OPEN ACCESS

TRANSLATIONAL SCIENCE

Large-scale meta-analysis across East Asian and European populations updated genetic architecture and variant-driven biology of rheumatoid arthritis, identifying 11 novel susceptibility loci

Eunji Ha ![ORCID],[1] Sang-Cheol Bae ![ORCID],[2,3] Kwangwoo Kim ![ORCID] [1]

![Check for updates]

ABSTRACT

Objectives Nearly 110 susceptibility loci for rheumatoid arthritis (RA) with modest effect sizes have been identified by population-based genetic association studies, suggesting a large number of undiscovered variants behind a highly polygenic genetic architecture of RA. Here, we performed the largest-ever trans-ancestral meta-analysis with the aim to identify new RA loci and to better understand RA biology underlying genetic associations.

Methods Genome-wide RA association summary statistics in three large case–control collections consisting of 311 292 individuals of Korean, Japanese and European populations were used in an inverse-variance-weighted fixed-effects meta-analysis. Several computational analyses using public omics resources were conducted to prioritise causal variants and genes, RA variant-implicating features (tissues, pathways and transcription factors) and potentially repurposable drugs for RA treatment.

Results We identified 11 new RA susceptibility loci that explained 6.9% and 1.8% of the single-nucleotide polymorphism-based heritability in East Asians and Europeans, respectively, and confirmed 71 known non-human leukocyte antigens (HLA) susceptibility loci, identifying 90 independent association signals. The RA variants were preferentially located in binding sites of various transcription factors and in cell type-specific transcription–activation histone marks that simultaneously highlighted the importance of CD4+ T-cell activation and the potential role of non-immune organs in RA pathogenesis. A total of 615 plausible effector genes, based on gene-based associations, expression-associated variants and chromatin interaction, included targets of drugs approved for RA treatments and potentially repurposable drugs approved for other indications.

Conclusion Our findings provide useful insights regarding RA genetic aetiology and variant-driven RA pathogenesis.

INTRODUCTION

Rheumatoid arthritis (RA) is a common autoimmune disease characterised mainly by pain, swelling and deformity of joints due to prolonged inflammation.[1] A substantial portion of patients with RA are seropositive for rheumatoid factor or anticitrullinated protein antibodies (ACPA).[2] RA is most common in women aged 30–50 years and affects

Key messages

What is already known about this subject?
► Genetic factors are strongly involved in the aetiology of rheumatoid arthritis (RA).
► Although approximately 110 loci were reported as RA susceptibility loci by genome-wide association studies, a large fraction of genetic heritability estimated in twin studies is still hidden behind many genetic variants with weak effect sizes.

What does this study add?
► We discovered 11 novel loci (*DGUOK-AS1*, *DAP*, *BAD*, *TPCN2*, *LOC107984408*, *LOC105369698*, *IQGAP1*, *PRKCB*, *ZNF689*, *C20orf181* and *SMC1B*) associated with RA surpassing the genome-wide significance level (p=5×10⁻⁸) in a large-scale meta-analysis integrating genetic associations in East Asians and Europeans.
► We catalogued possible causal variants and genes, RA-relevant biological features (tissues, pathways and transcription factors) and repurposable drugs for RA treatment from the genetic association results.

How might this impact on clinical practice or future developments?
► The genetic liability explained by these novel variants will improve precision medicine in RA and provide novel druggable targets for RA treatment.

approximately 1% of the general population.[3] Multifactorial causes, including genetic and environmental factors, are involved in the development of RA.[1] Twin concordance studies demonstrated that genetic predisposition plays a major role in the pathogenesis of RA, accounting for approximately 60% of the liability for RA.[4] In the past decade, genome-wide association studies (GWASs) in various populations identified nearly 110 RA susceptibility loci across the human genome.[5] However, the heritability explained by the identified variants was estimated to be ~20%,[5,6] suggesting that a much larger number of disease variants remain unidentified. Consistent with these results, Bayesian inference predicted hundreds of common susceptibility

variants with modest effect sizes and rare variants with large effect sizes.[7]

Here, we carried out a large-scale meta-analysis on RA with summary association data from East Asian and European cohorts comprising 22 628 RA cases and 288 664 controls and identified 11 new RA susceptibility loci. In addition, the integration of accumulated knowledge of RA variants with emerging high-throughput omics data facilitated various so-called post-GWAS approaches that helped unravel the biology of RA based on disease-risk variants in actual human patients, suggesting potentially repurposable drugs for RA treatment.

METHODS

Association summary statistics

RA association summary statistics were retrieved from previous GWASs in Korean, Japanese and European populations.[6 8 9] The sample size of each GWAS was considerable (4068 RA cases and 36 487 controls in the Korean population[8]; 4199 cases and 208 254 controls in the Japanese population[9]; 14 361 cases and 43 923 controls in the European population[6]), which brought the final sample size to more than 310 000. The summary statistics from Japanese and European populations were available at the Japanese Encyclopedia of Genetic associations by RIKEN (http://jenger.riken.jp/en/result). All data included over 13 million imputed variants with reliable imputation quality scores and minor allele frequency (MAF) of ≥0.5%. Each study performed a whole-genome imputation using the imputation reference panel from the 1000 Genomes Projects (1KGP) phase III data with or without ancestry-matched whole-genome sequencing data.

Genome-wide meta-analysis

An inverse-variance-weighted fixed-effects model with genomic control was used in a meta-analysis of the association summary statistics in three studies by meta-analysis of genome-wide association studies (METAL)[10] to calculate the disease effect sizes and SEs of query variants (n=13 810 676). Heterogeneity of effect sizes between East Asian and European populations was assessed using Cochran's Q test.[11] We defined an RA-associated locus by merging the neighbouring significant variants (p≤5×10⁻⁸), between which the minimum distance should not exceed 250 kb. Downstream analyses excluded the extended major histocompatibility complex (MHC) region (24–37 Mb in chromosome 6 in the human genome assembly GRCh37).

