2015

High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci

K. Kim

S. Y. Bang

H. S. Lee

S. K. Cho

C. B. Choi

See next page for additional authors

Authors
K. Kim, S. Y. Bang, H. S. Lee, S. K. Cho, C. B. Choi, Y. K. Sung, T. H. Kim, P. K. Gregersen, S. C. Bae, and +32 additional authors

HHS Public Access

Author manuscript

Ann Rheum Dis. Author manuscript; available in PMC 2015 June 15.

Published in final edited form as:

Ann Rheum Dis. 2015 March ; 74(3): e13. doi:10.1136/annrheumdis-2013-204749.

High-Density Genotyping of Immune Loci in Koreans and Europeans Identifies Eight New Rheumatoid Arthritis Risk Loci

Kwangwoo Kim[1,2,3,4], **So-Young Bang**[1], **Hye-Soon Lee**[1], **Soo-Kyung Cho**[1], **Chan-Bum Choi**[1], **Yoon-Kyoung Sung**[1], **Tae-Hwan Kim**[1], **Jae-Bum Jun**[1], **Dae Hyun Yoo**[1], **Young Mo Kang**[5], **Seong-Kyu Kim**[6], **Chang-Hee Suh**[7], **Seung-Cheol Shim**[8], **Shin-Seok Lee**[9], **Jisoo Lee**[10], **Won Tae Chung**[11], **Jung-Yoon Choe**[6], **Hyoung Doo Shin**[12], **Jong-Young Lee**[13], **Bok-Ghee Han**[13], **Swapan K. Nath**[14], **Steve Eyre**[15], **John Bowes**[15], **Dimitrios A. Pappas**[16], **Joel M. Kremer**[17], **Miguel A Gonzalez-Gay**[18], **Luis Rodriguez-Rodriguez**[19], **Lisbeth Ärlestig**[20], **Yukinori Okada**[2,3,4,21,22], **Dorothée Diogo**[2,3,4], **Katherine P. Liao**[2], **Elizabeth W. Karlson**[2], **Soumya Raychaudhuri**[2,3,4], **Solbritt Rantapää-Dahlqvist**[20], **Javier Martin**[23], **Lars Klareskog**[24], **Leonid Padyukov**[24], **Peter K. Gregersen**[25], **Jane Worthington**[14], **Jeffrey D. Greenberg**[26], **Robert M. Plenge**[2,3,4], and **Sang-Cheol Bae**[1]

[1]Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of Korea

[2]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

[3]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

[4]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA

[5]Division of Rheumatology, Department of Internal Medicine, Kyungpook National University School of Medicine, Daegu, Republic of Korea

[6]Division of Rheumatology, Department of Internal Medicine, Arthritis & Autoimmunity Research Center, Catholic University of Daegu School of Medicine, Daegu, Republic of Korea

[7]Department of Rheumatology, Ajou University School of Medicine, Suwon, Republic of Korea

[8]Division of Rheumatology, Daejeon Rheumatoid & Degenerative Arthritis Center, Chungnam National University Hospital, Daejeon, Republic of Korea

Correspondence to Robert M. Plenge, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; rplenge@partners.org or Sang-Cheol Bae, Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of Korea; scbae@hanyang.ac.kr.

Contributors KK, RMP and SCB designed the study. KK, SYB, HSL, HDS, RMP and SCB analyzed the data. SYB, HSL, SKC, CBC, YKS, THK, JBJ, DHY, YMK, SKK, CHS, SCS, SSL, JL, WTC, JYC, JYL, BGH, SKN, SE, JB, DAP, JMK, MAG-G, LR, LA, YO, DD, KPL, EWK, SR, SR-D, JM, LK, LP, PKG, JW, JDG, RMP and SCB recruited, characterized the cases and controls or analyzed European genotype data. KK, RMP and SCB wrote the manuscript. All authors reviewed and approved the manuscript.

Competing Interests None

Patient consent Obtained.

Ethics approval The study was approved by the Institutional Review Board of all participating institutions.

[9]Division of Rheumatology, Department of Internal Medicine, Chonnam National University Medical School and Hospital, Gwangju, Republic of Korea

[10]Division of Rheumatology, Department of Internal Medicine, Ewha Womans University School of Medicine, Seoul, Republic of Korea

[11]Division of Rheumatology, Department of internal medicine, Dong-A University, Busan, Republic of Korea

[12]Department of Life Science, Sogang University, Seoul, Republic of Korea

[13]Center for Genome Science, Korea National Institute of Health, Osong Health Technology, Chungcheongbuk-do, Republic of Korea

[14]Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, USA

[15]Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester Academic Health Sciences Centre, Manchester, UK

[16]Department of Medicine, Division of Rheumatology, Columbia University, New York, New York, USA

[17]The Center of Rheumatology, Albany, New York, USA

[18]Department of Rheumatology, Hospital Marques de Valdecilla, IFIMAV, Santander, Spain

[19]Department of Rheumatology, Hospital Clinico San Carlos, Madrid, Spain

[20]Department of Clinical Medicine/Rheumatoloy, Umeå University, Umeå, Sweden

[21]Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan

[22]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

[23]Instituto de Parasitologia y Biomedicina Lopez-Neyra, CSIC, Granada, Spain

[24]Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden

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

[26]Division of Rheumatology, New York University School of Medicine, New York, New York, USA

Abstract

Objective—A highly polygenic etiology and high degree of allele-sharing between ancestries have been well-elucidated in genetic studies of rheumatoid arthritis. Recently, the high-density genotyping array Immunochip for immune disease loci identified 14 new rheumatoid arthritis risk loci among individuals of European ancestry. Here, we aimed to identify new rheumatoid arthritis risk loci using Korean-specific Immunochip data.

Methods—We analyzed Korean rheumatoid arthritis case-control samples using the Immunochip and GWAS array to search for new risk alleles of rheumatoid arthritis with anti-citrullinated peptide antibodies. To increase power, we performed a meta-analysis of Korean data with previously published European Immunochip and GWAS data, for a total sample size of 9,299 Korean and 45,790 European case-control samples.

