ARTICLE

Contribution of a Non-classical HLA Gene, *HLA-DOA*, to the Risk of Rheumatoid Arthritis

Yukinori Okada,[1,2,3,*] Akari Suzuki,[4] Katsunori Ikari,[5,6] Chikashi Terao,[7,8,9,10,11] Yuta Kochi,[4,6]
Koichiro Ohmura,[12] Koichiro Higasa,[10] Masato Akiyama,[2] Kyota Ashikawa,[13] Masahiro Kanai,[1,2]
Jun Hirata,[1,14] Naomasa Suita,[1,15] Yik-Ying Teo,[16] Huji Xu,[17] Sang-Cheol Bae,[18] Atsushi Takahashi,[2]
Yukihide Momozawa,[13] Koichi Matsuda,[19] Shigeki Momohara,[5] Atsuo Taniguchi,[5] Ryo Yamada,[10]
Tsuneyo Mimori,[12] Michiaki Kubo,[13] Matthew A. Brown,[20] Soumya Raychaudhuri,[7,8,9,21,22]
Fumihiko Matsuda,[10] Hisashi Yamanaka,[5] Yoichiro Kamatani,[2] and Kazuhiko Yamamoto[4,23]

Despite the progress in human leukocyte antigen (HLA) causal variant mapping, independent localization of major histocompatibility complex (MHC) risk from classical HLA genes is challenging. Here, we conducted a large-scale MHC fine-mapping analysis of rheumatoid arthritis (RA) in a Japanese population (6,244 RA cases and 23,731 controls) population by using HLA imputation, followed by a multi-ethnic validation study including east Asian and European populations (n = 7,097 and 23,149, respectively). Our study identified an independent risk of a synonymous mutation at *HLA-DOA*, a non-classical HLA gene, on anti-citrullinated protein autoantibody (ACPA)-positive RA risk (p = 1.4×10^{-9}), which demonstrated a cis-expression quantitative trait loci (cis-eQTL) effect on *HLA-DOA* expression. Trans-ethnic comparison revealed different linkage disequilibrium (LD) patterns in *HLA-DOA* and *HLA-DRB1*, explaining the observed *HLA-DOA* variant risk heterogeneity among ethnicities, which was most evident in the Japanese population. Although previous HLA fine-mapping studies have identified amino acid polymorphisms of the classical HLA genes as driving genetic susceptibility to disease, our study additionally identifies the dosage contribution of a non-classical HLA gene to disease etiology. Our study contributes to the understanding of HLA immunology in human diseases and suggests the value of incorporating additional ancestry in MHC fine-mapping.

Introduction

Rheumatoid arthritis (RA [MIM: 180300]), which affects up to 0.5% of the global population, is a common autoimmune disease characterized by inflammation and destruction of synovial joints.[1] Anti-citrullinated protein autoantibody (ACPA) is a highly specific biomarker of RA and is related to disease severity; positivity of ACPA is an important component of RA diagnosis, as well as its titer in serum.[1] ACPA-positive and -negative RA, which constitute approximately 70% and 30% of RA cases, respectively, are considered to be clinically distinct disease subsets.[1] Genetic and environmental factors explain the majority of RA susceptibility, and the major histocompatibility complex (MHC) region at 6p21 substantially contrib-

utes to 30%–50% of RA heritability.[2,3] In particular, polymorphisms in *HLA-DRB1* (MIM: 142857), one of the class II classical human leukocyte antigen (HLA) genes, are strongly associated with risk of ACPA-positive and -negative RA and with a quantitative trait locus (QTL) effect on serum ACPA titer.[4–12]

Whether a genetic risk from these ACPA-defined RA phenotypes exists in the MHC region independently of *HLA-DRB1* has long been a research question.[13] Recently, fine-mapping of the variants in the MHC region via HLA imputation[14,15] applied to genome-wide association studies (GWASs) or Immunochip SNP data has successfully identified an independent risk of ACPA-positive RA at multiple classical HLA genes in European populations (*HLA-A* [MIM: 142800], *HLA-B* [MIM: 142830], and *HLA-DPB1*

[1]Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, 113-0034 Tokyo, Japan; [2]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan; [3]Department of Statistical Genetics, Graduate School of Medicine, Osaka University, Osaka 565-0871, Japan; [4]Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan; [5]Institute of Rheumatology, Tokyo Women's Medical University, Tokyo 162-0054, Japan; [6]Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST), Tokyo 162-0054, Japan; [7]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; [8]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; [9]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA; [10]Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan; [11]Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, Kyoto 606-8501, Japan; [12]Department of Rheumatology and Clinical Immunology, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan; [13]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan; [14]Pharmaceutical Discovery Research Laboratories, Teijin Pharma, Hino 191-8512, Japan; [15]Advanced Medicinal Research Laboratories, Tsukuba Research Institute, Ono Pharmaceutical, Tsukuba 300-4247, Japan; [16]Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117549; [17]Department of Rheumatology and Immunology, Shanghai Changzheng Hospital and The Second Military Medical University, Shanghai 200433, China; [18]Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul 04763, South Korea; [19]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 108-8639 Tokyo, Japan; [20]University of Queensland Diamantina Institute, Translational Research Institute, Princess Alexandra Hospital, Brisbane, Queensland 4072, Australia; [21]Partners Center for Personalized Genetic Medicine, Boston, MA 02115, USA; [22]Faculty of Medical and Human Sciences, University of Manchester, Manchester M13 9PL, UK; [23]Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo 113-0033, Japan
*Correspondence: yokada@sg.med.osaka-u.ac.jp
http://dx.doi.org/10.1016/j.ajhg.2016.06.019.
© 2016

[MIM: 142858]).[7,8] The MHC is one of the most polymorphic genetic sites in the human genome.[16] A number of HLA-related genes (such as non-classical HLA genes, HLA-like genes, and pseudo HLA genes) and immune-related genes (such as *TNF* [MIM: 191160] and *BTNL2* [MIM: 606000]) are also located in the MHC region. As empirical examples, previous studies suggested additional RA genetic risk independent of the classical HLA genes.[17–19]

Due to strong selection pressure, the MHC region has a complex ancestry-related linkage disequilibrium (LD) structure that is characterized by population-specific long-range HLA haplotypes.[20] Given that HLA fine-mapping analysis results reflect the LD structure of examined populations, fine-mapping studies in additional populations, such as non-Europeans, could contribute to identification of additional independent association signals in the MHC region.[20]

Here, we performed a large-scale HLA fine-mapping analysis of RA in the Japanese population (6,244 RA cases with ACPA information and 23,731 controls), followed by a validation analysis in east Asian (n = 7,097)[10] and European (n = 23,149) populations.[8]

Material and Methods

RA GWAS Data in the Japanese Population
We used four GWAS datasets consisting of 6,244 RA cases (5,119 ACPA-positive RA cases and 1,125 ACPA-negative RA cases; Table S1) and 23,731 controls, all of Japanese ancestry. Of these, two GWAS datasets were previously constructed (BBJ1 and Kyoto University; 3,042 RA cases and 15,374 controls), and the other two sets were genotyped in this study (BBJ2 and IORRA; 3,202 RA cases and 8,357 controls).[21] All RA cases fulfilled the 1987 criteria of the American College of Rheumatology for RA diagnosis.[22] Serum ACPA levels were assayed with the second-generation ACPA ELISA kit (DIASTAT anti-CCP; MBL, Japan) with a positivity cutoff value of 4.5 U/mL. All subjects provided written informed consent approved by the institutional ethical committees.