Estimation of RA heritability

Single-nucleotide polymorphism (SNP)-based heritability (h^2_{SNP}) of RA in East Asian (Korean+Japanese) or European populations was calculated in a liability scale using LD score regression (LDSC)[12] with precalculated linkage disequilibrium (LD) scores, regression weights and allele frequencies from the 1KGP phase III data in a relevant ancestral population (https://data.broadinstitute.org/alkesgroup/LDSCORE/), setting the disease prevalence parameter to 1%.

Correlation analysis of genome-wide SNP effects between ancestries

We employed Popcorn[13] to calculate the trans-ethnic genetic correlation of effect sizes of the genome-wide variants based on genetic effect scores via LD optimisation similar to LDSC.

Conditional association analysis in RA loci

To identify independent association signals in 82 non-MHC loci, a stepwise approximate conditional association analysis was performed for each population by genome-wide complex trait analysis (GCTA)[14] using ancestry-matched LD matrices. Genotype data of 5288 Korean and 2502 European individuals were used to estimate LD matrices. The Korean genotype/imputation data were previously described in our recent study,[8] and the European genotype data were obtained from European Genome-phenome Archive (EGAD00010000290).[15] We newly imputed untyped variants in the European individuals from quality-controlled variants (with missing rate≤0.05, MAF≥0.5%, and p value in Hardy-Weinberg equilibrium exact test≤1×10⁻⁷) in each RA locus using SHAPEIT2 and Minimac3 with the 1KGP imputation reference panel. The polymorphic variants with imputation quality score of ≥0.3 and MAF of ≥0.5% were used to calculate variant–variant correlations. A conditional analysis with adjustment of a primary association signal (the lead variant) in each RA locus was performed for each ancestral group and followed by a cross-ancestry meta-analysis to identify a secondary association signal with conditional p_{meta} of ≤5×10⁻⁸. This procedure conditioning on all independent association signals was repeated until no more signals were significant. When a lead variant was not available in one ancestry, we performed a stepwise conditional analysis in the other ancestry. All annotation of lead variants and their proxies (r²≥0.9) were retrieved from HaploReg[16] and RegulomeDB.[17]

Heritability-partitioning analysis using transcription factor-binding sites (TFBSs)

We calculated RA heritability attributed to SNPs within binding sites of 161 transcription factors (TFs) using stratified LDSC[18] to estimate relative heritability enrichment. All tested TFBSs were provided by an annotation library from PAINTOR (https://github.com/gkichaev/PAINTOR_V3.0/wiki/2b.-Overlapping-annotations). For each TF, heritability enrichment value is equal to the proportion of heritability (= $h^2_{TFBS\_SNP}$ / h^2_{SNP}) divided by the proportion of variants (= *number of SNPs in TFBS* / *number of total SNPs*). The enrichment scores and significance levels were estimated in East Asian and European populations separately using ancestry-matched LD scores, and the cross-ancestry significance level for each TF was calculated by Fisher's combined probability test.[19] A statistically significant threshold was set at a false discovery rate (FDR) of <5%.

Enrichment analysis of RA variants using histone modification marks

We assessed the statistical significance for the colocalisation of RA-associated lead variants (plus their proxies in r²≥0.8 with the lead variants in a relevant ancestry) with tissue-specific epigenomic regulatory elements in East Asians and Europeans separately using GREGOR[20] by comparing 1000 feature-matched control sets. The genomic locations of six histone marks (H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K36me3 and H3K27ac) of various cells or tissues were retrieved from the Roadmap Epigenomics Project data.[21] The number of lead variants in RA loci outside of the MHC regions was 38 in East Asians and 43 in Europeans. The feature-matched control sets were randomly selected based on the 1KGP phase III data to be matched for three properties of RA-associated lead variants (the number of LD proxies, MAF and distance to the most proximal gene). The enrichment p value was defined as a probability of detecting more than the observed number of overlaps between the query variant sets and a selected histone mark, based on the control distribution approximated using feature-matched control

sets by the saddle point method.[20] Fisher's combined probability test was used to calculate the combined p value.

Nomination of potential effector genes in RA loci

We catalogued potential effector genes (=disease-driving genes) from the meta-analysis association summary statistics using FUMA[22] and MAGMA[23] with LD information in all 1KGP individuals. First, FUMA analysis mapped candidate genes using the SNP2GENE module with default parameters when RA-risk variants were located within 10 kb of the genes, known to have expression quantitative trait loci (eQTLs) of the genes or interacted with the genes via chromatin interaction. Briefly, we retrieved reported eQTLs in selected tissues (blood cells, spleen, small intestine and lung) from the EBI eQTL catalogue (https://www.ebi.ac.uk/eqtl), single-cell RNA eQTLs,[24] DICE,[25] GTEx[26] and other blood eQTL sets.[27–30] Chromatin interaction mapping was performed using publicly available Hi-C data[31] in the same four tissues and FANTOM5 enhancers and promoters.[32] To retain more likely candidate effector genes, physically mapped genes were narrowed down at the Bonferroni-corrected gene-based p value cut-off of 0.05 in MAGMA.

Prediction of repurposable drugs for RA

We investigated potential drug candidates for RA treatment using Genome for REPositioning drugs (GREP).[33] Proteins directly interacting with the products of 615 potential RA effector genes were obtained from HumanNet V.2 resources.[34] The genes of these interacting proteins were used as query genes, along with the 615 RA genes in the GREP analysis to search for potential repurposable drugs that target effector proteins or their network members. Anatomical therapeutic chemical (ATC) classes[35] (n=85) were tested for the enrichment of the identified drugs. The most likely repurposable drugs were defined as belonging to any ATC groups with FDR-corrected enrichment p value of ≤0.05.