Results—We identified 8 new rheumatoid arthritis susceptibility loci (*TNFSF4, LBH, EOMES, ETS1–FLI1, COG6, RAD51B, UBASH3A* and *SYNGR1*) that passed a genome-wide significance threshold ($p<5\times10^{-8}$), with evidence for three independent risk alleles at 1q25/*TNFSF4*. The risk alleles from the 7 new loci except for the *TNFSF4* locus (monomorphic in Koreans), together with risk alleles from previously established RA risk loci, exhibited a high correlation of effect sizes between ancestries. Further, we refined the number of SNPs that represent potentially causal variants through a trans-ethnic comparison of densely genotyped SNPs.

Conclusion—This study demonstrates the advantage of dense-mapping and trans-ancestral analysis for identification of potentially causal SNPs. In addition, our findings support the importance of T cells in the pathogenesis and the fact of frequent overlap of risk loci among diverse autoimmune diseases.

Keywords

Rheumatoid arthritis; Gene polymorphism; Anti-CCP

Introduction

Rheumatoid arthritis (RA; OMIM180300) is a chronic and systemic autoimmune disease affecting up to 1% of the adult population worldwide.[1] Total heritability of RA was estimated to be ~65% from a previous twin study that compared the disease discordance in monozygotic twins with dizygotic twins.[2] To date, >50 risk loci have been discovered among individuals of European and Asian ancestry,[3,4] but only ~16% of the heritability (or ~25% of the genetic heritability) could be explained by the confirmed risk loci.[5] Current RA genetic studies have revealed a highly polygenic etiology by an inferred genetic architecture that hundreds (if not thousands) of common SNPs with modest effect and smaller number of rare causal variants account for total genetic heritability of RA.[6]

A previous study reported the majority of the RA susceptibility alleles are shared among individuals of European and Japanese ancestry (correlation coefficient for effect sizes between ancestries=0.82).[3] This observation suggests that there are shared causal variants between the two populations, and further suggests that a large cohort study using multiple ancestries can be powerful to detect new RA risk loci.[7]

Recently, 14 new RA risk loci were identified by integrating high-density genotype data of immune loci from the Immunochip (iChip) and imputed data from genome-wide association studies (GWAS) among individuals of European ancestry.[4] Here, we generate a new iChip + GWAS dataset among individuals of Korean ancestry. We perform a meta-analysis with a previously published European iChip + GWAS dataset,[4] as well as new iChip data from

individuals of European ancestry (n=2,840), for a total sample size of 9,299 Korean and 45,790 European case-control samples (online supplementary table S1).

METHODS

Korean Subjects, Genotyping, Imputation and Quality Control

A total of 4,689 subjects including 1,525 RA cases [average age=52.7 (17–82); female=83.4%] with anti-citrullinated peptide antibodies (ACPA) and 3,164 healthy controls [average age=39.4 (16–79); female=60.7%] were genotyped by the customized iChip array at multiple centers (SNP Genetics, Inc. Korea and Feinstein Institute for Medical Research for BAE#1; Oklahoma Medical Research Foundation for BAE#2; University of Queensland Diamantina Institute for BAE#3) (online supplementary table S1). Each sub-collection was filtered using standard criteria based on minor allele frequency (MAF), Hardy-Weinberg equilibrium (HWE), call rate per SNP, call rate per individual, unique mapping, genetic homogeneity, population stratification, and cryptic relatedness (duplicate and cryptic 1st degree relative) before and/or after merging the sub-collections (online supplementary table S2).

An independent set of 3,700 Korean controls [average age = 59.8 (38–89); female = 63.0%] was examined by Illumina HumanOmni1-Quad BeadChip GWAS array at Korea National Institute of Health. A sample was removed if it showed low call rate (<0.95), heterozygosity <0.26 or >0.30, cryptic relatedness (cryptic 1st degree relative) or sex inconsistency. The genotype data was merged with a previously reported Korean RA GWAS dataset of 709 ACPA-positive RA cases [average age=52.4 (19–82); female=89.7%] and 201 controls [average age=39.8 (17–76); female=74.6%].[8]

After merging the GWAS datasets, we applied the quality control criteria of MAF ≥0.01, p value of HWE<10^{-4} and call rate>95% in controls and cases. The quality control-passed control-case GWAS data were pre-phased to construct haplotypes of the autosomal chromosomes by SHAPEIT[9] and subsequent imputation was performed by IMPUTE2[10] using the 1000 Genomes Project phase I reference panel. The imputed data was filtered by the criteria of MAF>1%, Maximum probability ≥0.9, p value of HWE<10^{-4} and Call rate per SNP ≥0.95. A total of 65,014 iChip markers were extracted from the imputed data (online supplementary table S3).

Principal component analysis was performed by the SNP & Variation Suite 7 software using the SNPs with MAF>5% and linkage disequilibrium (LD) r^2<0.2 with other SNPs in a +/−250kb window. Cryptic relatedness with high kinship coefficient (>0.177) corresponding to the duplicate or 1st-degree relative was estimated by the KING software.[11] Quantile-quantile (Q-Q) plots were generated using the SNPs associated with reading and writing ability to estimate inflation.

In order to examine the third-effect SNP rs2027498 in *TNFSF4,* an independent Korean cohort including 1,254 RA cases [average age=53.1 (17–88); female=87.2%; ACPA(+)=71.1%, ACPA(−)=2.0%, ACPA(not tested)=26.9%] and 1,011 healthy controls [average age=54.0 (40–86); female=80.0%] was recruited. The genotyping was performed

using Sequenom iPlex platform at LabGenomics Co. LTD, showing a genotype call rate of 98.1% and no deviation from HWE.

All patients fulfilled the 1987 criteria of the American College of Rheumatology. The level of ACPA in the Korean RA patients was measured by using the ImmuLisa CCP2 enzyme-linked immunosorbent assay kit. Anti-cyclic citrullinated peptide (Anti-CCP) >25 units/ml was considered as positive for ACPA.