Characteristics of the study cohorts, including SNP genotyping platforms and GWAS quality control (QC) criteria for the subjects and SNPs, are described in Table S1. After applying the QC filter separately for each GWAS dataset, as described elsewhere,[21] we estimated the top ten principal components (PCs) by using LD-pruned whole-genome SNP genotype data (excluding the SNPs in the MHC region), which were used as covariates to correct potential population structure.

HLA Imputation of the RA GWAS Data
We applied HLA imputation separately to the RA GWAS datasets by using the HLA imputation reference panel of the Japanese population (n = 908), as described elsewhere.[20] We extracted the genotypes of the SNPs located in the MHC region (defined as from 29 Mbp to 34 Mbp at chromosome 6 [NCBI Genome browser, build 37]) and imputed genotypes of the two-digit and four-digit HLA alleles and amino acid polymorphisms of the seven classical HLA genes (*HLA-A*, *HLA-B*, and *HLA-C* [MIM: 142840] for class I and *HLA-DRB1*, *HLA-DQB1* [MIM: 604305], *HLA-DPA1* [MIM: 142880], and *HLA-DPB1* for class II) by using the SNP2HLA software.[15] *HLA-DRA1* (MIM: 142860) and *HLA-DQA1* (MIM: 146880) were not included in the reference panel. We also imputed additional SNPs not genotyped in the original GWAS but genotyped in the reference panel. We applied post-imputation QC criteria of minor-allele frequency (MAF) $\geq$ 0.5% in each subject group (ACPA-positive and -negative RA cases and controls) and an imputation score of $Rsq \geq 0.5$ for fine-mapping analysis. High-resolution (four digit) genotyping of the *HLA-DRB1* allele was performed for some of the subjects in the BBJ1 and BBJ2 cohorts (n = 2,153), via a WAKFlow HLA typing kit (Wakunaga, Japan) with a Luminex multi-analyte profiling system (xMAP, Luminex). Accuracy of imputation was evaluated by assessment of concordances between imputed and genotyped *HLA-DRB1* alleles, as described elsewhere.[20]

Fine-Mapping Analysis of the HLA Variants with ACPA-Defined RA Phenotypes
We defined four ACPA-defined RA phenotypes: (1) that of ACPA-positive RA cases as compared to controls, (2) that of ACPA-negative RA cases as compared to controls, (3) that of ACPA-positive RA cases as compared to ACPA-negative RA cases, and (4) that of ACPA titer QTL in ACPA-positive RA cases. For the ACPA titer QTL analysis, we defined the trait rank value of the individual (= t_{CCP}) as any of (1), (2), or (3) as follows, considering the RA classification criteria and that the ACPA titers over the upper bound of the ELISA assay were difficult to precisely quantify: (1) when ACPA titer was between the cutoff value (= 4.5 U/mL) and three times the cutoff value (= 13.5 U/mL), $t_{CCP} = 1$; (2) when ACPA titer was between three times the cutoff value and the upper bound of the ELISA assay (= 100 U/mL), $t_{CCP} = 2$; and (3) when ACPA titer exceeded the upper bound of the ELISA assay, $t_{CCP} = 3$.

We evaluated associations of the HLA variants with the risk of the binary phenotypes (phenotypes 1–3) by using a logistic regression model and the effect of the quantitative phenotype (phenotype 4) by using a linear regression model, as described elsewhere.[10,20,23] We assumed additive effects of the allele dosages on the log-odds scale (phenotypes 1–3) or the linear scale (phenotype 4). In this study, we defined the HLA variants as bi-allelic SNPs in the MHC region (29 Mbp to 34 Mbp at chromosome 6 [NCBI Genome browser, build 37]), two-digit and four-digit bi-allelic HLA alleles, bi-allelic HLA amino acid polymorphisms corresponding to the respective residues, and multi-allelic HLA amino acid polymorphisms for each amino acid position. To robustly account for potential population stratification, we included the top ten PCs obtained from each genome-wide SNP data as covariates in the regression models. Given that we merged the imputed genotype data of the GWAS datasets into a single dataset, we also included dummy categorical variables as covariates that distinguished the respective GWAS datasets.

An omnibus p value for each HLA amino acid site was calculated by a log-likelihood ratio test comparing the likelihood of the null model against the likelihood of the fitted model, as described elsewhere.[10,20,23] We evaluated the significance of the improvement in model fitting by calculating the deviance, which follows a χ^2 distribution with m-1 degree(s) of freedom for an amino acid position with m residues.

Conditional Association Analysis of the HLA Variants
To evaluate independent risk among the HLA variants, we conducted a conditional regression analysis that additionally included

the HLA variant genotypes as covariates. When conditioned on the HLA gene(s), we robustly included all the two-digit and four-digit alleles and identified HLA risk variants as covariates. When conditioning on the specific HLA amino acid position(s), we included the multi-allelic variants of the amino acid residues. When conditioning on the HLA gene(s), we included all the two- and four-digit alleles of the corresponding HLA gene(s). We applied a forward stepwise conditional analysis for the HLA amino acid positions and then HLA genes, based on the genome-wide association significance threshold (p = 5.0 × 10^{-8}).

Explained Phenotypic Variance Considering Non-additive Effect of the HLA Variants

We estimated the phenotypic variances explained by the identified risk HLA variants. Variance of the binary phenotype (phenotypes 1–3) was estimated on the basis of a liability threshold model assuming a disease prevalence of 0.5%.[10,20,23] Variance of the quantitative phenotype was estimated on the basis of the explained variance in a linear regression model. For each phenotype, we performed a multivariate regression analysis that included four-digit alleles of the risk HLA genes and the risk SNP allele in the regression models. The HLA allele with the highest frequency for each HLA gene and non-risk SNP allele were used as references. Genetic risk scores (GRSs) were calculated separately for each phenotype on the basis of the effect sizes of the risk HLA variants obtained from the multivariate regression analysis on each phenotype. To comprehensively evaluate phenotypic variances, we also considered non-additive effects of the HLA variants, as described elsewhere.[24] We conducted a multivariate regression analysis that additionally included non-additive genotype dosages of the HLA variants with MAF ≥ 5% in the subject groups.

