





Citation: Kim K, Bang S-Y, Yoo DH, Cho S-K, Choi C-B, Sung Y-K, et al. (2016) Imputing Variants in HLA-DR Beta Genes Reveals That HLA-DRB1 Is Solely Associated with Rheumatoid Arthritis and Systemic Lupus Erythematosus. PLoS ONE 11(2): e0150283. doi:10.1371/journal.pone.0150283

Editor: Jianming Tang, University of Alabama at Birmingham, UNITED STATES

Received: December 17, 2015

Accepted: February 11, 2016

Published: February 26, 2016

Copyright: © 2016 Kim et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1C1A1A02036527 to Dr. Kwangwoo Kim), the US National Institutes of Health (R01MD007909 and R01AR060366 to Dr. Swapan K. Nath), and the Korea Healthcare Technology R&D Project of the Ministry for Health & Welfare (HI13C2124 to Dr. Sang-Cheol Bae). The funders had no role in study design, data collection

RESEARCH ARTICLE

Imputing Variants in HLA-DR Beta Genes Reveals That HLA-DRB1 Is Solely Associated with Rheumatoid Arthritis and Systemic Lupus Erythematosus

Kwangwoo Kim¹, So-Young Bang¹, Dae Hyun Yoo¹, Soo-Kyung Cho¹, Chan-Bum Choi¹, Yoon-Kyoung Sung¹, Tae-Hwan Kim¹, Jae-Bum Jun¹, Young Mo Kang², Chang-Hee Suh³, Seung-Cheol Shim⁴, Shin-Seok Lee⁵, Jisoo Lee⁶, Won Tae Chung⁷, Seong-Kyu Kim⁸, Jung-Yoon Choe⁸, Swapan K. Nath⁹, Hye-Soon Lee¹, Sang-Cheol Bae¹*

1 Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Republic of Korea, 2 Division of Rheumatology, Department of Internal Medicine, Kyungpook National University School of Medicine, Daegu, Republic of Korea, 3 Department of Rheumatology, Ajou University School of Medicine, Suwon, Republic of Korea, 4 Division of Rheumatology, Daejeon Rheumatoid & Degenerative Arthritis Center, Chungnam National University Hospital, Daejeon, Republic of Korea, 5 Division of Rheumatology, Department of Internal Medicine, Chonnam National University Medical School and Hospital, Gwangju, Republic of Korea, 6 Division of Rheumatology, Department of Internal Medicine, Ewha Womans University School of Medicine, Seoul, Republic of Korea, 7 Division of Rheumatology, Department of Internal Medicine, Dong-A University, Busan, Republic of Korea, 8 Division of Rheumatology, Department of Internal Medicine, Arthritis & Autoimmunity Research Center, Catholic University of Daegu School of Medicine, Daegu, Republic of Korea, 9 Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma, United States of America

* scbae@hanyang.ac.kr

Abstract

The genetic association of *HLA-DRB1* with rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) is well documented, but association with other HLA-DR beta genes (*HLA-DRB3*, *HLA-DRB4* and *HLA-DRB5*) has not been thoroughly studied, despite their similar functions and chromosomal positions. We examined variants in all functional HLA-DR beta genes in RA and SLE patients and controls, down to the amino-acid level, to better understand disease association with the HLA-DR locus. To this end, we improved an existing HLA reference panel to impute variants in all protein-coding HLA-DR beta genes. Using the reference panel, HLA variants were inferred from high-density SNP data of 9,271 RA-control subjects and 5,342 SLE-control subjects. Disease association tests were performed by logistic regression and log-likelihood ratio tests. After imputation using the newly constructed HLA reference panel and statistical analysis, we observed that *HLA-DRB1* variants better accounted for the association between MHC and susceptibility to RA and SLE than did the other three HLA-DRB variants. Moreover, there were no secondary effects in *HLA-DRB3*, *HLA-DRB4*, or *HLA-DRB5* in RA or SLE. Of all the HLA-DR beta chain paralogs, those encoded by *HLA-DRB1* solely or dominantly influence susceptibility to RA and SLE.



and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

HLA-DR is a key molecule implicated in conferring risk for rheumatoid arthritis (RA), systemic lupus erythematosus (SLE) and other diseases. HLA-DR is present on the surface of antigen-presenting cells as heterodimers consisting of an alpha chain (HLA-DR α ; encoded by *HLA-DRA*) and a beta chain (HLA-DR β ; encoded by *HLA-DRB1*, *HLA-DRB3*, *HLA-DRB4*, or *HLA-DRB5*) [1]. HLA-DR β has variable coding variations especially in its peptide-binding groove, in contrast to the beta chain proteins [1].

Recently, genetic studies have fine-mapped the primary association within the major histocompatibility complex (MHC) locus with RA and SLE to HLA-DRB1, and further narrowed it down to specific amino-acid positions [2–4]. However, these studies did not investigate the other functional HLA-DR beta genes (HLA-DRB3, HLA-DRB4, or HLA-DRB5) due to the lack of a reference panel suitable for imputing their genetic variants, although all the HLA-DRB genes are in strong linkage disequilibrium and encode beta chains functionally the same as HLA-DRB1. It is thus very important to examine the associations of all the HLA-DRB genes simultaneously with HLA-DRB1- associated diseases.

