-specific gene expression in 54 tissues in Genotype-Tissue Expression (GTEX) v8 RNA Sequencing (RNA-Seq) data19 were tested according to the MAGMA17 regression models.

**Enrichment analysis for tissue-specific histone modifications**

GREGOR20 was deployed to calculate the enrichment estimates of RA variants within four tissue-specific histone marks (H3K4me1, H3K4me3, H3K27ac and H3K27me3) from Roadmap Epigenomics Project.21

**Gene prioritisation in RA loci**

For each RA locus, we identified a lead variant with the lowest p_{min} value within a physically (±300 kb) or genetically (r^2>0.1) defined region. FUMA22 was then employed to identify the most likely RA-relevant genes based on the lead RA-risk variants and the meta-analysis association summary statistics according to the following three mapping strategies—(1) Positional mapping: this strategy found genes within and 10 kb around the region containing a lead variant and its proxy variants in each RA locus. (2) Expression quantitative trait locus (eQTL) mapping: this mapping collected the genes regulated by known eQTL in LD with lead RA-risk variants in blood, spleen, lung and small intestine. The known eQTL was retrieved from GTEx,19 single-cell RNA-seq data in peripheral blood mononuclear cells,23 and DICE immune-cell data.24 (3) Chromatin interaction mapping: this approach found genes making promoter-involved or enhancer-involved chromatin interactions with the region with RA-risk variants in blood, spleen, lung and small intestine based on the GSE87112 (Hi-C) data25 and FANTOM526 annotations.

**Variant-based heritability**

An LD score regression method27 was performed to estimate the genome-wide and partitioned heritability,28 using East Asian LD scores from the 1KGP phase 310 and the association summary statistics, assuming an RA prevalence of 0.01.

**Ancestral genetic correlation of RA associations between Koreans and Europeans**

A correlation of RA GWAS between Korean and European5 populations was examined by the Popcorn programme accounting for the ancestry-specific LD scores and cross-covariance scores in two ancestries in the 1KGP.

**Correlation analysis of genetic architectures among diseases**

We assessed the correlation of genome-wide disease association between RA and other diseases (n=28; from Biobank Japan Project30 31 and Bentham et al32) using the LD score regression method.32 The association summary statistics in 27 diseases excluding systemic lupus erythematosus were estimated from Japanese populations, whose LD structure is highly consistent with Korean ancestry.33 The data for systemic lupus erythematosus were from the 101136 Japanese Project.28 On 28 July 2020.
erythematous were generated from European case–control populations. Each set of genome-wide association summary statistics was generated using >1000 cases and displayed at least one disease locus with p<5×10⁻⁸.

RESULTS
Identification of six novel RA loci
To identify new RA risk loci in a Korean population, we performed a fixed-effects meta-analysis of the RA association statistics of the newly generated, imputed variants in two independent Korean cohorts comprising 4068 RA cases and 36487 controls. There were 10404202 intersecting variants between the two cohorts. The genomic control inflation factors (λGC) of the meta-analysis results including and excluding the major histocompatibility complex (MHC) region were estimated to be 1.035 and 1.020, respectively, showing no substantial population stratification in a quantile–quantile plot (online supplementary figure 1).

We identified 19 loci surpassing the genome-wide significance threshold (figure 1 and table 1). Of these loci, six (SLAMF6, CXCL13, SWAP70, NFKBIA, ZFP36L1 and LINC00158; figure 2) were novel loci that showed highly consistent effect sizes on the risk of RA in the two cohorts (table 1). The other 13 loci including the MHC region with the most significant association in the human genome had been reported in other RA genetic studies14 (figure 1 and table 1). No secondary association signals in novel RA loci were detected in conditional analyses.

The overall variant-based heritability for RA on the liability scale was estimated to be 43.04%. The contribution of the non-MHC regions to heritability was 18.60%. When partitioned by the known non-MHC risk loci and the six novel loci, the heritability for RA was 6.65% from the known loci and 0.65% from the novel loci. The genetic correlation of RA associations between Korean and European populations was strong in a genome-wide level (correlation coefficient r=0.76; p=1.3×10⁻⁸) and in 100 known and six novel RA variants (r=0.76; online supplementary table 3 and figure 4).