Cis-eQTL Analysis of HLA-DOA

A cis-eQTL effect of rs369150 genotypes on HLA-DOA (GenBank: NM_002119 [MIM: 142930]) mRNA expression in Japanese individuals was obtained from a previously conducted eQTL study of peripheral blood cells (PBCs) of the 298 individuals.[25] Expression was measured with a microarray technique (the A_32_P356316 probe for SurePrint G3 Human Gene Expression 8660K Microarray Kits [Agilent Technologies]).[25] No common variants (MAF ≥ 0.01 in the Japanese population) were located within the probe sequence region. A PBC cis-eQTL effect in Europeans was obtained from the Blood eQTL browser (n = 2,775).[26] Cell-type-specific expression profiles of HLA-DOA were obtained from the FANTOM database.[27]

A Multi-ethnic Validation Study of the HLA-DOA Risk SNP on ACPA-Positive RA Cases

We conducted a multi-ethnic validation study of the identified HLA-DOA risk SNP on ACPA-positive RA cases. We referred to the results of the previously conducted HLA imputation fine-mapping studies on ACPA-positive RA cases versus controls in east Asian, including Chinese and Korean (2,782 ACPA-positive RA cases and 4,315 controls),[10] and European (7,279 ACPA-positive RA cases and 15,870 controls) populations.[8] Given that the top associated HLA-DOA risk SNP (rs378352) was not available in these results, the association results of the proxy SNP (rs369150, r^2 = 0.99 in Japanese) were assessed. Trans-ethnic LD comparison between rs369150 genotypes and HLA-DRB1 four-digit alleles was performed by evaluation of the HLA imputation reference panels of the Japanese (n = 908),[20] east Asian (n = 350),[10,28] and European populations (T1DGC consortium; n = 5,225).[8]

Results

HLA Imputation of the RA GWAS Data for the Japanese Population

We performed HLA imputation for the large-scale RA GWAS data on individuals with Japanese ancestry, which included 5,119 ACPA-positive RA cases, 1,125 ACPA-negative RA cases, and 23,731 controls obtained from four datasets (BBJ1, BBJ2, IORRA, and Kyoto University; Table S1).[21] We used the population-specific HLA imputation reference panel of individuals with Japanese ancestry.[20] We assessed imputation accuracy by using separately genotyped HLA-DRB1 alleles from a subset of the subjects and confirmed genotype concordance rates of 97.5% in 4-digit alleles. We obtained imputed genotypes of 143 two-digit or four-digit alleles and 685 amino acid polymorphisms of seven classical HLA genes (HLA-A, HLA-B, and HLA-C for class I and HLA-DRB1, HLA-DQB1, HLA-DPA1, and HLA-DPB1 for class II) and 6,504 SNPs in the MHC region (from 29 Mbp to 34 Mbp at chromosome 6 [NCBI Genome browser, build 37]).

HLA-DRB1 Has the Largest Impact on ACPA-Defined RA Phenotypes

To comprehensively assess the HLA variant risk on RA, we evaluated four ACPA-defined RA phenotypes: (1) ACPA-positive RA cases versus controls, (2) ACPA-negative RA cases versus controls, (3) ACPA-positive RA cases versus ACPA-negative RA cases (i.e., ACPA positivity in overall RA cases), and (4) ACPA titer QTL in ACPA-positive RA cases. When we tested the associations of the imputed HLA variants, we found the most significant associations of HLA-DRB1 for each of the four ACPA-defined RA phenotypes (Figure 1 and Table S2). Within HLA-DRB1, amino acid polymorphisms of HLA-DRβ1 showed the most significant risk, as previously reported;[7,8,10,11] the amino acid position 11 (or 13, in strong LD with the position 11 [r^2 = 1.00 between Pro11 and Arg13 and 0.95 between Val11 and His13; Figure S1]) for ACPA-positive RA risk (p = 1.7 × 10^{-282}; Figure 1A), position 71 for ACPA-negative RA risk (p = 9.8 × 10^{-11}; Figure 1B), position 11 [or 13] for ACPA-positive versus -negative RA (p = 3.6 × 10^{-57}; Figure 1C), and position 74 for ACPA titer QTL (p = 2.3 × 10^{-45}; Figure 1D). We note that HLA-DRβ1 Ala71 is in strong LD with Pro11 and Arg13 (r^2 = 0.95; Figure S1).

When we performed a stepwise conditional analysis of HLA-DRβ1 amino acid polymorphisms, ACPA-positive RA risk at amino acid positions 74, 37, 67, and 86, and ACPA-positive versus -negative RA risk at amino acid position 74, was independently observed (p < 5.0 × 10^{-8}; Figure S2). We note that, though, parsimonious localization of causal amino acid positions is relatively difficult when multiple amino acid positions with moderate LDs are fitted in a regression model,[10] while additional accumulation of the subjects could increase statistical power to distinguish the causal position(s). No significant additional risk was
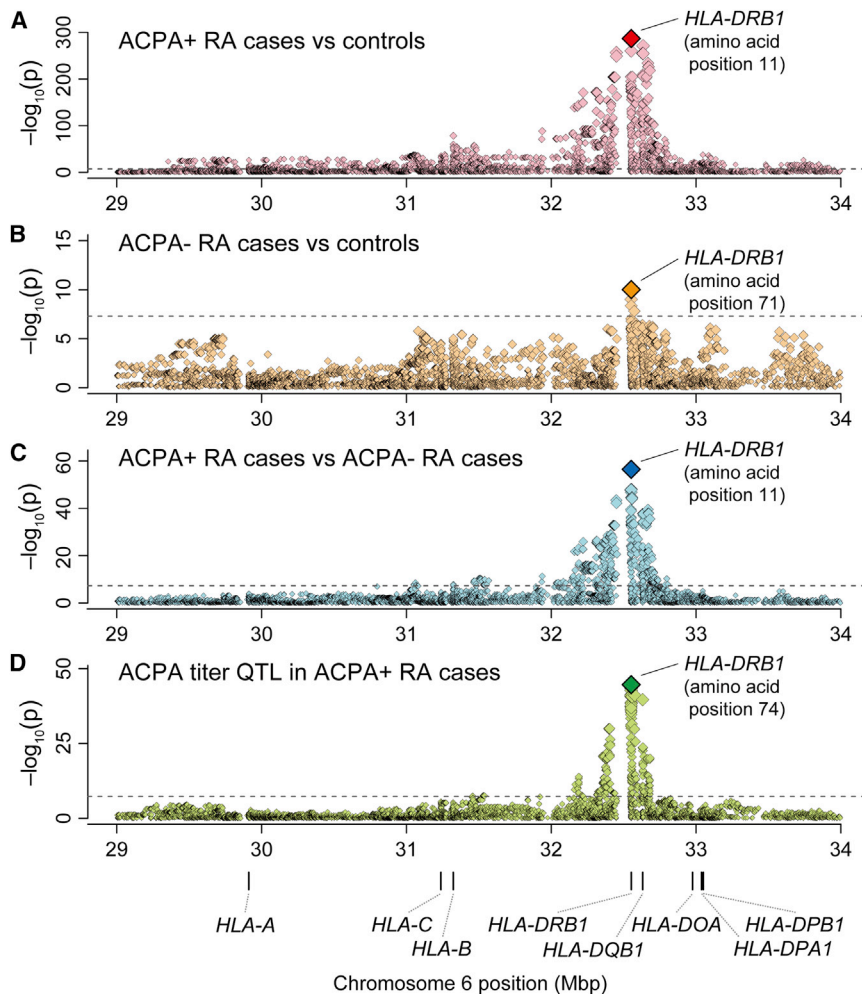
Figure 1. Regional Association Plots of HLA Variants to RA Risk in Japanese Populations

Each diamond represents $-\log_{10}(p)$ of the tested HLA variants, including SNPs, HLA alleles, and HLA amino acid polymorphisms. The dotted horizontal line represents the study-wide significance threshold of $p = 5.0 \times 10^{-8}$. The bottom part shows the physical positions of the HLA genes on chromosome 6 (NCBI Genome browser, build 37). Each panel shows an association plot of the ACPA-defined RA phenotypes: (A) ACPA-positive RA cases versus controls, (B) ACPA-negative RA cases versus controls, (C) ACPA-positive RA cases versus ACPA-negative RA cases, and (D) ACPA titer QTL in ACPA-positive RA cases.

observed for the other two ACPA-defined RA phenotypes when these phenotypes were conditioned on the top associated HLA-DRβ1 amino acid positions ($p > 5.0 \times 10^{-8}$).