RESULTS

### Identification of 11 novel susceptibility loci for RA

A genome-wide meta-analysis of RA associations in 22 628 patients with RA and 288 664 healthy controls in East Asian (8267 cases and 244 741 controls) and European (14 361 cases and 43 923 controls) populations detected 11 novel loci associated with RA risk at the genome-wide significance level ($4.6 \times 10^{-8} \leq p_{meta} \leq 1.5 \times 10^{-10}$), replicating the known RA associations in 71 non-MHC loci[8 36 37] (figure 1 and online supplemental table S1). The meta-analysis association summary statistics in this study are available online (https://doi.org/10.5061/dryad.ns1rn8pr0). Excluding MHC variants, we found that the meta-analysis results displayed an inflation factor λ of 1.01, indicating little systemic bias in our study (online supplemental figure S1). The most significant variants in the newly identified 11 loci (with the following nearest genes: *DGUOK-AS1*, *DAP*, *BAD*, *TPCN2*, *LOC107984408*, *LOC105369698*, *IQGAP1*, *PRKCB*, *ZNF689*, *C20orf181* and *SMC1B*) showed modest effect sizes on RA development (OR for a risk allele ≤1.13, figure 2, table 1). The observed effect sizes were highly consistent between Korean and Japanese populations (heterogeneity p value≥0.38, online supplemental table S2) and between East Asians and Europeans (heterogeneity p value≥0.67, table 1) for all the tested lead variants, although four lead variants were not tested in Europeans. Of the 11 lead variants, four were insertion/deletion variants. The lead variant in *DGUOK-AS1* was not common in Europeans
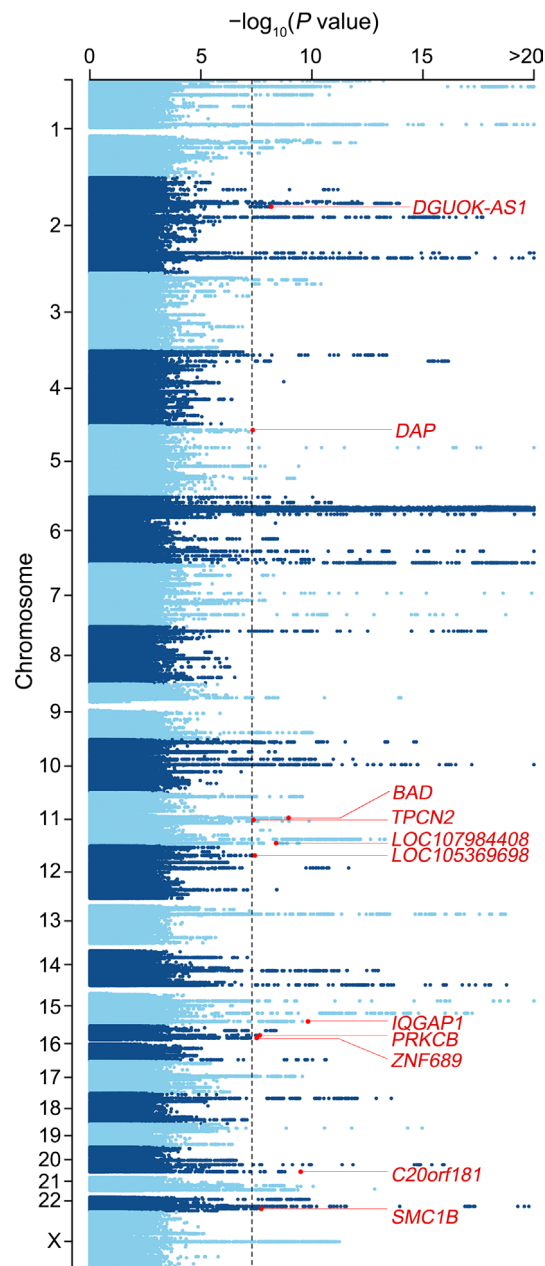
Figure 1 Manhattan plot of trans-ancestral genome-wide association meta-analysis results for RA. The y-axis represents negative natural logarithms of the p value for each variant in a meta-analysis using an inverse-variance-weighted fixed-effects model for the genomic-control associations in Korean, Japanese and European populations. The x-axis indicates chromosomal positions. The dashed line represents the genome-wide significance threshold ($p=5 \times 10^{-8}$). Eleven novel loci are marked in red with the nearest genes. RA, rheumatoid arthritis.

(MAF=0.7%), yielding an insignificant association p value but a consistent effect size (table 1).