Asian-specific RA-risk missense variant in the SH2B3 locus
Among the replicated known loci in this study, the RA association of the SH2B3 locus was previously explained by the intergenic variant rs10774624 in European populations5 that were neither common (MAF=0.56%) nor associated with RA (ptext=0.64) in Korean populations (online supplementary table 3). However, we discovered that the SH2B3 missense variant (rs78894077, c.724C>T of exon 2; p.Phe242Ser of the PH domain) was associated with RA in the same locus (pgen=3.89×10⁻⁷). Interestingly, the RA-risk allele T (p.242Ser) of rs78894077 was frequent in Korean populations (MAF=7.43%) but absent in other non-Asian ancestries according to the 1KGP phase 3 data. The RA-risk allele of the same variant in SH2B3 was also reported to associate with the increased risk of myeloproliferative neoplasms in East Asian populations.14 35

Tissues and gene sets implicated by genome-wide association results
We calculated gene-level disease association p values from the variant-level association summary statistics using MAGMA19 and tested for the association between gene-level Z scores (converted from the gene-level p values) and tissue-specific expression level to provide deeper insights into the underlying disease-driving tissues in RA development. The analysis revealed five significant tissues with tissue-specific expression of RA-associated genes, including spleen (p=1.50×10⁻⁹), Epstein-Barr virus (EBV)-transformed lymphocytes (p=5.32×10⁻⁹), whole blood (p=1.12×10⁻⁸), lung (p=2.99×10⁻⁸) and small intestine (p=7.36×10⁻⁸), among 54 different tissues based on GTEx v8 RNA-Seq data19 (figure 3A and online supplementary table 1). Consistent with the association between the tissue specificity of gene expression and gene-level associations, we identified a significant enrichment of the identified lead and proxy variants within three transcription-activating histone marks (H3K4me1, H3K4me3 and H3K27ac) in the same tissues in a GREGOR30 enrichment analysis using Roadmap Epigenomic data at a false discovery rate (FDR) threshold of 5%.
Rheumatoid arthritis

These results are supported by the well-documented knowledge about the aetiological roles of immune tissues and non-immune tissues in RA (eg, the increased risk of RA by EBV infection to B cells, the early production of RA autoantibody in lungs, and the enhanced inflammation in small intestine in RA).