Han et al. have pointed out that individuals affected with other diseases clinically indistinguishable from, or possibly misdiagnosed as, ACPA-negative RA might be embedded within ACPA-negative RA cases and thus could confound associations.[8] We note that risk HLA alleles of candidate confounding diseases in Europeans, such as HLA-B*27 for ankylosing spondylitis (AS [MIM: 106300])[29] and HLA-C*06:02 for psoriasis vulgaris (PsV [MIM: 177900]),[23] have very low frequencies in Japanese populations (= 0.002 and 0.011, respectively).[20] When we conditioned on the GRS of ACPA-positive RA to adjust for possible misdiagnosis of ACPA-positive RA cases as ACPA-negative RA cases,[8] we also found the most significant association of ACPA-negative RA at the HLA-DRβ1 amino acid position 71 ($p = 6.7 \times 10^{-8}$).

Independent Risk of Multiple Classical and Non-classical HLA Genes on ACPA-Positive RA

We then investigated ACPA-defined RA phenotype risk of the other HLA genes independently of *HLA-DRB1* (Figure 2 and Figure S3). When we conditioned on the *HLA-DRB1*

gene variants, we observed a significant independent association of ACPA-positive RA at HLA-DPβ1 amino acid position 84 ($p = 2.2 \times 10^{-33}$; Figure 2B). When we conditioned on *HLA-DRB1* and *HLA-DPB1*, we observed a significant independent association at a synonymous coding SNP of *HLA-DOA*, one of the non-classical HLA genes (rs378352, $p = 1.4 \times 10^{-9}$; Figure 2C). When we conditioned on *HLA-DRB1*, *HLA-DPB1*, and *HLA-DOA*, we observed a significant independent association at HLA-B*40:02 ($p = 2.8 \times 10^{-9}$; Figure 2D), which is known as a risk mutation allele in AS[29] and a protective allele in acquired aplastic anemia (MIM: 609135).[30] No independent association was observed when we conditioned on these risk HLA genes ($p > 5.0 \times 10^{-8}$; Figure 2E). A stepwise regression analysis for each of the risk HLA genes did not identify additional independent risk variants (data not shown). Observed ACPA-positive RA risk HLA gene combinations were generally consistent with those previously reported by large-scale European studies,[7,8] providing additional findings of independent ACPA-positive RA risk at the *HLA-DOA* SNP (discussed later). For the other three ACPA-defined RA comparison patterns, no independent association signal was observed when we conditioned on *HLA-DRB1* (Figure S3).

To comprehensively evaluate the contribution of the multiple classical and non-classical HLA genes, we conducted a multivariate regression analysis incorporating the associated HLA variants (Table S3). Increased risk on ACPA-positive RA (odds ratio [OR] = 8.00, 95% confidence interval [95% CI] = 6.18–10.4; $p = 2.6 \times 10^{-56}$) and ACPA-positive versus -negative RA (OR = 3.38, 95% CI = 2.75–4.17; $p = 2.3 \times 10^{-30}$) were observed at HLA-DRβ1 Asp11, the amino acid polymorphism corresponding to the Asian-specific RA risk allele of HLA-DRB1* 09:01.[9,17] HLA-DRβ1 Glu71 demonstrated a protective
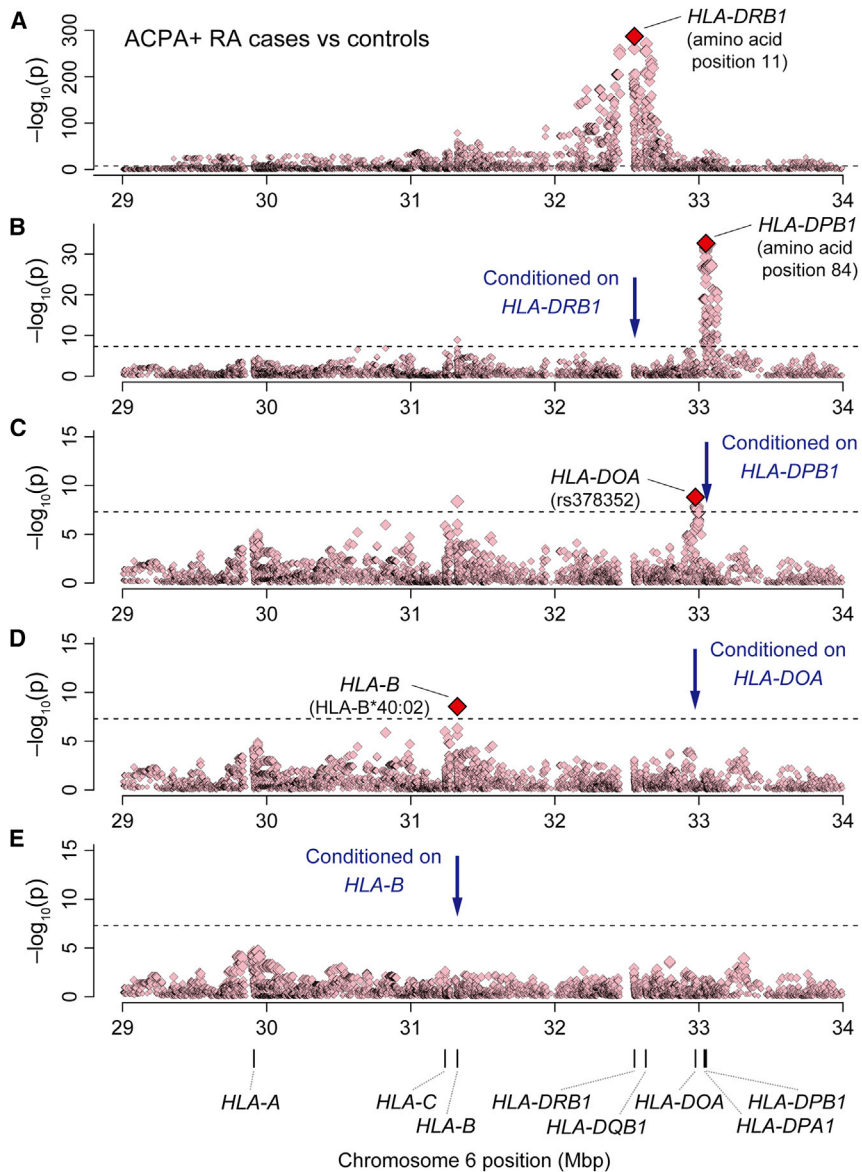
Figure 2. Stepwise Conditional Association Plots of the HLA Variants with ACPA-Positive RA Risk

Each diamond represents $-\log_{10}(p)$ of the tested HLA variants, including the SNPs, HLA alleles, and HLA amino acid polymorphisms. The dotted horizontal line represents the study-wide significance threshold of $p = 5.0 \times 10^{-8}$. The bottom part shows the physical positions of the HLA genes on chromosome 6 (NCBI Genome browser, build 37). Each panel shows the stepwise conditional association plot with ACPA-positive RA risk.

homozygous; $p = 8.7 \times 10^{-9}$ for non-additive effect; Table S4).