Genetic liability explained by the tested genome-wide variants was estimated as 0.176 in East Asians and 0.275 in Europeans in an LDSC heritability analysis[12] (table 2). The variants in 11 novel loci accounted for 6.9% and 1.8% of the non-MHC SNP-based heritability in East Asians and Europeans, respectively. Although the estimated SNP-based heritability was higher in Europeans than in East Asians, we observed a highly strong association correlation between the two ancestries in whole-genome
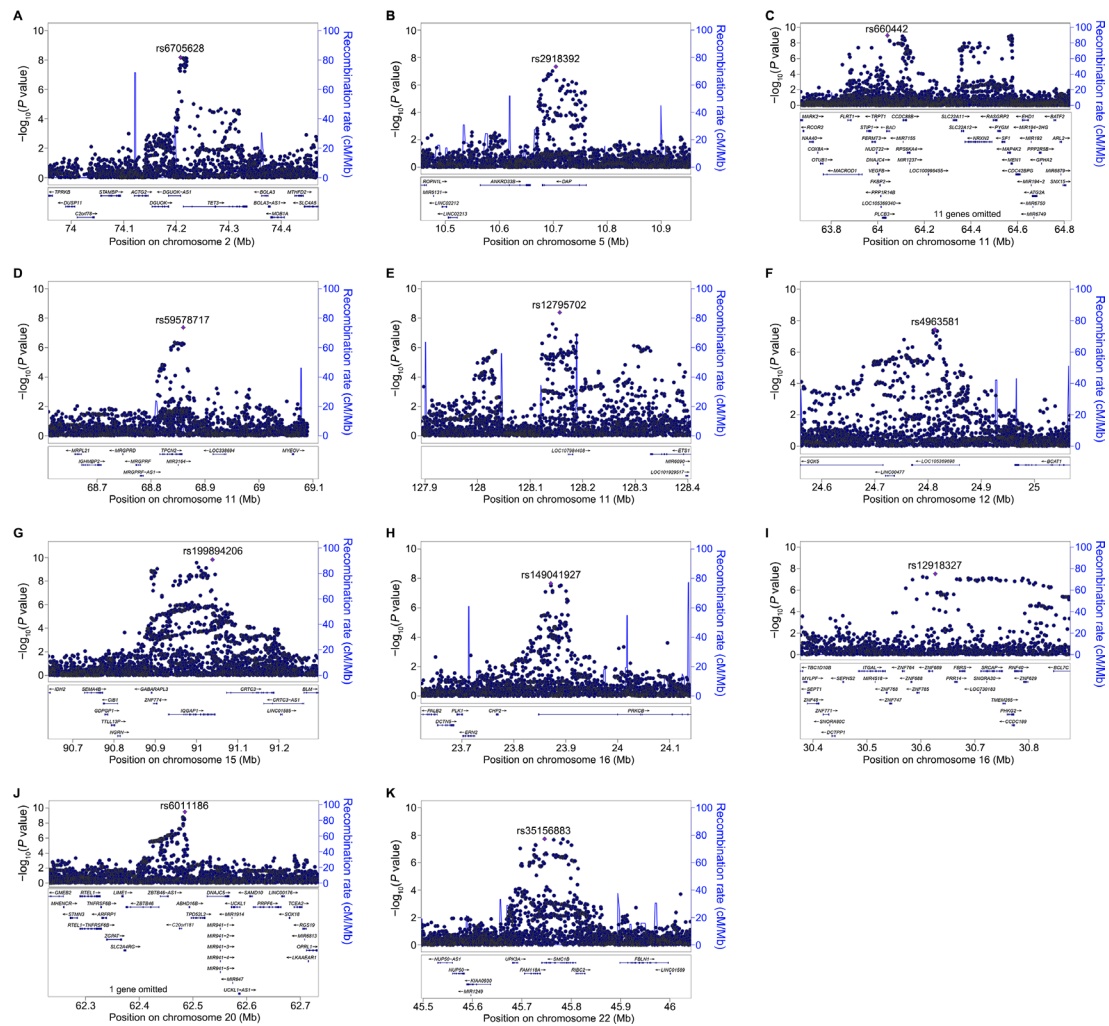
Figure 2 Regional association plots for the newly identified RA loci. The association significance levels in the loci of interest were plotted in negative logarithm scale according the chromosomal position of variants. The most significant variants are denoted as purple diamonds. (A) *DGUOK-AS1*, (B) *DAP*, (C) BAD, (D) *TPCN2*, (E) *LOC107984408*, (F) *LOC105369698*, (G) *IQGAP1*, (H) *PRKCB*, (I) *ZNF689*, (J) *C20orf181* and (K) *SMC1B*. RA, rheumatoid arthritis.

scale (r=0.69±0.09, p value=2.12×10^{-14}), which supports a high degree of risk allele sharing between the two populations.

To narrow down the potentially functional variants, we found 130 proxy variants in high LD ($r^2 \geq 0.9$) with lead variants in both East Asian and European populations (for eight loci tested in both populations) or in East Asians (for four loci tested in only East Asians). Among the 141 potentially causal variants, two (rs11556482 and rs6007594) were missense variants in *FAM118A* (near *SMC1B*). In addition, 29 variants in six loci were likely to affect TF binding or linked to gene expression in an allele-specific manner, being located with highly functional annotations (RegulomeDB[17] category=1 or 2, online supplemental table S3).

Dissecting association signals
To determine the number and sources of association signals in 82 non-MHC loci, we performed a stepwise approximate conditional association analysis for each ancestral group followed by a meta-analysis of conditional association results in two groups. There were at least two independent association signals in each of seven loci (*PADI4*, *CTLA4*, *TNFAIP3*, *IL2RA*, *PRKCQ*, *ARID5B* and *LOC145837*) with a conditional p value$\leq 5 \times 10^{-8}$ (online supplemental table S4).

Enrichment of RA variants on TFBSs and tissue-specific epigenetic features
The degree of enrichment of RA heritability on binding sites of 161 TFs was assessed using the population-specific LDSC followed by Fisher's combined probability tests. We observed that RA heritability was significantly enriched in variants within binding sites of 29 TFs (p value for heritability enrichment ≤ 0.05 in both populations and FDR-corrected $p_{meta} \leq 0.05$, figure 3A and online supplemental table S5). Among the identified TFs, 12 displayed extremely large heritability enrichment in their TFBSs (enrichment>40 in both populations), and these TFs have been significantly associated with T-cell receptor (TCR) signalling transduction mediated by mitogen-activated protein kinases, nuclear factor-kappa B and nuclear factor of activated T-cells[38] (online supplemental table S6). These results reinforce the importance of CD4$^+$ T-cell activation in RA pathogenesis,[5] suggesting that heritability-explaining RA variants may play an allele-specific transcription-regulatory role in CD4$^+$ T cells, especially on activation.