Table 1: RA associations of 19 loci in a meta-analysis

<table>
<thead>
<tr>
<th>Variant</th>
<th>Chr:Pos*</th>
<th>Gene†</th>
<th>Cohort #1</th>
<th>Cohort #2</th>
<th>Meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)‡</td>
<td>P value</td>
<td>OR (95% CI)‡</td>
<td>P value</td>
<td>OR (95% CI)‡</td>
</tr>
<tr>
<td>rs11464953</td>
<td>1:1745185</td>
<td>PADI4</td>
<td>1.23 (1.16–1.29)</td>
<td>1.5×10⁻¹³</td>
<td>1.25 (1.12–1.40)</td>
</tr>
<tr>
<td>rs148363003</td>
<td>1:160443973</td>
<td>SLAMF6</td>
<td>1.63 (1.39–1.92)</td>
<td>3.5×10⁻⁵</td>
<td>1.59 (1.12–2.27)</td>
</tr>
<tr>
<td>rs11889341</td>
<td>1:219943742</td>
<td>STAT4</td>
<td>1.21 (1.14–1.28)</td>
<td>1.3×10⁻¹¹</td>
<td>1.26 (1.12–1.41)</td>
</tr>
<tr>
<td>rs17605225</td>
<td>4:78508197</td>
<td>CXCL13</td>
<td>0.66 (0.57–0.76)</td>
<td>1.6×10⁻⁶</td>
<td>0.59 (0.40–0.87)</td>
</tr>
<tr>
<td>rs112062732</td>
<td>6:32543936</td>
<td>HLA-DRB1</td>
<td>2.38 (2.25–2.52)</td>
<td>1.1×10⁻¹⁹</td>
<td>2.79 (2.45–3.19)</td>
</tr>
<tr>
<td>rs1044690</td>
<td>6:44245241</td>
<td>NFKBIE</td>
<td>1.11 (1.22–1.37)</td>
<td>1.4×10⁻¹⁰</td>
<td>1.20 (1.06–1.37)</td>
</tr>
<tr>
<td>rs9494893</td>
<td>6:138223490</td>
<td>SLAMF6</td>
<td>1.40 (1.27–1.55)</td>
<td>4.7×10⁻¹¹</td>
<td>1.15 (0.92–1.45)</td>
</tr>
<tr>
<td>rs3093019</td>
<td>6:167539655</td>
<td>CCR6</td>
<td>0.83 (0.78–0.87)</td>
<td>2.3×10⁻¹²</td>
<td>0.84 (0.75–0.94)</td>
</tr>
<tr>
<td>rs113066392</td>
<td>7:74026152</td>
<td>GT2R1D1</td>
<td>1.46 (1.35–1.57)</td>
<td>6.3×10⁻²²</td>
<td>1.22 (0.95–1.57)</td>
</tr>
<tr>
<td>rs71508903</td>
<td>10:63779871</td>
<td>ARID5B</td>
<td>1.27 (1.19–1.36)</td>
<td>2.9×10⁻¹³</td>
<td>1.12 (0.98–1.28)</td>
</tr>
<tr>
<td>rs360136</td>
<td>11:9773517</td>
<td>SWAP70</td>
<td>1.14 (1.08–1.20)</td>
<td>1.4×10⁻⁶</td>
<td>1.31 (1.18–1.46)</td>
</tr>
<tr>
<td>rs78994077</td>
<td>12:111856673</td>
<td>SH2B3</td>
<td>1.30 (1.19–1.43)</td>
<td>3.6×10⁻⁸</td>
<td>1.42 (1.20–1.45)</td>
</tr>
<tr>
<td>rs111597524</td>
<td>14:35861800</td>
<td>NFKBIA</td>
<td>1.18 (1.12–1.26)</td>
<td>1.3×10⁻⁸</td>
<td>1.15 (1.01–1.31)</td>
</tr>
<tr>
<td>rs194757</td>
<td>14:69213851</td>
<td>ZFP36L1</td>
<td>0.87 (0.83–0.92)</td>
<td>2.9×10⁻⁶</td>
<td>0.81 (0.72–0.91)</td>
</tr>
<tr>
<td>rs2841269</td>
<td>14:105386149</td>
<td>PLD4</td>
<td>0.80 (0.76–0.85)</td>
<td>5.9×10⁻¹⁵</td>
<td>0.82 (0.72–0.92)</td>
</tr>
<tr>
<td>rs1032939</td>
<td>15:38834033</td>
<td>RASGRP1</td>
<td>0.84 (0.80–0.89)</td>
<td>1.6×10⁻¹⁰</td>
<td>0.93 (0.83–1.04)</td>
</tr>
<tr>
<td>rs1547233</td>
<td>21:26769507</td>
<td>LINC0015B</td>
<td>1.17 (1.10–1.24)</td>
<td>8.4×10⁻⁸</td>
<td>1.18 (1.05–1.33)</td>
</tr>
<tr>
<td>rs2049985</td>
<td>22:39756650</td>
<td>SYNGR1</td>
<td>0.78 (0.72–0.85)</td>
<td>1.2×10⁻⁷</td>
<td>0.89 (0.75–1.05)</td>
</tr>
<tr>
<td>rs1059702</td>
<td>X:153284192</td>
<td>IRAK1</td>
<td>0.86 (0.81–0.91)</td>
<td>8.2×10⁻⁷</td>
<td>0.74 (0.65–0.84)</td>
</tr>
</tbody>
</table>

*Based on hg19.
†Novel RA loci are indicated in bold.
‡Frequencies, ORs and CIs for EAs.
§P values for cross-cohort association heterogeneity (pHet) were calculated by Cochrane’s Q tests.
Chr, chromosome; EAs, effect alleles; Freq, frequency; NEA, non-effect allele; Pos, Position; RA, rheumatoid arthritis.

Figure 2: Regional association plots for novel RA-associated loci. Regional association plots for the six new RA loci: (A) SLAMF6, (B) CXCL13, (C) SWAP70, (D) NFKBIA, (E) ZFP36L1 and (F) LINC0015B, represent the minus log₁₀-transformed p values for variants in each locus according to their genomic positions. The lead variants are indicated by purple points and the other variants are coloured according to r² values with the lead variant in each locus. The names of lead variants and their closest genes are given on the top of the plots. RA, rheumatoid arthritis.
In addition, a MAGMA gene set analysis was performed using the gene-level Z scores and the predefined gene sets in MSigDB. The results suggested important roles of several immune-related gene sets such as the CD40 pathway (p=1.62×10^{-14}), citrullination (p=8.60×10^{-15}), IL-21-mediated signalling pathway (p=3.66×10^{-8}) and lymphocyte activation (p=4.53×10^{-8}) in developing RA (online supplementary table 2).