Here, we constructed an HLA reference panel [5] to impute all functional HLA-DRB genes and dissected their associations with RA and SLE to better understand HLA-DR association with these diseases.

Materials and Methods

Genotyping *HLA-DRB3*, *HLA-DRB4*, and *HLA-DRB5* to construct an HLA reference panel

413 unrelated Korean individuals were genotyped for four-digit classical HLA alleles and for copy numbers of *HLA-DRB3*, *HLA-DRB4*, and *HLA-DRB5* using a Roche GS 454 sequencing system at the Institute for Immunology and Infectious Diseases (IIID; Murdoch WA, Australia) and IIID's institution-specific calling algorithms that were accredited by the American Society for Histocompatibility and Immunogenetics (ASHI). These 413 individuals are the same individuals who were used in a previous Korean HLA reference panel [5]. The analysis was approved by the Institutional Review Board of Hanyang University, and written consent was obtained from the participants.

Constructing a Korean HLA reference panel for imputation

The previous Korean HLA reference panel had haplotype-level data of 2- and 4-digit classical alleles and amino acid residues of 6 HLA genes: *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DPB1* and *HLA-DQB1*, from 413 unrelated Korean individuals [5]. In this study, we additionally merged the data for copy number, classical allele, and amino-acid residue of *HLA-DRB3*, *HLA-DRB4*, and *HLA-DRB5* of the same 413 Korean subjects with the existing data in the previous HLA reference panel [5]. Amino-acid residues were defined based on the amino-acid sequence of 4-digit alleles in the IMGT/HLA database [6]. In the panel, a bi-allelic marker was encoded as allele *A/B*, and a multi-allelic marker was encoded as presence/absence for each allele of the marker. The HLA-DR locus houses zero or one gene copy of *HLA-DRB3*, *HLA-DRB4*, and *HLA-DRB5*, which allowed us to encode the information on copy number as presence/absence of each gene in a haplotype.

The new HLA imputation reference panel was constructed by phasing 5,858 MHC SNPs and amino acid residues, 2-digit and 4-digit HLA alleles, and copy number of 9 HLA genes using the Beagle 3.0.4 imputation program [7] powered by the SNP2HLA method [8] with some modifications.



Imputing HLA variants in Korean case-control cohorts

We extracted MHC SNPs with minor allele frequency $\geq 1\%$ from previous genome-wide and Immunochip SNP datasets [4,9] of Korean case-control cohorts for RA (n = 9,271; 2,234 cases and 7,036 controls; 2 independent cohorts) and SLE (n = 5,342; 849 cases and 4,493 controls; 1 cohort). Information on these cohorts and data has been presented in previous reports [4,9]. The study involving case-control participants was approved by the Institutional Review Board of Hanyang University, and written consent was obtained from the participants. Datasets for each cohort were used to impute HLA variants by SNP2HLA [8] and the new Korean HLA reference panel. Imputed markers with minor allele frequency $\geq 1\%$ and imputation quality (PLINK INFO) ≥ 0.8 were used in disease association tests.

Testing disease association

Disease association at each imputed marker was tested by logistic regression, adjusting the top 10 principal components (PCs) calculated from genome-wide SNP data [4,9]. The disease association of each amino-acid position with multiple residues was examined by log-likelihood ratio tests [4,9]. The null logistic regression model included only the top 10 PCs as predictors, and the full model additionally included the dosage of the tested markers, excluding the most frequent allele. In addition, because the datasets for RA were separately imputed from two different array datasets (a genome-wide SNP array and an immunochip array dataset), we used a dummy variable indicating datasets in fitting a logistic regression model.

Results and Discussion

We improved an existing HLA reference panel [5] to additionally impute copy numbers (0 to 2 copies for each gene), classical alleles, and amino-acid residues of *HLA-DRB3*, *HLA-DRB4* and *HLA-DRB5*, as well as variants in pre-existing HLA target genes. In brief, the three HLA-DRB genes were genotyped by sequencing 413 Korean subjects who were used in a previous Korean HLA reference panel [5]. Long-range haplotypes were then constructed by phasing the HLA-DRB variants with all pre-existing variants of SNP and HLA (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DPB1*, and HLA-*DQB1*).