The regulatory effects of RA-risk variants even in relevant TFBSs highly depend on chromatin accessibility associated with highly cell type-specific histone modification marks. Given this knowledge, we searched for RA-relevant tissues, in which histone

Table 1 RA associations of the newly identified loci in a trans-ancestral genome-wide association meta-analysis

Variant	Chr:position*	Nearest gene	EA/NEA	East Asian Freq_ASN	OR_KOR	OR_JAP	OR_ASN (95% CI)	P_ASN	European Freq_EUR	OR_EUR (95% CI)	P_EUR	Meta-analysis OR_META (95% CI)	P_META	p_{HET}
rs6705628	2:74208362	DGUOK-AS1	T/C	0.18	0.87	0.89	0.88 (0.84 to 0.92)	2.1×10^{-8}	0.007	0.90 (0.79 to 1.03)	0.12	0.88 (0.85 to 0.92)	6.7×10^{-9}	0.95
rs2918392	5:10704797	DAP	T/C	0.29	0.95	0.95	0.95 (0.91 to 0.98)	3.9×10^{-3}	0.32	0.93 (0.90 to 0.96)	1.6×10^{-6}	0.94 (0.91 to 0.96)	4.6×10^{-8}	0.89
rs660442	11:6404297	BAD	A/G	0.08	0.82	0.86	0.84 (0.79 to 0.89)	3.9×10^{-8}	0.10	0.93 (0.89 to 0.96)	2.0×10^{-4}	0.90 (0.87 to 0.93)	1.1×10^{-9}	0.67
rs59578717	11:68859848	TPCN2	Ins1†/C	0.51	0.90	0.92	0.91 (0.88 to 0.94)	4.3×10^{-8}	0.86‡	NA	NA	0.91 (0.88 to 0.94)	4.3×10^{-8}	NA
rs12795702	11:128156314	LOC107984408	A/G	0.77	1.14	1.07	1.11 (1.06 to 1.15)	5.8×10^{-7}	0.18	1.07 (1.03 to 1.11)	8.5×10^{-4}	1.09 (1.06 to 1.12)	4.1×10^{-9}	0.89
rs4963581	12:24813281	LOC105369698	A/G	0.75	1.09	1.09	1.09 (1.05 to 1.13)	6.6×10^{-6}	0.09	1.09 (1.03 to 1.14)	1.4×10^{-3}	1.09 (1.06 to 1.12)	3.8×10^{-8}	0.98
rs199894206	15:91038883	IQGAP1	CT/C	0.36	1.12	1.12	1.12 (1.08 to 1.16)	1.5×10^{-10}	0.19‡	NA	NA	1.12 (1.08 to 1.16)	1.5×10^{-10}	NA
rs1490041927§	16:23871206	PRKCB	A/Ins2¶	0.38	0.90	0.91	0.91 (0.88 to 0.94)	2.2×10^{-8}	0.44‡	NA	NA	0.91 (0.88 to 0.94)	2.2×10^{-8}	NA
rs12918327	16:30626616	ZNF689	T/C	0.10	1.06	1.12	1.09 (1.03 to 1.16)	1.6×10^{-3}	0.12	1.09 (1.05 to 1.13)	3.9×10^{-6}	1.09 (1.06 to 1.12)	3.0×10^{-8}	0.98
rs6011186	20:62484008	C20orf181	T/C	0.38	0.87	0.92	0.90 (0.87 to 0.93)	3.7×10^{-9}	0.035	0.90 (0.82 to 0.99)	2.6×10^{-2}	0.90 (0.87 to 0.93)	3.2×10^{-10}	0.99
rs35156883	22:45746152	SMC1B	A/AT	0.52	1.12	1.08	1.10 (1.06 to 1.13)	1.9×10^{-8}	0.63	NA	NA	1.10 (1.06 to 1.13)	1.9×10^{-8}	NA

*Based on hg19.
†CGGCTAAGGGAGGAGAGCCCGGAAGCCACCGCACCGGCCAAGGGGCGGATA.
‡Frequency data were retrieved from European statistics in the 1000 Genomes Project phase III.
§or rs59279623.
¶AAGGGAG.

Chr, chromosome; EA, effect allele; EUR, European; Freq, frequency of EA; JAP, Japanese; KOR, Korean; NEA, non-effect allele; P_{HET}, p value from a Cochrane's Q test for heterogeneity in effect sizes; RA, rheumatoid arthritis.

Table 2 Liability-scale RA heritability h^2 estimated from the tested genome-wide variants in each ancestry

Ancestry	All tested variants h^2	SE of h^2	Non-MHC variants* h^2	SE of h^2	Proportion of h^2 explained by Known RA loci	Novel RA loci
East Asian	0.176	0.053	0.123	0.012	43.0%	6.9%
European	0. 275	0.096	0.185	0.021	38.1%	1.8%

*The MHC region was defined as spanning as the 24–37 Mb region of chromosome 6 in hg19.

MHC, major histocompatibility complex; RA, rheumatoid arthritis.

marks colocalise with significantly more RA-risk variants. We employed the GERGOR algorithm[20] to test the enrichment on four transcription-activating histone modifications (H3K4me1, H3K27ac, H3K4me3 and H3K36me3) and two repressing histone modifications (H3K27me3 and H3K9me3) in diverse human cell types. Transcription-activating histone marks were strongly associated with RA-risk variants in various immune cells, especially in CD4$^+$ T-cell subtypes (figure 3B). Among the CD4$^+$ T cells, memory CD4$^+$ T cells (E037 and E040) rather than naïve CD4$^+$ T cells (E038 and E039) presented relatively strong significance levels for the RA-variant enrichment. Furthermore, chromatin changes on T-cell activation and T$_{reg}$ differentiation were strongly associated with RA variants (figure 3B). In addition, this analysis replicated our recent findings on the involvement of two non-immune organs,[8] lung and small intestine, in disease pathogenesis (figure 3B and online supplemental table S7).