Dosage Contribution of *HLA-DOA* to ACPA-Positive RA Risk

Whereas previous MHC fine-mapping studies have pointed at disease risk associated with amino acid polymorphisms in classical HLA genes,[7,8,10,14,20,23] our study reveals ACPA-positive RA risk independently associated with a synonymous mutation (rs378352) in the non-classical HLA gene of *HLA-DOA* (OR = 1.20, 95% CI = 1.13–1.28; $p = 1.4 \times 10^{-9}$; conditioned on the nearby RA-risk HLA genes *HLA-DRB1* and *HLA-DPB1*; Table 1). The *HLA-DOA* risk SNP also demonstrated independent association when we conditioned on all the imputed classical HLA genes ($p = 7.4 \times 10^{-9}$). To rule out false-positive results due to inaccurate HLA imputation, we performed a separate association analysis using directly genotyped data of the *HLA-DOA* SNP, four-digit *HLA-DRB1* alleles, and a *HLA-DPB1* proxy SNP (rs9277396 for the amino acid position 84; $r^2 = 0.99$ in Japanese), and confirmed significant independent risk of the *HLA-DOA* SNP (OR = 1.27; $p = 0.0048$; 1,710 ACPA-positive RA cases from BBJ1 and BBJ2 and 877 controls from the reference panel). We note that our data did not indicate an independent RA risk of a HLA class II SNP reported by a previous RA GWAS in a Chinese population (rs7765379; $r^2 = 0.89$ with an originally reported risk SNP of rs12525220; $p = 0.83$ when conditioned on all the imputed HLA genes).[19]

To assess the functional contribution of the risk *HLA-DOA* SNP, we assessed the eQTL database of the Japanese individuals (n = 298).[25] We found that rs369150, a proxy SNP of rs378352 ($r^2 = 0.99$ in Japanese), was also one of the top associated SNPs with a cis-eQTL effect on *HLA-DOA* mRNA expression ($p = 1.2 \times 10^{-7}$; Figure 3). The ACPA-positive RA risk allele of rs369150-A (equivalent to

effect on ACPA-negative RA (OR = 0.70, 95% CI = 0.58–0.84; $p = 9.4 \times 10^{-5}$), and HLA-DRβ1 Glu74 showed a reducing effect on ACPA titer (β = −0.220, SE = 0.016; $p = 2.0 \times 10^{-44}$). GRSs of ACPA-defined RA phenotypes showed almost no correlation between ACPA-negative RA risk and ACPA-positive versus -negative RA ($r = 0.03$) and negative correlation between ACPA-negative RA risk and ACPA titer QTL ($r = -0.48$), whereas other phenotype pairs showed positive correlations ($r = 0.22$–0.94; Figure S4).

Variants of the identified risk HLA genes explained 9.2% (6.4% for *HLA-DRB1* and 2.8% for other HLA genes), 1.5%, 4.0%, and 5.0% of the phenotypic variance of ACPA-positive RA, ACPA-negative RA, ACPA-positive versus ACPA-negative, and ACPA titer, respectively (Figure S5). Applying a multivariate regression model fitting non-additive effects of the HLA alleles, we found a significant non-additive effect of the HLA-DRB1*04:05 allele that relatively increased homozygous risk (OR = 3.74 for heterozygous and 25.2 for

Table 1. Association of a *HLA-DOA* SNP with ACPA-Positive RA Risk

SNP	Position[a]	Gene	Reference/ Alternative	Population	Cohort	No. Subjects		Reference Allele Frequency		OR (95% CI)[b]	p Value[b]
						ACPA + RA Cases	Controls	ACPA + RA Cases	Controls		
rs378352	32,974,934	*HLA-DOA*	A/G	Japanese	BBJ1	1,820	14,209	0.279	0.258	1.24 (1.13–1.36)	4.6×10^{-6}
					BBJ2	710	3,105	0.263	0.261	1.13 (0.96–1.33)	0.15
					IORRA	1,955	5,252	0.266	0.258	1.16 (1.04–1.29)	0.0053
					Kyoto	634	1,165	0.297	0.245	1.27 (1.03–1.58)	0.027
					Meta-analysis	5,119	23,731	0.274	0.258	1.20 (1.13–1.28)	1.4×10^{-9}
rs369150[c]	32,975,720	*HLA-DOA*	A/G	east Asian[d]	–	2,782	4,315	0.342	0.310	1.15 (1.05–1.27)	0.0040
				European[d]	–	7,279	15,870	0.300	0.260	1.06 (1.01–1.12)	0.031

ACPA, anti-citrullinated protein autoantibody; RA, rheumatoid arthritis; OR, odds ratio; 95% CI, 95% confidence interval.
[a]NCBI Genome browser, build 37.
[b]Conditioned on *HLA-DRB1* and *HLA-DPB1*.
[c]A proxy SNP of rs378352 ($r^2 = 0.99$ in Japanese).
[d]Obtained from previous reports.[8,10]

rs378352-A) reduced expression of *HLA-DOA*. rs369150-A was also the top cis-QTL allele that decreased *HLA-DOA* expression in Europeans (n = 2,775; p = 1.4×10^{-161}).[26]

Cell-type-specific expression profiles indicated relatively high expression levels of *HLA-DOA* in immune-related cells (Figure S6).[27]
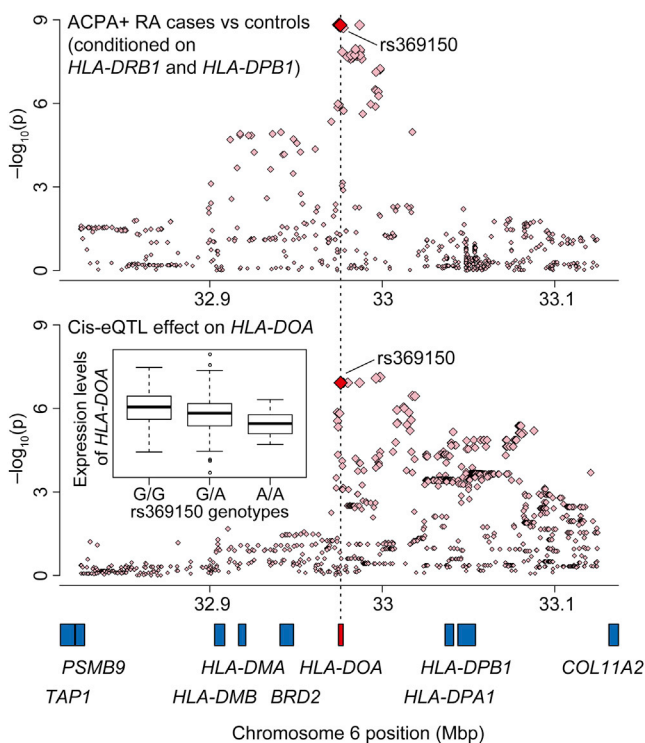
A Multi-ethnic Validation Study of the *HLA-DOA* SNP Risk