**Prioritisation of RA-relevant genes in RA loci**

Using the genome-wide association results, we further identified 122 potentially RA-relevant genes within 18 non-MHC RA loci by mapping the genes positioned on, cis-regulated by, and chromatin-interacted with RA-risk variants in blood, spleen, lung and small intestine (online supplementary table 3). Of the novel loci, the RA-risk variant in SLAMF6 (rs148363003) showed multiple chromatin interactions with several genes including five signalling lymphocytic activation molecule family (SLAMF) genes: SLAMF1, CD48 (also known as SLAMF2), CD84 (also known as SLAMF5), LY9 (also known as SLAMF3) and SLAMF7 (figure 3C). Interestingly, the expression levels of SLAMF6 and other interacting SLAMF genes were tissue-specifically over-expressed mostly in spleen, EBV-transformed B lymphocytes, whole blood, lung and small intestine terminal ileum. Similar tissue-specific expression patterns were observed in other well-known RA genes including PADI4, CCR6 and NCF1 (figure 3D).

**Genetic correlation with other diseases**

To assess the shared genetic risk between RA and other complex diseases using GWAS summary statistics, we calculated the genetic correlation of genome-wide disease associations between RA (in our Korean cohorts) and other diseases (27 diseases in Japanese populations^30^ and systemic lupus erythematosus in a European population^31^). The genetic architecture of RA was positively correlated with that of RA (in Japanese populations, Graves' disease, systemic lupus erythematosus and atrial fibrillation, while a negative correlation was found with chronic hepatitis B infection and gastric cancer (online supplementary figure 2). The degrees of genetic correlations with RA were high in autoimmune diseases, systemic lupus erythematosus and atrial fibrillation, while a negative correlation was found with chronic hepatitis B infection and gastric cancer (online supplementary figure 2). The degrees of genetic correlations with RA were high in autoimmune diseases, systemic lupus erythematosus and atrial fibrillation, while a negative correlation was found with chronic hepatitis B infection and gastric cancer (online supplementary figure 2).
diseases such as Graves’ disease (r=0.39) and systemic lupus erythematosus (r=0.33) (online supplementary figure 2).

Consistently, the prioritised RA-relevant genes in non-MHC RA loci in this study were significantly enriched in the susceptibility loci of inflammatory diseases such as systemic lupus erythematosus and inflammatory bowel disease at an FDR threshold of 0.05 (p≤2.43×10⁻⁶) in hypergeometric tests, according to GWAS catalog (online supplementary figure 3).

**DISCUSSION**

In this study, we employed a new genome-wide array, KoreanChip, for the majority of the study subjects. The customised array KoreanChip has an extensive genomic coverage and powerful imputation performance for low-frequency variants in Korean ancestry compared with other commercial GWAS arrays with a similar number of probes. A high-density variant association result and a meta-analysis with another independent cohort helped us perform a powerful analysis to assess the genomic boundaries, the reliability (or cross-cohort heterogeneity) of disease association signals and functional annotations of RA loci.

The meta-analysis of GWAS data in Korean cohorts identified 19 susceptibility loci for RA, of which 6 loci have not been reported before, although there was considerably large missing heritability and a lack of functional studies. Of these new loci for RA, previous functional studies on the products of SLAMF6, CXCL13 and SWAP70 have been associated with RA or other inflammatory diseases. SLAMF6 is a member of superfamly immunoglobulin (Ig) domain-containing molecules that are involved in haemophilic interactions between naive B and T cells. SLAMF6 acts as a costimulatory molecule on the surface of T lymphocytes to increase T-cell adhesiveness by activating the small GTPase Rap1, promoting the TCR downstream signal transduction and the differentiation to Th17 cells expressing IL-17A. SLAMF6 members including SLAMF6 have been proposed as potential therapeutic targets for inflammatory and autoimmune diseases. In this study, RA-risk variants in SLAMF6 including rs148363003 had chromatin interaction with five SLAMF genes and the expression level of these six SLAMF genes in the locus was high in most of the potentially RA-relevant tissues. In other studies on rheumatic diseases, a SLAMF6 mutation (p.Phe238Cys) was found in T cells in RA and the increase in disease activity index in patients with systemic lupus erythematosus was associated with the enhanced levels of SLAMF3 and SLAMF6 on the surface of T lymphocytes and the high IL-17 level in serum.