Candidates for repurposable drugs targeting RA genes

We narrowed down the potential effector genes to 615 genes based on three categories: gene-level association significance levels (estimated from genome-wide variant associations, gene-level p value≤0.05/19 644), known eQTL effects, and chromatin interactions between RA-variants and neighbouring genes (online supplemental table S8). A total of 132 genes belonged to more than two categories. For example, DAP in a novel locus encodes a member of mTOR signal transduction[39] and was identified as a new plausible effector for RA, as the gene-based association of DAP with RA was significant (p$_{MAGMA}$=1.62×10^{-6}) and a lead variant (rs2918392) was a known eQTL for DAP in blood cells.[25 26 29]

We further investigated potentially repurposable drugs for RA that target the 615 effector gene products and their 1543 direct interactors (=2158 RA-relevant genes). We found that the tested genes were significantly enriched in the targets of immunosuppressants, immunostimulants and antineoplastic agents in Fisher's exact tests (FDR-corrected p values≤3.34×10^{-4}). For example, 18 RA-relevant genes are targeted by 21 immunosuppressants, including known RA drugs (eg, abatacept, tocilizumab, tofacitinib, etanercept, sarilumab, baricitinib, infliximab, adalimumab, certolizumab pegol, golimumab and azathioprine) and potential repurposable drugs previously approved for other indications, such as systemic lupus erythematosus and multiple sclerosis (eg, eculizumab, alefacept, belatacept, daclizumab, siltuximab, mycophenolic acid, lenalidomide, basiliximab and pomalidomide; figure 4 and online supplemental table S9).

DISCUSSION

This study had the advantage of analysing two distinct ancestral populations. As two ancestral populations have highly different
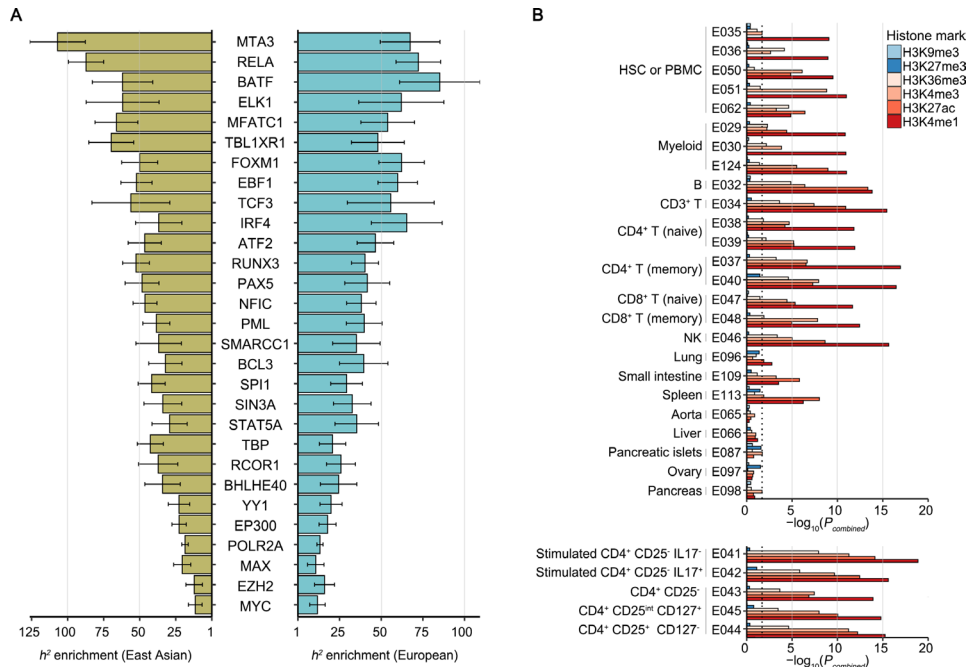
Figure 3 Transcription factors and cell types implicated by genetic associations in RA. (A) RA heritability was significantly enriched in binding sites of 29 TFs with p<0.05 in both populations and FDR-corrected p_{meta} <0.05. Erichment estimates in East Asians (left) and Europeans (right) were plotted for each TF (y-axis). (B) Enrichment of RA-risk variants on roadmap histone modification marks in diverse immune and non-immune cell types are assessed in each population. The combined p values for the enrichment were plotted in negative logarithm scale according to 25 selected cell types and 5 additional CD4$^+$ subtypes, with roadmap epigenome codes and cell-type names. FDR, false discovery rate; RA, rheumatoid arthritis; TF, transcription factor.

LD architectures, most analyses in this study (eg, association meta-analysis, conditional association analysis, TFBS heritability analysis and cell-type histone mark enrichment analysis) were performed for each population, separately, and the independent results were merged. Results in each population were mutually validated by the results in the other population, which provided population-shared, reliable insights into RA aetiology and pathogenesis.

This study increased the explained proportion of genetic liability for RA, especially more in East Asians. We detected 82