To further confirm *HLA-DOA* SNP risk, we conducted a multi-ethnic validation study by using previous results in east Asian (n = 7,097)[10] and European (n = 23,149) populations.[8] rs369150 showed significant independent risk in both populations, with similar OR in east Asians (OR = 1.15, 95% CI = 1.05-1.27; p = 0.0040) but weaker OR in Europeans (OR = 1.06, 95% CI = 1.01–1.12; p = 0.031) as compared to the Japanese population.

To dissect observed risk heterogeneity among populations, we assessed a trans-ethnic LD comparison of the *HLA-DOA* SNP and *HLA-DRB1* alleles (Figure 4), using HLA reference panels of Japanese (n = 908),[20] east Asian (n = 350),[10,28] and European (n = 5,225) populations.[8] Notably, in Europeans, the strongest ACPA-positive RA risk allele of HLA-DRB1*04:01 (OR = 3.81) was in the highest positive LD with the *HLA-DOA* SNP risk allele (r = 0.21), which could confound the spuriously reduced *HLA-DOA* SNP OR after conditioning on *HLA-DRB1*. In contrast, in the Japanese population, the *HLA-DOA* SNP risk allele was in weak LD (−0.1 < r < 0.1) with any of the moderate or strong *HLA-DRB1* risk alleles (OR < 0.50 or OR > 1.20), which suggests that the *HLA-DOA* SNP OR could be more precisely observed in Japanese individuals when conditioning on *HLA-DRB1*. Moderate LD was observed in east Asians, as an intermediate between the Europeans and Japanese. These results suggest that population-specific LD structures of the HLA variants could explain observed risk heterogeneity among ethnicities and thus could provide empirical evidence of the *HLA-DOA* SNP risk on ACPA-positive RA.
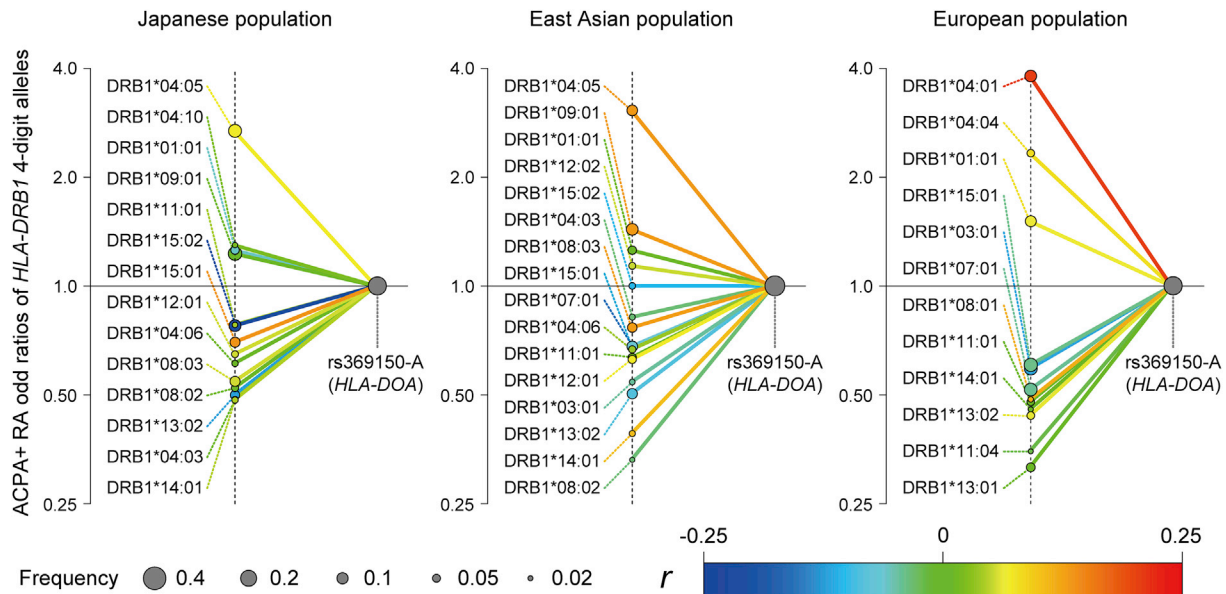

Figure 3. Regional Association Plots at *HLA-DOA* with ACPA-Positive RA Risk
Each diamond represents –log$_{10}$(p) of the tested SNPs at the *HLA-DOA* locus. The upper panel shows the SNP associations of ACPA-positive RA risk (conditioned on *HLA-DRB1* and *HLA-DPB1*). The middle panel shows SNP associations in the cis-eQTL study on *HLA-DOA* expression in Japanese populations.[25] The sub-panel shows a boxplot of the *HLA-DOA* expression based on genotypes of rs369150.[25] The bottom part shows the physical positions of the local genes (NCBI Genome browser, build 37).

Figure 4. Trans-ethnic Comparison of LD between *HLA-DRB1* and *HLA-DOA*
Two dimensional visualization of LD structures between the *HLA-DRB1* four-digit alleles and the *HLA-DOA* SNP risk allele in Japanese (left),[20] east Asian (middle),[10,28] and European (right) populations.[8] Each point represents the respective HLA alleles or SNP allele, and the size of the point corresponds to its frequency as indicated in the legend. Points with frequencies < 0.02 are not shown. Vertical positions of the points for *HLA-DRB1* alleles represent their binominal OR for ACPA-positive RA risk. Colors of the points for *HLA-DRB1* alleles and the segments connecting to the points correspond to the LD values of *r* as indicated in the legend.

Discussion

Fine-mapping of the MHC region and identification of causal HLA variants should enhance our genetic and biological knowledge about immune-related human diseases. Given their relatively strong effect sizes, the identified causal HLA variants could contribute to integration of human genetics into personalized medicine.[31,32]

In this study, we conducted a large-scale MHC fine-mapping study of ACPA-defined RA phenotype risks in Japanese individuals. In addition to validation of previously reported contributions of multiple classical HLA genes, most evidently exemplified by *HLA-DRB1*, we identified an independent risk of *HLA-DOA*, one of the non-classical HLA genes, on ACPA-positive RA risk as well as a cis-eQTL effect of the causal variant. Although most of the previous HLA fine-mapping studies have identified contributions of amino acid polymorphisms of classical HLA genes,[7,8,10,14,20,23] ours reports dosage contribution of a non-classical gene to immune-related human diseases.