European Subjects, Imputation and quality control

The seven European iChip data (ES, NL, SE-E, SE-U, UK, US and i2b2/CORRONA) listed in supplementary table S1 were reported in two previous studies.[4,12] ACPA-negative samples were additionally removed from all data passing the quality control criteria in the previous studies. The ACPA-positive iChip data were further filtered for missingness per SNP/Individual (<0.01), MAF (<0.01) and HWE ($p<10^{-7}$).

Each European GWAS data (BRASS, CANADA, NARAC2 and WTCCC) independent of the Immunochip data was pre-phased and imputed by MaCH[13] and Minimac[14] using the high-density phased reference data from the 1000 Genome Project instead of the reference phase of the HapMap2 data used in the previous study.[4] The imputed data was filtered to restrict the analysis to ACPA positive cases and SNPs with MAF ≥1% and imputation-r^2 ≥0.5. Q-Q plots were generated for each European collection using the SNPs associated with reading and writing ability. All patients in the reported datasets fulfilled the 1987 criteria of the American College of Rheumatology or were diagnosed by a professional rheumatologist.

RA Association Test

odds ratio (OR) and 95% confidence interval (CI) were calculated by PLINK[15] with adjustment for top 5 or 10 PCs in the logistic regression to test whether each autosomal SNP was associated with susceptibility to RA. A fixed-effects inverse-variance meta-analysis of the association results from each collection and a heterogeneity test of the SNP effects among the collections were conducted by the GWAMA software.[16] Statistical power to detect RA association was calculated using the CaTS Power calculator.

Secondary-Effect Analysis

The presence of independent effect within the new RA loci with dense iChip markers was examined by conditioning on a lead SNP and/or other independent-effect SNPs and PCs as covariates. The conditional analysis using forward stepwise logistic regression for each collection and subsequent meta-analysis was performed by PLINK[15] and GWAMA[16], respectively. When a SNP was significant with $p<5\times10^{-5}$ in the conditional analysis, the SNP was considered to have independent effect.

The genotypes of the independent-effect SNPs (rs61828284, rs4090392 and rs2027498) in 1q25/*TNFSF4* were phased using PLINK[15]. The haplotypes were examined in the European collections that were successfully genotyped or imputed for all three SNPs. In Koreans,

rs61828284 and rs4090392 were too rare to infer the haplotype or to perform statistical analysis.

Proxy SNP Analysis

The HaploReg v2 software[17] was used to search the proxy SNPs of the RA-associated SNPs and to annotate the functional effects at the SNP position. Association of each SNP in the novel RA loci with the expression of the genes within the locus was evaluated using published eQTL data from the lymphoblastoid cell lines of the 856 healthy female twins of the MuTHER resource[18] and the HapMap population[19] in the Genevar v3.2.0 web-based software[20].

RESULTS

Korean-only Analysis

We first analyzed the Korean iChip + GWAS data from 2,234 cases and 7,065 controls. A total of 96,952 iChip SNPs passed quality control in the Korean iChip dataset (online supplementary table S2) and the GWAS-based imputed dataset (online supplementary table S3 and figure S1). Principal components analysis was used to correct for population stratification, showing no outliers (online supplementary figure S2). We performed logistic regression analyses to calculate OR and 95% CI adjusted for the top 10 principal components in the iChip and GWAS datasets independently, followed by an inverse-variance-weighted fixed-model meta-analysis. Q-Q plots of p values and inflation factors indicated little evidence of systematic bias (λ_{1000} = 1.010; online supplementary figure S3).

While we were able to replicate four known RA risk loci (*HLA, PADI4, STAT4* and *RASGRP1*), no new RA risk loci were identified in the Korean-only analysis at a genome-wide level of significance ($p<5\times10^{-8}$; online supplementary figure S4). However, this analysis was underpowered to achieve the genome-wide significance for alleles with modest effect (online supplementary figure S5).

Korean-European meta-analysis: Identification of 8 new RA risk loci

Since most known rheumatoid arthritis risk alleles are shared among individuals of European and Asian ancestry,[3] we performed a meta-analysis with recently published iChip + GWAS data[4] and new iChip data[12] derived from case-control samples of European ancestry. The GWAS data were imputed to iChip markers using 1000 Genome Project data. A total of 133,816 SNPs were analyzed in the Korean-European meta-analysis (λ_{1000} = 1.003) (online supplementary figure S1 and S3).

In a combined analysis, we found 8 new RA susceptibility loci passing a genome-wide significance threshold of $p<5\times10^{-8}$ (figure 1). For 7 of the 8 loci, the new signal of association was driven by adding the Korean data to the previously published European iChip + GWAS data. One locus on 1q25 (containing the *TNFSF4* gene), was identified by adding new European iChip data to the previously published iChip + GWAS European dataset (table 1). A Cochran's Q test with a forest plot for each associated locus indicated no

obvious evidence of heterogeneity across sample collections (figure 2 and online supplementary table S4).

The lead SNP at the *UBASH3A* locus, rs1893592, demonstrated evidence of genetic heterogeneity among the collections, ($p_{q\text{-}test}$=0.0029). However, this heterogeneity was derived mainly from single collection of European ancestry (NL), rather than heterogeneity between Asian and European populations (figure 2).

In addition to the novel loci, we confirmed association at 35 known non-HLA loci at $p<5\times10^{-8}$ (figure 1). Another 12 known RA risk loci included on the iChip array demonstrated consistent direction of effects at a suggestive level of significance ($5\times10^{-8}<p\leq1.31\times10^{-3}$) (online supplementary table S4). The only previously reported RA risk allele in a Japanese population that did not demonstrate an association in our analysis ($p>0.05$) was at the *CD244* locus (online supplementary table S4).