non-MHC RA-risk loci, including 11 novel loci, and identified 90 distinct signals by combining conditional results for both ancestry groups. The gene-level approaches nominated 615 plausible effector genes from the 82 RA loci based on genic variant association, eQTL and chromatin interaction data. In addition, we provided complete lists of the most likely causal variants in novel loci based on LD in both populations and regulatory annotation/statistics. The catalogues of the most likely RA-risk variants and genes may be useful in choosing targets for experimental validation to deepen understanding of variant-driven
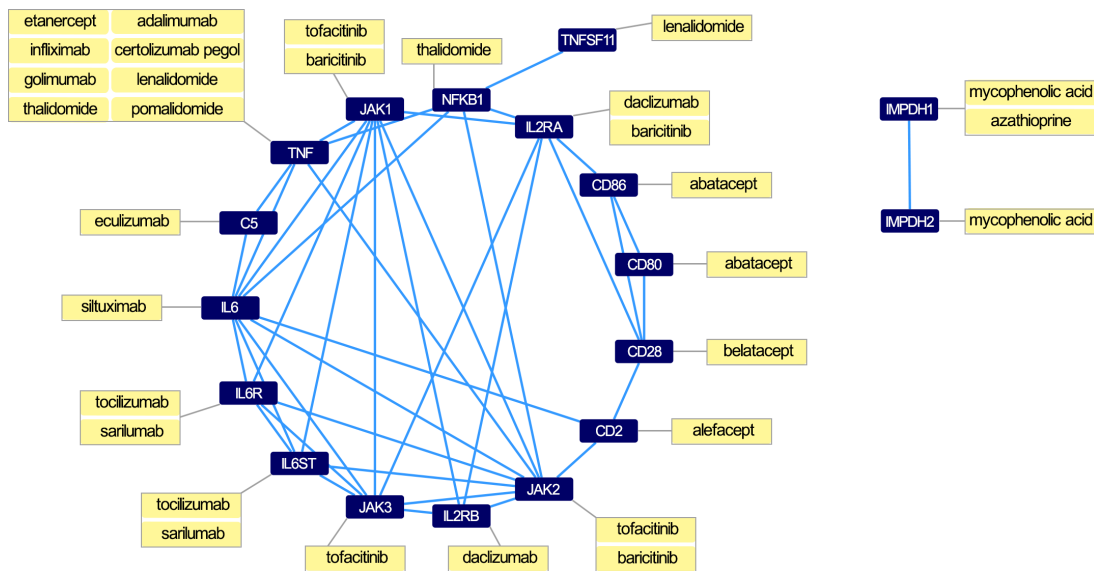
Figure 4 Approved or potentially repurposable immunosuppressants for RA treatment. Existing immunosuppressants (shown in yellow boxes) are connected with drug targets (shown in dark blue boxes) that were extracted from potential RA effector gene products and their interaction partners. A network of the drug targets was retrieved from HumanNet v2-XN and visualised by Cytoscape. RA, rheumatoid arthritis.

pathological processes and to identify druggable targets. Indeed, we showed that the RA effector genes and their interaction partners were actual targets of known RA drugs and suggested other drugs that may be repurposable for RA treatment.

CD4$^+$ T-cell biology was emphasised in RA pathogenesis by large-heritability variants that preferentially spanned binding sites of various TFs related with TCR-mediated signalling and that were preferentially located with transcription-activating histone marks in CD4$^+$ T cells, including stimulated, memory and/or regulatory T cells.

In the novel loci, 88 genes were detected as potential effector genes (8 genes per locus). The number of nominated genes are quite large but likely include true RA genes. It is possible to further narrow them down based on the number of categories (gene-based p value, eQTL and chromatin interaction) to which they are assigned (online supplemental table S8). For example, the number of genes with more than two categorical hits is decreased to 16 in 11 novel loci (*DAP*, *CCDC88B*, *RPS6KA4*, *NRXN2*, *MEN1*, *ZNF774*, *IQGAP1*, *CRTC3*, *PRKCB*, *AC002310.12*, *PRR14*, *FBRS*, *SRCAP*, *BCL7C*, *FAM118A* and *SMC1B*). Some have been documented for their functional relationship with immune cells or immune disorders (*DAP*,[40] *CCDC88B*,[41] *RPS6KA4*,[42] *IQGAP1*,[43] *CRTC3*[44] and *PRKCB*[45]). For example, *IQGAP1*, encoding a controller of tumour nectosis factor costimulatory receptor CD134,[43] modulates immune responses (eg, T-cell cosignalling pathway) possibly by an allele-specific regulatory effect of an RA-risk eQTL mediated by chromatin interaction in relevant tissues (online supplemental table S8).

CCDC88B is known to be essential for T-cell maturation and activation.[41] Variants in *CCDC88B* have been associated with the risk of inflammatory bowel diseases,[46] possibly leading to CD4$^+$ T-cell-induced colitis.[47] The lead variant rs660442 in our study was demonstrated to regulate expression of *CCDC88B* in immune cells (online supplemental table S3).

Another variant rs3826259 in *PRKCB*, an LD proxy (r^2=0.98) of a lead variant, is located in a highly conserved genomic element and is likely to influence binding affinity of several TFs according to the RegulomeDB.[17] The regulatory effect of the variant on *PRKCB* expression in RA-relevant cells is supported by eQTL catalogues (online supplemental table S3). *PRKCB* was a hub gene of the gene network constructed by differentially expressed genes in CD4$^+$ T cells in RA, involved in diverse signalling pathways.[45]

In summary, we performed the largest genome-wide meta-analysis using RA associations in three large cohorts comprising >300 000 East Asian and European individuals. Our computational analyses provided new insights and enhanced evidence regarding the genetic architecture/liability, disease-driving variants/genes/TFs/pathways/tissues and potential therapeutic targets.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs
Eunji Ha http://orcid.org/0000-0001-8618-8347
Sang-Cheol Bae http://orcid.org/0000-0003-4658-1093
Kwangwoo Kim http://orcid.org/0000-0001-8926-6216