C-X-C motif chemokine ligand 13 (CXCL13), a B-cell chemotactic receptor, binds to its receptor CXCR5 to promote B-cell migration. The RA-risk variant (rs17605225) was located in a highly conserved region in the intron of CXCL13 according to two conservation scores (called GERP and SiPhy). The RA-risk CXCL13 variant showed consistent effect sizes in the two independent cohorts and had no proxy variants based on the LD indices in the Korean cohorts and the 1KGP populations. Circulating CXCL13 levels had positive correlations with joint destruction rates and disease activity scores. Similarly, a transcriptome analysis displayed the enriched CXCL13 transcript levels in T cells in synovial fluid of patients with RA.

SWAP70 is a Rho GTPase-regulatory protein that controls both cytoskeletal dynamics and the induction of interferon regulatory factor 4 (IRF4). SWAP70 is necessary for normal B-cell migration and immobilises F-actin filaments on phagosomes in human dendritic cells. Although a further investigation would be required to pinpoint and confirm SWAP70 as an RA-driving gene in the SWAP70 locus, an experimental murine model showed that genetic defects of SWAP70 resulted in lupus-like symptoms and the accumulation of T<sub>reg</sub> cells.

This study emphasised the crucial role of the immune and non-immune tissues (lung and small intestine) for the development of RA by analysing the tissue-specific expression of RA-relevant genes, especially SLAMF member genes in the SLAMF6 locus and several known RA-associated genes including PADI4, CCR6 and NCF1. Among the RA-relevant tissues, lung has long been suggested as an early site of inflammation and ACpA in developing RA in an interaction of genetic factors (eg, HLA-DRB1, PADI4, CCR6) with cigarette smoking and air pollution. Our previous bioinformatic study also identified the significant enrichment of RA-risk variants within lung-specific enhancer regions, indicating that non-coding RA-variants may regulate gene expression in allele-specific manner in lung.

Small intestine has large surfaces that are exposed to the gut environment and affects the human immune response with intestinal immune cells. As IgA-producing plasma cells of gut-associated lymphoid tissues are rich in lamina propria, the immune response by ACpA and rheumatoid factor of I gA isotype can be initiated in the intestinal mucosal tissue in RA. Interestingly, IL-21, highlighted by our gene set analysis, is primarily produced in the small intestine and is important in IgA responses to gut atypical commensals and T helper cell responses in intestinal inflammation in lamina propria in mice. In gut microbiota-induced arthritis in a murine model, T follicular helper cells were induced in lamina propria of gut before the onset of experimental arthritis and migrated to systemic lymphoid tissues, resulting in autoantibody production and arthritis.

EBV infection has been associated with an increased risk of RA and other inflammatory diseases. An in silico study by Harley et al analysed the binding sites of EBV-encoded EBNA2 proteins on the human genome in EBV-transformed B cells. The authors identified the significant enrichment of RA variants within the EBNA2-binding elements, which allowed the expression of RA-relevant genes to be highly allele-specific.

Functional studies on novel RA variants are needed in the future. Our bioinformatics approaches provide a list of 122 potentially RA-relevant genes in RA loci that may facilitate follow-up functional studies of candidate genes, patient stratification and drug target validation. The gene set analysis highlighted several immune-related pathways such as the CD40 signalling and IL-21-mediated pathways that had showed altered activities in patients with RA.

In summary, this study discovered six novel RA loci in Korean populations and demonstrated variant-driven pathological features, evoking the importance of the identification of new RA loci and the integration of disease association results with other omics data.

**Author affiliations**

1. Hanyang University Institute for Rheumatology Research, Seoul, Republic of Korea
2. Department of Biology and Department of Life and Nanopharmaceutical Sciences, Kyung Hee University, Seoul, Republic of Korea
3. Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of Korea
4. Division of Genome Research, Center for Genome Science, National Institute of Health, Osong Health Technology Administration Complex, Cheongju, Republic of Korea
5. Department of Rheumatology, Catholic University of Daegu School of Medicine, Daegu, Republic of Korea
6. Division of Rheumatology, Department of Internal Medicine, Chonnam National University Medical School and Hospital, Gwangju, Republic of Korea
7. Division of Rheumatology, Ewha Womans University School of Medicine, Seoul, Republic of Korea

### References