The SNP with the strongest signal of association at each of the new and known loci (excluding the *CD244* locus) exhibited a high correlation of effect sizes between populations, with the same allele conferring RA risk in individuals of Korean and European ancestry for most loci (correlation coefficient=0.91; online supplementary figure S6). There were, however, 10 SNPs in which the European risk allele was present at a very low MAF among individuals of Korean ancestry, as has been observed for the risk allele of *PTPN22*.[21]

Three independent risk alleles in 1q25/*TNFSF4*

For the 8 novel loci, we investigated for evidence of independent risk alleles by performing conditional analysis on the lead SNP in the combined Korean-European sample set. We found evidence for two additional risk alleles at the 1q25/*TNFSF4* locus: both rs4090392 ($p_{cond}=3.82\times10^{-7}$) and rs2027498 ($p_{cond}=1.16\times10^{-5}$) contributed signals independent of the lead SNP rs61828284 (online supplementary figure S7, S8 and table S5). However, 2 of the 3 1q25/*TNFSF4* SNPs were very rare among Koreans (rs61828284 and rs4090392); the third SNP (rs2027498) was common in Koreans and showed the same direction of effect in both populations (OR_{EUR}=0.90 [0.86–0.94] and OR_{KOR}=0.95 [0.84–1.09]). Among Europeans, these three SNPs are in weak LD ($r^2\leq0.1$, $|D'|\leq0.66$). The *TAA* haplotype carrying the three protective alleles exhibited the greatest protection against RA (OR=0.67 [0.57–0.77]; online supplementary figure S9).

It was possible that third effect SNP was associated in Koreans as well as Europeans because the effect size of the SNP was similar between populations. In order to increase power (52%), we attempted additional genotyping of rs2027498 for an independent set of 1,254 RA cases and 1,011 controls from Korea, but did not find evidence of association (p_{meta}=0.87; OR=1.01; online supplementary Table S6).

Identification of potentially causal SNPs by trans-ethnic mapping

To find the most likely causal variants in the 7 new loci with consistent signals of association in both Koreans and Europeans (excluding 1q25/*TNFSF4*), we examined local patterns of LD and annotated putative functional variants. The lead SNPs at these 7 loci were in LD with 55 and 51 SNPs at $r^2\geq0.9$ in the 1000 Genome Project data of East Asian

(CHB+JPT) and European (CEU) ancestry, respectively. Considering the overlap between the two populations, we observed only 37 SNPs in LD with the lead SNPs in both populations, none of which was a nonsynonymous, nonsense or splicing-site variant (online supplementary table S7).

When we applied the trans-ethnic mapping to 21 known RA loci and 7 new RA loci showing association in meta-analysis ($p<5\times10^{-8}$) and each of populations ($p<0.05$), we could narrow down the number of proxy SNPs from 492 SNPs (445 SNPs in Koreans and 231 SNPs in European) to 182 SNPs in the 21 known loci. Among the 28 RA loci, 16 and 19 loci had the decreased number of the proxy SNPs in Koreans and Europeans, respectively (online supplementary figure S10 and table S8).

An example of the trans-ethnic mapping approach is shown in figure 3. Only 2 out of 8 SNPs were in LD with the lead SNP in both populations: rs909685 (lead SNP) and rs2069235 (r^2=0.92 in Asian and r^2=1.00 in European). This locus contains *SYNGR1* (an integral membrane protein associated with presynaptic vesicles in neuronal cells, which is also known as a susceptibility locus for primary biliary cirrhosis.[22] Both variants lie within putative functional sequences (e.g., transcriptional factor-binding motif alteration, histone marks, DNase hypersensitivity, DNA-binding proteins). Furthermore, both SNPs were identical to the most significant eQTL associated with *SYNGR1* expression from lymphoblastoid cell lines in previous four studies (online supplementary figure S11).[18,19,23,24]

DISCUSSION

Based on previous Asian-European studies[3,25] and theoretical estimates of the polygenic architecture of RA[6], a large number of common genetic variants with modest effect size on risk of RA are expected to be discovered by a trans-ethnic approach. Here, we identified 8 new loci with relatively modest OR ($0.81 \leq OR \leq 0.91$ and $1.10 \leq OR \leq 1.13$) using a large cohort of Korean and European populations, bring the number of RA loci to >60.

There are several important observations from our study. First, the association of the new loci exhibits a similar trend of ORs between Korean and European populations. Further, there were similar trends of ORs among the 13 data collections and no remarkable deviation between observed and reported MAF, which indicates little possibility of systemic bias or error.

Second, our study continues to emphasize the importance of T cell biology in the pathogenesis of RA.[26] Four of the 8 new loci contain genes with established function in CD4+ T cells. *TNFSF4*, which encodes a cytokine OX40L in the TNF ligand family, is involved in the regulation of T cell-mediated immune responses.[27] *EOMES* encodes a transcription factor with a crucial role in differentiation of CD8+ T cells and homeostasis of effector and memory T-cells.[28] *ETS1* encodes a transcription factor that regulates the differentiation of Th1 cells, regulatory T cells and B cells by affecting the function of other key regulators like T-bet, Foxp3 and Blimp-1.[29] *UBASH3A* encodes a T-cell ubiquitin ligand that acts as a negative regulator of T-cell receptor (TCR)-mediated signaling.[30]

Third, many of new RA risk loci are also associated with other complex diseases, especially autoimmune diseases (*TNFSF4* with systemic lupus erythematosus (SLE)[31,32] and Crohn's disease[33]; *EOMES* with multiple sclerosis[34]; *ETS1* with SLE[31,32,35] and celiac disease[36]; *RAD51B* with primary biliary cirrhosis[37] and breast cancer[38]; *UBASH3A* with type 1 diabetes[39,40] and vitiligo[41]; *SYNGR1* with primary biliary cirrhosis[22]). *LBH* and *UBASH3A* have been suggested as shared risk loci of RA and celiac diseases in a meta-analysis of RA and celiac datasets,[42] but not established in each disease alone. Although functional connection between the new loci and RA pathogenesis is not elucidated yet, the overlap suggests a shared etiology across autoimmune diseases.[43] All associated genes are known to be highly expressed in immune cells like T cells or broadly expressed across tissues including blood cells (online supplementary table S9).