REFERENCES

1 Firestein GS, Gabriel SE, McInnes IB, *et al*. *Kelley and Firestein's textbook of rheumatology*. 10th edn. Philadelphia, PA: Elsevier, 2017.
2 de Brito Rocha S, Baldo DC, Andrade LEC. Clinical and pathophysiologic relevance of autoantibodies in rheumatoid arthritis. *Adv Rheumatol* 2019;59:2.
3 Scott DL, Wolfe F, Huizinga TWJ. Rheumatoid arthritis. *Lancet* 2010;376:1094–108.
4 MacGregor AJ, Snieder H, Rigby AS, *et al*. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000;43:30–7.
5 Kim K, Bang S-Y, Lee H-S, *et al*. Update on the genetic architecture of rheumatoid arthritis. *Nat Rev Rheumatol* 2017;13:13–24.
6 Okada Y, Wu D, Trynka G, *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;506:376–81.
7 Stahl EA, Wegmann D, Trynka G, *et al*. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 2012;44:483–9.
8 Kwon Y-C, Lim J, Bang S-Y, *et al*. Genome-Wide association study in a Korean population identifies six novel susceptibility loci for rheumatoid arthritis. *Ann Rheum Dis* 2020;79:1438–45.
9 Ishigaki K, Akiyama M, Kanai M, *et al*. Large-Scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat Genet* 2020;52:669–79.
10 Willer CJ, Li Y, Abecasis GR. Metal: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–1.
11 Cochran WG. The comparison of percentages in matched samples. *Biometrika* 1950;37:256–66.
12 Bulik-Sullivan BK, Loh P-R, Finucane HK, *et al*. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015;47:291–5.
13 Brown BC, Ye CJ, *et al*, Asian Genetic Epidemiology Network Type 2 Diabetes Consortium. Transethnic Genetic-Correlation estimates from summary statistics. *Am J Hum Genet* 2016;99:76–88.
14 Yang J, Lee SH, Goddard ME, *et al*. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76–82.
15 Dubois PCA, Trynka G, Franke L, *et al*. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;42:295–302.
16 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–4.
17 Boyle AP, Hong EL, Hariharan M, *et al*. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012;22:1790–7.
18 Finucane HK, Bulik-Sullivan B, Gusev A, *et al*. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* 2015;47:1228–35.
19 Fisher RA. *Statistical methods for research workers*. 4th. Edinburgh etc: Oliver and Boyd, 1932.
20 Schmidt EM, Zhang J, Zhou W, *et al*. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics* 2015;31:2601–6.
21 , Kundaje A, Meuleman W, *et al*, Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–30.

22 Watanabe K, Taskesen E, van Bochoven A, *et al*. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017;8:1826.

23 de Leeuw CA, Mooij JM, Heskes T, *et al*. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015;11:e1004219.

24 van der Wijst MGP, Brugge H, de Vries DH, *et al*. Single-Cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat Genet* 2018;50:493–7.

25 Schmiedel BJ, Singh D, Madrigal A, *et al*. Impact of genetic polymorphisms on human immune cell gene expression. *Cell* 2018;175:e16:1701–15.

26 GTEx Consortium. The Genotype-Tissue expression (GTEx) project. *Nat Genet* 2013;45:580–5.

27 Bonder MJ, Luijk R, Zhernakova DV, *et al*. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* 2017;49:131–8.

28 Westra H-J, Peters MJ, Esko T, *et al*. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013;45:1238–43.

29 Zhernakova DV, Deelen P, Vermaat M, *et al*. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet* 2017;49:139–45.

30 Grundberg E, Small KS, Hedman Åsa K, *et al*. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* 2012;44:1084–9.

31 Schmitt AD, Hu M, Jung I, *et al*. A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep* 2016;17:2042–59.

32 , Forrest ARR, Kawaji H, *et al*, FANTOM Consortium and the RIKEN PMI and CLST (DGT). A promoter-level mammalian expression atlas. *Nature* 2014;507:462–70.

33 Sakaue S, Okada Y. GREP: genome for repositioning drugs. *Bioinformatics* 2019;35:3821–3.

34 Hwang S, Kim CY, Yang S, *et al*. HumanNet V2: human gene networks for disease research. *Nucleic Acids Res* 2019;47:D573–80.

35 Skrbo A, Zulić I, Hadzić S, *et al*. [Anatomic-therapeutic-chemical classification of drugs]. *Med Arh* 1999;53:57–60.

36 Gutierrez-Achury J, Zorro MM, Ricaño-Ponce I, *et al*. Functional implications of disease-specific variants in loci jointly associated with coeliac disease and rheumatoid arthritis. *Hum Mol Genet* 2016;25:180–90.

37 Acosta-Herrera M, Kerick M, González-Serna D, *et al*. Genome-wide meta-analysis reveals shared new *loci* in systemic seropositive rheumatic diseases. *Ann Rheum Dis* 2019;78:311–9.

38 Macian F. Nfat proteins: key regulators of T-cell development and function. *Nat Rev Immunol* 2005;5:472–84.

39 Koren I, Reem E, Kimchi A. Dap1, a novel substrate of mTOR, negatively regulates autophagy. *Curr Biol* 2010;20:1093–8.

40 Caza T, Landas S. Functional and Phenotypic Plasticity of CD4(+) T Cell Subsets. *Biomed Res Int* 2015;2015:1–13.

41 Kennedy JM, Fodil N, Torre S, *et al*. CCDC88B is a novel regulator of maturation and effector functions of T cells during pathological inflammation. *J Exp Med* 2014;211:2519–35.

42 Ananieva O, Darragh J, Johansen C, *et al*. The kinases MSK1 and MSK2 act as negative regulators of Toll-like receptor signaling. *Nat Immunol* 2008;9:1028–36.

43 Gorman JA, Babich A, Dick CJ, *et al*. The cytoskeletal adaptor protein IQGAP1 regulates TCR-mediated signaling and filamentous actin dynamics. *J Immunol* 2012;188:6135–44.

44 Kim J-H, Hedrick S, Tsai W-W, *et al*. Creb coactivators CRTC2 and CRTC3 modulate bone marrow hematopoiesis. *Proc Natl Acad Sci U S A* 2017;114:11739–44.

45 Ye H, Zhang J, Wang J, *et al*. Cd4 T-cell transcriptome analysis reveals aberrant regulation of STAT3 and Wnt signaling pathways in rheumatoid arthritis: evidence from a case-control study. *Arthritis Res Ther* 2015;17:76.

46 Jostins L, Ripke S, Weersma RK, *et al*. Host-Microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119–24.

47 Fodil N, Moradin N, Leung V, *et al*. CCDC88B is required for pathogenesis of inflammatory bowel disease. *Nat Commun* 2017;8:932.