Non-classical HLA genes, including *HLA-DOA*, are generally less polymorphic in amino acid sequences than classical HLA genes.[33] Instead, dosages of non-classical HLA molecules, such as HLA-DO (a heterodimer of HLA-DOA and HLA-DOB) and HLA-DM, could have regulatory effects on antigen presentation and cross-restriction of class II classical HLA molecules, including HLA-DR.[34] HLA-DO is known to have suppressive dosage effects on autoimmunity.[3,12,34] Overexpression of HLA-DO (namely, HLA-DOA and HLA-DOB) protects against autoimmune diabetes in non-obese diabetic (NOD) mice,[35] whereas HLA-DO defi-

cient mice develop autoantibody production.[36] Our results regarding the suppressive dosage effect of the RA-risk *HLA-DOA* allele on *HLA-DOA* expression might be consistent with these biological findings and suggest a potential functional mechanism for the observed dosage contribution of *HLA-DOA* to ACPA-positive RA etiology.

Our study provides an empirical example in which population-specific LD structure of the HLA variants could cause observed risk heterogeneity among populations when conditioned on variants with strong effect sizes, such as *HLA-DRB1* for RA risk. This result suggests the value of incorporating additional ethnic populations in MHC fine-mapping to further reveal embedded risk that is difficult to identify in a limited range of ancestral lineages.

Although our study clearly points to independent risk of one of the non-classical HLA genes, a larger number of additional HLA-related or immune-related genes are contained in the MHC region, and some of them should have impacts on disease etiologies.[17,18] To comprehensively assess the genetic risk of these unassessed variants, further construction of an HLA imputation reference panel covering a larger variety of gene polymorphisms is warranted. We also note that the SNPs included in the current HLA imputation reference panels are obtained by microarray genotyping, and thus, not all the variants in the MHC region were assessed. To this end, application of high-throughput next-generation sequencing of the MHC genes and region would be useful.[37]

In summary, we fine-mapped ACPA-defined RA risk in the MHC by applying HLA imputation to large-scale RA GWAS data from a Japanese population. Our study

identified an independent risk of *HLA-DOA* on ACPA-positive RA, suggesting a dosage contribution of this non-classical HLA gene to the disease etiology. Our study contributes to understanding HLA immunology in human diseases and suggests the incorporation of additional ancestry in MHC fine-mapping.

Supplemental Data

Supplemental Data include six figures and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.06.019.

Acknowledgments

Web Resources

Blood eQTL Browser, http://genenetwork.nl/bloodeqtlbrowser/
Broad Institute, SNP2HLA: imputation of amino acid polymorphisms in human leukocyte antigens (east Asian and European HLA imputation reference panels), http://www.broadinstitute.org/mpg/snp2hla/
FANTOM Consortium, http://fantom.gsc.riken.jp/
GenBank, http://www.ncbi.nlm.nih.gov/genbank/
Human Genetic Variation Database (HGVD), http://www.genome.med.kyoto-u.ac.jp/SnpDB/
JGA Meta Viewer, an imputation reference panel of HLA variants in Japanese, https://ddbj.nig.ac.jp/jga/viewer/view/study/JGAS00000000018
NCBI Genome, http://www.ncbi.nlm.nih.gov/genome/
OMIM, http://www.omim.org/

References

1. Neogi, T., Aletaha, D., Silman, A.J., Naden, R.L., Felson, D.T., Aggarwal, R., Bingham, C.O., 3rd, Birnbaum, N.S., Burmester, G.R., Bykerk, V.P., et al.; American College of Rheumatology; European League Against Rheumatism (2010). The 2010 American College of Rheumatology/European League Against Rheumatism classification criteria for rheumatoid arthritis: Phase 2 methodological report. Arthritis Rheum. *62*, 2582–2591.

2. Yamamoto, K., Okada, Y., Suzuki, A., and Kochi, Y. (2015). Genetic studies of rheumatoid arthritis. Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci. *91*, 410–422.

3. Terao, C., Ikari, K., Nakayamada, S., Takahashi, Y., Yamada, R., Ohmura, K., Hashimoto, M., Furu, M., Ito, H., Fujii, T., et al. (2016). A twin study of rheumatoid arthritis in the Japanese population. Mod. Rheumatol. *4*, 1–5.

4. Gregersen, P.K., Silver, J., and Winchester, R.J. (1987). The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum. *30*, 1205–1213.

5. Okada, Y., Suzuki, A., Yamada, R., Kochi, Y., Shimane, K., Myouzen, K., Kubo, M., Nakamura, Y., and Yamamoto, K. (2010). HLA-DRB1*0901 lowers anti-cyclic citrullinated peptide antibody levels in Japanese patients with rheumatoid arthritis. Ann. Rheum. Dis. *69*, 1569–1570.

6. Terao, C., Ohmura, K., Kochi, Y., Ikari, K., Maruya, E., Katayama, M., Shimada, K., Murasawa, A., Honjo, S., Takasugi, K., et al. (2011). A large-scale association study identified multiple HLA-DRB1 alleles associated with ACPA-negative rheumatoid arthritis in Japanese subjects. Ann. Rheum. Dis. *70*, 2134–2139.

7. Raychaudhuri, S., Sandor, C., Stahl, E.A., Freudenberg, J., Lee, H.S., Jia, X., Alfredsson, L., Padyukov, L., Klareskog, L., Worthington, J., et al. (2012). Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat. Genet. *44*, 291–296.

8. Han, B., Diogo, D., Eyre, S., Kallberg, H., Zhernakova, A., Bowes, J., Padyukov, L., Okada, Y., González-Gay, M.A., Rantapää-Dahlqvist, S., et al. (2014). Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. Am. J. Hum. Genet. *94*, 522–532.

9. Shimane, K., Kochi, Y., Suzuki, A., Okada, Y., Ishii, T., Horita, T., Saito, K., Okamoto, A., Nishimoto, N., Myouzen, K., et al. (2013). An association analysis of HLA-DRB1 with systemic lupus erythematosus and rheumatoid arthritis in a Japanese population: effects of *09:01 allele on disease phenotypes. Rheumatology (Oxford) *52*, 1172–1182.

10. Okada, Y., Kim, K., Han, B., Pillai, N.E., Ong, R.T.H., Saw, W.Y., Luo, M., Jiang, L., Yin, J., Bang, S.Y., et al. (2014). Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. Hum. Mol. Genet. *23*, 6916–6926.

11. Terao, C., Suzuki, A., Ikari, K., Kochi, Y., Ohmura, K., Katayama, M., Nakabo, S., Yamamoto, N., Suzuki, T., Iwamoto, T., et al. (2015). An association between amino acid position 74 of HLA-DRB1 and anti-citrullinated protein antibody levels in Japanese patients with anti-citrullinated protein antibody-positive rheumatoid arthritis. Arthritis Rheumatol. *67*, 2038–2045.