Fourth, we found evidence that multiple alleles contribute to risk at one of newly discovered loci. Using a conditional test among individuals of European ancestry, we found that the lead SNP rs61828284 in 1q25/*TNFSF4* could not account for the secondary effect at rs4090392, nor at the third-effect SNP, rs2027498. Notably, the association of the secondary-effect SNP but not the third effect SNP became strengthened by conditioning of the lead SNPs in the meta-analysis. A similar trend was observed among all the European-ancestry collections. The masked or strengthened significance of the secondary SNP before or after conditioning may be the result of Simpson's paradox, which has been shown in a celiac locus, *SOCS1*.[44]

Finally, we used a trans-ethnic approach to narrow the list of putative causal alleles at each of the 28 RA risk loci with association signals in both Asians and Europeans. We narrowed down 563 proxy SNPs (in LD with the lead SNP at r^2>0.90 in one or the other population) to 219 putative causal variants (in LD with the lead SNP in *both* populations). One example is shown at the *SYNGR1* locus (figure 3). Similarly, the lead SNP at *UBASH3A* (rs1893592) among the new loci shared only one proxy SNP in both populations, which is located in 3-bp downstream of the boundary between exon 10 and intron 10 of *UBASH3A*. The minor allele *C* of rs1893592 disrupts the consensus sequence *R* (=*A* or *G*) of the splice donor site in contrast to the major allele *A*, although biological validation is required for the potential alternative splicing.

A limitation of our study is the coverage of the iChip among individuals of Asian ancestry. The customized iChip array was largely designed to capture variants identified in the CEU cohort of the 1000 Genomes Project pilot study. As a result, we observe that a large portion (42%) of the iChip markers were not polymorphic or rare (MAF<0.01) in the Korean dataset, compare to 22% in the European iChip dataset. Nevertheless, the SNPs with MAF ≥0.01 in the Korean population in the iChip target regions were still dense and almost overlapped with the SNP with MAF ≥0.01 in the European population (online supplementary figure S2) so that we could perform a high-quality meta-analysis, identify new RA risk loci, and narrow the list of putative causal variants through a trans-ethnic approach.

In summary, we identified 8 new RA loci using a dense genetic mapping approach, providing additional insight into the pathogenesis of RA risk. Fine-mapping and trans-

ancestral comparison narrowed the list of putative causal variants that may explain the underlying signal of association at these 8 new loci.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to all participants who donated blood samples for this study.

Funding: This study is supported by grants from the Korea Healthcare technology R&D project of Ministry for Health & Welfare (A121983), NIH grant (AI094377), the Arthritis Research UK grant reference number 17552 and the Manchester Biomedical Research Centre.

REFERENCES

1. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. Lancet. 2010; 376:1094–1108. [PubMed: 20870100]
2. MacGregor AJ, Snieder H, Rigby AS, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum. 2000; 43:30–37. [PubMed: 10643697]
3. Okada Y, Terao C, Ikari K, et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet. 2012; 44:511–516. [PubMed: 22446963]
4. Eyre S, Bowes J, Diogo D, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet. 2012; 44:1336–1340. [PubMed: 23143596]
5. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. Nat Rev Rheumatol. 2013; 9:141–153. [PubMed: 23381558]
6. Stahl EA, Wegmann D, Trynka G, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet. 2012; 44:483–489. [PubMed: 22446960]
7. Marigorta UM, Navarro A. High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants. PLoS Genet. 2013; 9:e1003566. [PubMed: 23785302]
8. Freudenberg J, Lee HS, Han BG, et al. Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. Arthritis Rheum. 2011; 63:884–893. [PubMed: 21452313]
9. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods. 2013; 10:5–6. [PubMed: 23269371]
10. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5:e1000529. [PubMed: 19543373]
11. Manichaikul A, Mychaleckyj JC, Rich SS, et al. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010; 26:2867–2873. [PubMed: 20926424]
12. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2013 published online on Dec 25.
13. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34:816–834. [PubMed: 21058334]
14. Howie B, Fuchsberger C, Stephens M, et al. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet. 2012; 44:955–959. [PubMed: 22820512]
15. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
16. Magi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics. 2010; 11:288. [PubMed: 20509871]

17. 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–D934. [PubMed: 22064851]
18. Grundberg E, Small KS, Hedman AK, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat Genet. 2012; 44:1084–1089. [PubMed: 22941192]
19. Stranger BE, Montgomery SB, Dimas AS, et al. Patterns of cis regulatory variation in diverse human populations. PLoS Genet. 2012; 8:e1002639. [PubMed: 22532805]
20. Yang TP, Beazley C, Montgomery SB, et al. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics. 2010; 26:2474–2476. [PubMed: 20702402]
21. Lee HS, Korman BD, Le JM, et al. Genetic risk factors for rheumatoid arthritis differ in Caucasian and Korean populations. Arthritis Rheum. 2009; 60:364–371. [PubMed: 19180477]
22. Liu JZ, Almarri MA, Gaffney DJ, et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nat Genet. 2012; 44:1137–1141. [PubMed: 22961000]
23. Stranger BE, Nica AC, Forrest MS, et al. Population genomics of human gene expression. Nat Genet. 2007; 39:1217–1224. [PubMed: 17873874]
24. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature. 2010; 464:773–777. [PubMed: 20220756]
25. Kurreeman FA, Stahl EA, Okada Y, et al. Use of a multiethnic approach to identify rheumatoid-arthritis-susceptibility loci, 1p36 and 17q12. Am J Hum Genet. 2012; 90:524–532. [PubMed: 22365150]
26. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. N Engl J Med. 2011; 365:2205–2219. [PubMed: 22150039]
27. Ishii N, Takahashi T, Soroosh P, et al. OX40-OX40 ligand interaction in T-cell-mediated immunity and immunopathology. Adv Immunol. 2010; 105:63–98. [PubMed: 20510730]
28. Kallies A. Distinct regulation of effector and memory T-cell differentiation. Immunol Cell Biol. 2008; 86:325–332. [PubMed: 18362944]
29. Leng RX, Pan HF, Chen GM, et al. The dual nature of Ets-1: focus to the pathogenesis of systemic lupus erythematosus. Autoimmun Rev. 2011; 10:439–443. [PubMed: 21296190]
30. Tsygankov AY. TULA-family proteins: a new class of cellular regulators. J Cell Physiol. 2013; 228:43–49. [PubMed: 22689384]
31. Yang W, Tang H, Zhang Y, et al. Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. Am J Hum Genet. 2013; 92:41–51. [PubMed: 23273568]
32. Han JW, Zheng HF, Cui Y, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nat Genet. 2009; 41:1234–1237. [PubMed: 19838193]
33. Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010; 42:1118–1125. [PubMed: 21102463]
34. Sawcer S, et al. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011; 476:214–219. [PubMed: 21833088]
35. Yang W, Shen N, Ye DQ, et al. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS Genet. 2010; 6:e1000841. [PubMed: 20169177]
36. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 2010; 42:295–302. [PubMed: 20190752]
37. Mells GF, Floyd JA, Morley KI, et al. Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. Nat Genet. 2011; 43:329–332. [PubMed: 21399635]

38. Orr N, Lemnrau A, Cooke R, et al. Genome-wide association study identifies a common variant in RAD51B associated with male breast cancer risk. Nat Genet. 2012; 44:1182–1184. [PubMed: 23001122]

39. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009; 41:703–707. [PubMed: 19430480]

40. Grant SF, Qu HQ, Bradfield JP, et al. Follow-up analysis of genome-wide association data identifies novel loci for type 1 diabetes. Diabetes. 2009; 58:290–295. [PubMed: 18840781]

41. Jin Y, Birlea SA, Fain PR, et al. Variant of TYR and autoimmunity susceptibility loci in generalized vitiligo. N Engl J Med. 2010; 362:1686–1697. [PubMed: 20410501]

42. Zhernakova A, Stahl EA, Trynka G, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet. 2011; 7:e1002004. [PubMed: 21383967]

43. Parkes M, Cortes A, van Heel DA, et al. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet. 2013

44. Trynka G, Hunt KA, Bockett NA, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011; 43:1193–1201. [PubMed: 22057235]

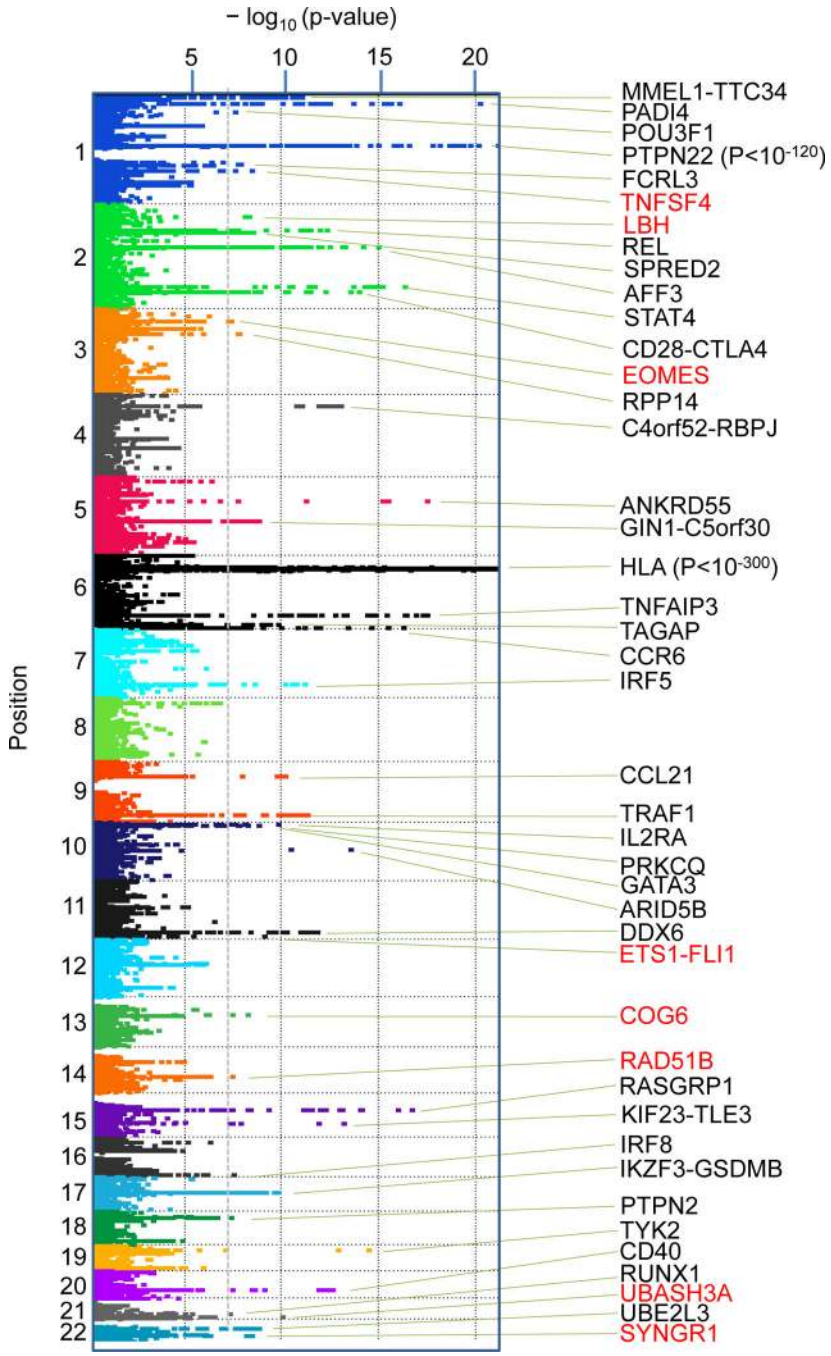

Figure 1.

Manhattan plot from the meta-analysis of the Korean and European datasets. Newly identified and known RA risk loci passing the genome-wide significance level ($p < 5\times10^{-8}$) are shown with the locus names in red and black, respectively. The dashed gray line indicates the genome-wide significance threshold.