12. Terao, C., Ikari, K., Ohmura, K., Suzuki, T., Iwamoto, T., Takasugi, K., Saji, H., Taniguchi, A., Momohara, S., Yamanaka, H., et al. (2012). Quantitative effect of HLA-DRB1 alleles to ACPA levels in Japanese rheumatoid arthritis: no strong genetic impact of shared epitope to ACPA levels after stratification of HLA-DRB1*09:01. Ann. Rheum. Dis. *71*, 1095–1097.

13. Okada, Y., Yamada, R., Suzuki, A., Kochi, Y., Shimane, K., Myouzen, K., Kubo, M., Nakamura, Y., and Yamamoto, K. (2009). Contribution of a haplotype in the HLA region to anti-cyclic citrullinated peptide antibody positivity in rheumatoid arthritis, independently of HLA-DRB1. Arthritis Rheum. *60*, 3582–3590.

14. The International HIV Controllers Study (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science 330, 1551–1557.

15. Jia, X., Han, B., Onengut-Gumuscu, S., Chen, W.M., Concannon, P.J., Rich, S.S., Raychaudhuri, S., and de Bakker, P.I. (2013). Imputing amino acid polymorphisms in human leukocyte antigens. PLoS ONE 8, e64683.

16. The MHC sequencing consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. Nature 401, 921–923.

17. Kochi, Y., Yamada, R., Kobayashi, K., Takahashi, A., Suzuki, A., Sekine, A., Mabuchi, A., Akiyama, F., Tsunoda, T., Nakamura, Y., and Yamamoto, K. (2004). Analysis of single-nucleotide polymorphisms in Japanese rheumatoid arthritis patients shows additional susceptibility markers besides the classic shared epitope susceptibility sequences. Arthritis Rheum. 50, 63–71.

18. Mitsunaga, S., Hosomichi, K., Okudaira, Y., Nakaoka, H., Kunii, N., Suzuki, Y., Kuwana, M., Sato, S., Kaneko, Y., Homma, Y., et al. (2013). Exome sequencing identifies novel rheumatoid arthritis-susceptible variants in the BTNL2. J. Hum. Genet. 58, 210–215.

19. Jiang, L., Yin, J., Ye, L., Yang, J., Hemani, G., Liu, A.J., Zou, H., He, D., Sun, L., Zeng, X., et al. (2014). Novel risk loci for rheumatoid arthritis in Han Chinese and congruence with risk variants in Europeans. Arthritis Rheumatol. 66, 1121–1132.

20. Okada, Y., Momozawa, Y., Ashikawa, K., Kanai, M., Matsuda, K., Kamatani, Y., Takahashi, A., and Kubo, M. (2015). Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. Nat. Genet. 47, 798–802.

21. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al.; RACI consortium; GARNET consortium (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376–381.

22. Arnett, F.C., Edworthy, S.M., Bloch, D.A., McShane, D.J., Fries, J.F., Cooper, N.S., Healey, L.A., Kaplan, S.R., Liang, M.H., Luthra, H.S., et al. (1988). The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum. 31, 315–324.

23. Okada, Y., Han, B., Tsoi, L.C., Stuart, P.E., Ellinghaus, E., Tejasvi, T., Chandran, V., Pellett, F., Pollock, R., Bowcock, A.M., et al. (2014). Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. Am. J. Hum. Genet. 95, 162–172.

24. Lenz, T.L., Deutsch, A.J., Han, B., Hu, X., Okada, Y., Eyre, S., Knapp, M., Zhernakova, A., Huizinga, T.W., Abecasis, G., et al. (2015). Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. Nat. Genet. 47, 1085–1090.

25. Narahara, M., Higasa, K., Nakamura, S., Tabara, Y., Kawaguchi, T., Ishii, M., Matsubara, K., Matsuda, F., and Yamada, R. (2014). Large-scale East-Asian eQTL mapping reveals novel candidate genes for LD mapping and the genomic landscape of transcriptional effects of sequence variants. PLoS ONE 9, e100924.

26. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243.

27. Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al.; FANTOM Consortium and the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. Nature 507, 462–470.

28. Pillai, N.E., Okada, Y., Saw, W.Y., Ong, R.T., Wang, X., Tantoso, E., Xu, W., Peterson, T.A., Bielawny, T., Ali, M., et al. (2014). Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations. Hum. Mol. Genet. 23, 4443–4451.

29. Cortes, A., Pulit, S.L., Leo, P.J., Pointon, J.J., Robinson, P.C., Weisman, M.H., Ward, M., Gensler, L.S., Zhou, X., Garchon, H.J., et al. (2015). Major histocompatibility complex associations of ankylosing spondylitis are complex and involve further epistasis with ERAP1. Nat. Commun. 6, 7146.

30. Inaguma, Y., Akatsuka, Y., Hosokawa, K., Maruyama, H., Okamoto, A., Katagiri, T., Shiraishi, K., Murayama, Y., Tsuzuki-Iba, S., Mizutani, Y., et al. (2016). Induction of HLA-B*40:02-restricted T cells possessing cytotoxic and suppressive functions against haematopoietic progenitor cells from a patient with severe aplastic anaemia. Br. J. Haematol. 172, 131–134.

31. Mallal, S., Nolan, D., Witt, C., Masel, G., Martin, A.M., Moore, C., Sayer, D., Castley, A., Mamotte, C., Maxwell, D., et al. (2002). Association between presence of HLA-B*5701, HLA-DR7, and HLA-DQ3 and hypersensitivity to HIV-1 reverse-transcriptase inhibitor abacavir. Lancet 359, 727–732.

32. Anderson, R.P., Henry, M.J., Taylor, R., Duncan, E.L., Danoy, P., Costa, M.J., Addison, K., Tye-Din, J.A., Kotowicz, M.A., Knight, R.E., et al. (2013). A novel serogenetic approach determines the community prevalence of celiac disease and informs improved diagnostic pathways. BMC Med. 11, 188.

33. Moon, S.M., Gu, H., Ryu, H.J., Kim, J.J., Kim, H.T., Han, B.G., Kimm, K., Lee, J.K., and Oh, B. (2005). Identification of four novel HLA-DOA alleles, DOA*010106, DOA*0102, DOA*0103, and DOA*0104N, by sequence-based typing*. Tissue Antigens 66, 242–245.

34. Guce, A.I., Mortimer, S.E., Yoon, T., Painter, C.A., Jiang, W., Mellins, E.D., and Stern, L.J. (2013). HLA-DO acts as a substrate mimic to inhibit HLA-DM by a competitive mechanism. Nat. Struct. Mol. Biol. 20, 90–98.

35. Yi, W., Seth, N.P., Martillotti, T., Wucherpfennig, K.W., Sant'Angelo, D.B., and Denzin, L.K. (2010). Targeted regulation of self-peptide presentation prevents type I diabetes in mice without disrupting general immunocompetence. J. Clin. Invest. 120, 1324–1336.

36. Gu, Y., Jensen, P.E., and Chen, X. (2013). Immunodeficiency and autoimmunity in H2-O-deficient mice. J. Immunol. 190, 126–137.

37. Hosomichi, K., Shiina, T., Tajima, A., and Inoue, I. (2015). The impact of next-generation sequencing technologies on HLA research. J. Hum. Genet. 60, 665–673.