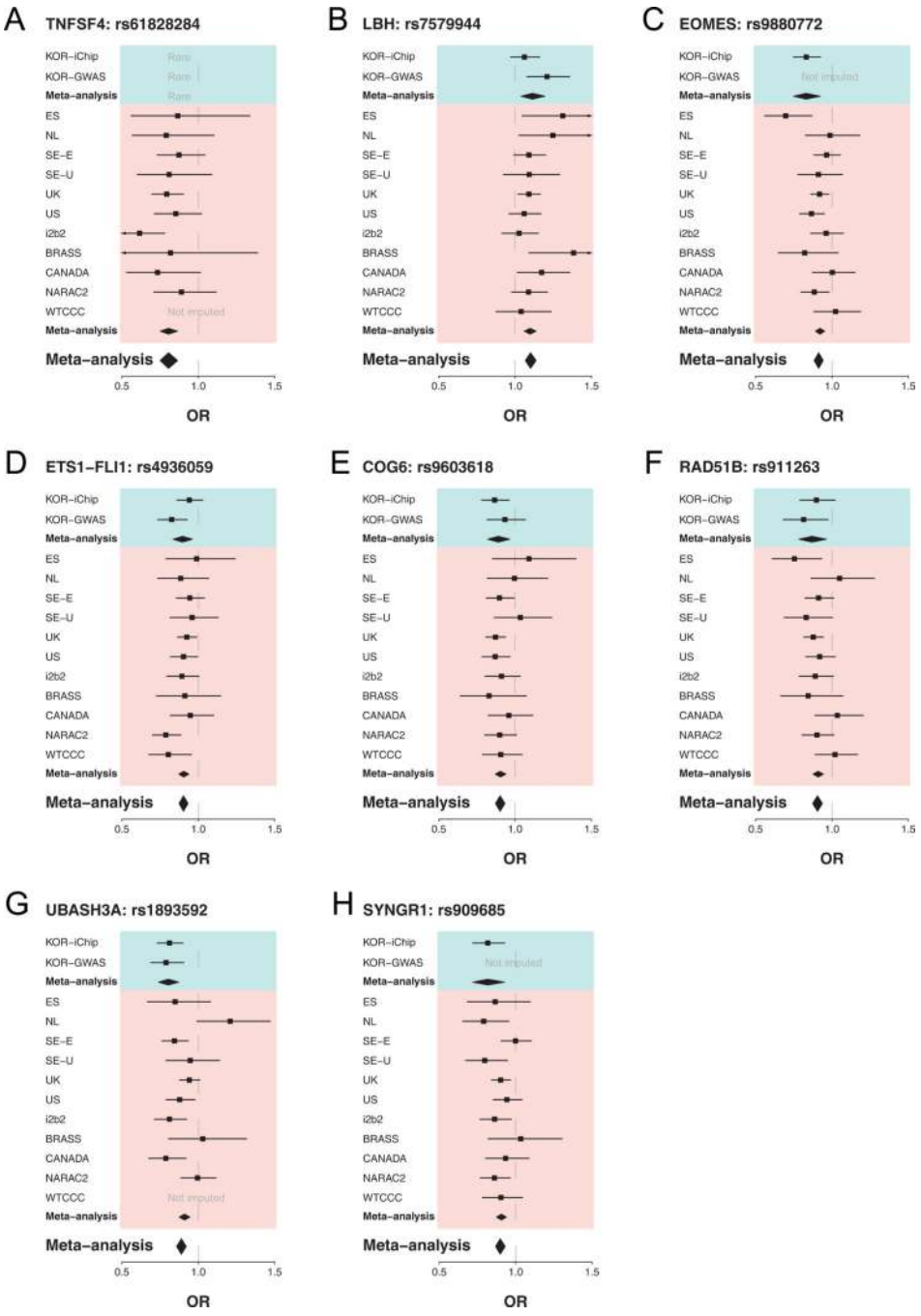

Figure 2.

Forest plots of the 8 lead SNPs for the 13 independent collections. ORs of the lead SNP in each RA susceptibility locus (A-H) were shown for each individual collection as well as the meta-analysis across the sample collections. Locus and SNP name are shown in the top of each plot. The OR patterns for Korean and European collections are on the blue and pink background, respectively.

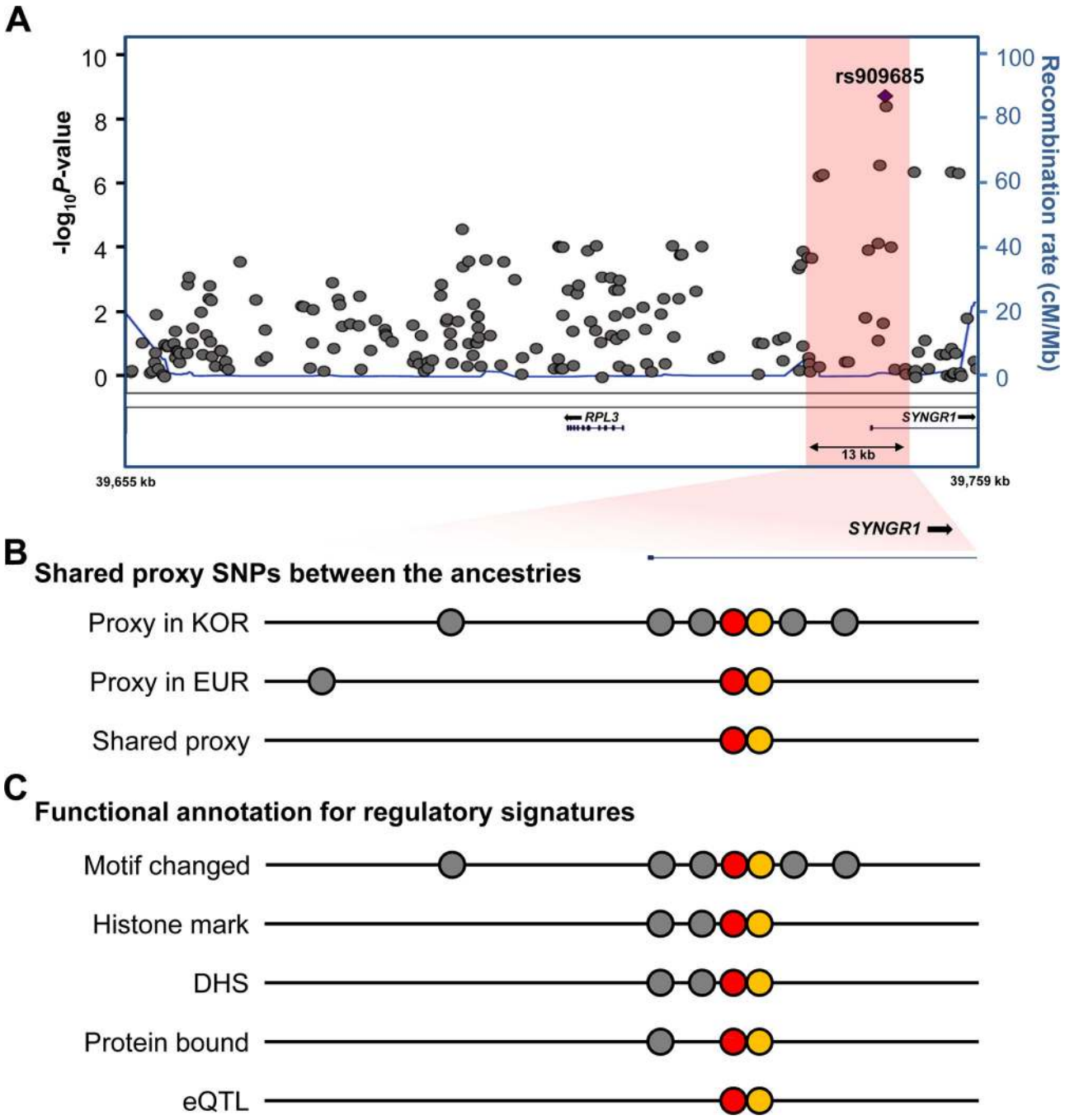

Figure 3.

Overlap of proxy SNPs shared in the ancestries and regulatory SNPs in the *SYNGR1* locus. (A) Regional association plot for the meta-analysis for the *RPL3-SYNGR1* locus in the chromosome 22. The best hit was found at rs909685 in the intron of *SYNGR1*. The 13-kb region containing all proxy SNPs of rs909685 was highlighted with pink. Coordinates are based on the hg19 assembly. (B) The proxy SNPs (circle) correlated with the lead SNP rs909685 in Korean and/or European (r^2 ≥0.9). The two proxy SNPs, rs909685 (red) and rs2069235 (orange), were observed in both population in contrast to the others (grey). (C)

The functional annotation of the proxy SNPs. Each proxy SNP was annotated for that it significantly alters the transcription factor binding motifs by the alleles and that it is in histone-mark region, DNase-hypersensitive site (DHS), Protein-binding site and eQTL.

Table 1

Association results for SNPs in novel RA loci

					Korean (n = 9,299)		European (n = 45,790)		Meta-analysis (n = 55,089)		LD region
rs_Number	**Chr**	**Position***	**Gene**	**EA/NEA**	**EAF**	**OR (95% CI)**	**EAF**	**OR (95% CI)**	**OR (95% CI)**	**p Value**	**(r^2 ≥0.9)**
rs1893592	21	43855067	*UBASH3A*	C/A	0.25	0.80 (0.74–0.87)	0.28	0.91 (0.87–0.95)	0.89 (0.86–0.92)	8.18×10^{-11}	Intron 10 of *UBASH3A*
rs4936059	11	128502496	*ETS1-FLI1*	A/G	0.38	0.90 (0.83–0.96)	0.34	0.90 (0.87–0.94)	0.90 (0.87–0.93)	6.94×10^{-10}	45kb to 47kb 5' of *ETS1*
rs61828284	1	173299743	*TNFSF4*	A/G	0.00	NA	0.08	0.81 (0.75–0.86)	0.81 (0.75–0.86)	2.66×10^{-9}	18kb 3' of *TNFSF4* to complete *TNFSF4* and intron 2 of *LOC100506023*†
rs909685	22	39747671	*SYNGR1*	A/T	0.15	0.82 (0.72–0.93)	0.31	0.91 (0.87–0.94)	0.90 (0.87–0.93)	2.95×10^{-9}	6.8kb 5' to intron 1 of *SYNGR1*
rs7579944	2	30445026	*LBH*	G/A	0.38	1.11 (1.04–1.20)	0.36	1.10 (1.06–1.14)	1.10 (1.07–1.14)	3.83×10^{-9}	12.0kb to 3.1kb 5' of *LBH*
rs9603618	13	40371377	*COG6*	A/G	0.24	0.89 (0.82–0.97)	0.27	0.90 (0.87–0.94)	0.90 (0.87–0.93)	4.53×10^{-9}	5.3kb to 18kb 3' of *COG6*
rs911263	14	68753593	*RAD51B*	G/A	0.14	0.87 (0.78–0.96)	0.28	0.91 (0.88–0.94)	0.90 (0.87–0.94)	3.09×10^{-8}	Intron 4 to 8 of *RAD51B*
rs9880772	3	27777779	*EOMES*	G/A	0.22	0.83 (0.75–0.92)	0.47	0.92 (0.89–0.95)	0.91 (0.88–0.94)	3.46×10^{-8}	3'UTR to 29kb 5' of *EOMES*

*Coordinates are based on the hg19 assembly.

†LD region was expanded by the proxy SNPs of independent-effect SNPs in the 1q25/TNFSF4 locus.

RA, rheumatoid arthritis; Chr, chromosome; EA, effect allele; NEA, non-effect allele; EAF, effect allele frequency; OR, odds ratio; CI confidence interval; LD, linkage disequilibrium; and NA, not applicable.