Imputation accuracy in imputation using the new HLA reference panel was examined by cross-validation comparing the imputed and actual genotypes of classical alleles of HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DPB1, HLA-DQB1, HLA-DRB3, HLA-DRB4 and HLA-DRB5, as described following. The Korean HLA reference panel subjects (n = 413) were randomly divided into 10 almost equal sized subgroups ($n\approx41$). Of the 10 subgroups, a single subgroup ($n\approx41$) is retained as the test group to be imputed for HLA variants, and the remaining 9 subgroups (n≈372) are used for an HLA reference panel. Therefore, each of the 10 subgroups used exactly once as the test data in the cross-validation process. The classical HLA alleles in a test group were then imputed from MHC SNPs in the test group using SNP2HLA and the reference panel that was constructed from the matched reference group. Average concordance rates were calculated from concordance rates between imputed and actual 2-digit and 4-digit classical alleles (including deleted alleles in HLA-DRB3, HLA-DRB4, and HLA-DRB5) of each HLA gene. The average concordance rates between the best-guess imputed and actual 4-digit alleles of HLA-DRB3, HLA-DRB4, and HLA-DRB5 were 91.7 to 98.4% (Table 1). The imputation accuracy for HLA-DRB1 was 89.6% at 2-digit resolution and 81.6% at 4-digit resolution, which was similar to that of previous Asian panels [3,5]. In addition, we checked the correlation of imputed dosage (0 to 2) with actual dosage (0, 1, or 2) for each HLA allele. For the preexisting HLA variants in the original reference panel, a high correlation between imputed and actual dosage was observed (average Pearson's correlation coefficient (r) = 0.909 for alleles



Table 1. Concordance rate between imputed and actual allele of HLA genesa.

Allelic resolution	esolution HLA genes									
	A	В	С	DRB1	DPB1	DQB1	DRB3	DRB4	DRB5	Total
2-digit	0.970	0.916	0.970	0.892	0.966	0.954	0.919	0.959	0.979	0.947
4-digit	0.913	0.867	0.930	0.815	0.949	0.908	0.918	0.958	0.979	0.915

a Imputation accuracy was examined by cross-validation comparing the imputed and actual genotypes of classical alleles of HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB1, HLA-DRB3, HLA-DRB4 and HLA-DRB5, as described following. The Korean HLA reference panel subjects (n = 413) were randomly divided into 10 almost equal sized subgroups (n≈41). Of the 10 subgroups, a single subgroup (n≈41) is retained as the test group to be imputed for HLA variants, and the remaining 9 subgroups (n≈372) are used for an HLA reference panel. The classical HLA alleles in a test group were then imputed from MHC SNPs in the test group using SNP2HLA and the reference panel that was constructed from the matched reference group. Average concordance rates were calculated from concordance rates between imputed and actual 2-digit and 4-digit classical alleles (including deleted alleles in HLA-DRB3, HLA-DRB4, and HLA-DRB5) of each HLA gene.

doi:10.1371/journal.pone.0150283.t001

with frequency \geq 0.01), which was consistent with that of the original Korean HLA reference panel (r = 0.887 for alleles with frequency \geq 0.01) [5]. Our new reference panel also showed good correlation coefficients for the alleles of *HLA-DRB3*, *HLA-DRB4*, and *HLA-DRB5* (r = 0.891 for alleles with frequency \geq 0.01; S1 Fig).

The HLA reference panel is publicly available at https://sites.google.com/site/scbaehanyang/ hla panel.

We then revisited our previous SNP datasets from Korean case-control cohorts for RA (9,271; 2,234 cases and 7,036 controls) [9] and SLE (5,342; 849 cases and 4,493 controls) [4] to examine the disease associations of all functional HLA-DRB genes. After imputation and association tests using logistic regression and log-likelihood ratio tests, we identified the primary disease association to be, among the HLA-DRB genes in the extended MHC region, at *HLA-DRB1*. The most significant association was identified at the two previously reported linked amino-acid positions, 11 and 13, of HLA-DRB1 (at 11, $p = 4.69 \times 10^{-112}$ in RA and $p = 5.64 \times 10^{-17}$ in SLE) which accounted for disease risk better than any of the variants in the other HLA-DRB genes ($p \ge 3.62 \times 10^{-62}$ in RA and $p \ge 1.03 \times 10^{-13}$ in SLE) (Fig 1A and 1C). We note that the observed effects of the residues at these positions were consistent with previous reports [3,4].

To determine whether any secondary signals exist in the other HLA-DRB genes, we performed a conditional analysis controlling for HLA-DRB1 association effects at the reported disease-associated amino-acid positions (positions 11, 13, 71, and 74 in RA; 11, 13, and 26 in SLE). There was no independent association in HLA-DRB3, HLA-DRB4, or HLA-DRB5 passing the significance threshold $p < 5 \times 10^{-8}$ (Fig 1B and 1D).

In addition, we looked for haplotypic effects among *HLA-DRB* genes. It is well known that the presence of *HLA-DRB3*, *HLA-DRB4*, and *HLA-DRB5* perfectly correlates with two-digit alleles of *HLA-DRB1* [10–13]. For example, if the *HLA-DRB1* alleles *01, *08, or *10 are present, none of the other three HLA-DRB genes are found on the same chromosome phase. In contrast, *HLA-DRB3* is present whenever any of alleles *03, *11, *12, *13, or *14 of *HLA-DRB1* is. Similarly, *HLA-DRB4* is present whenever any of alleles *04, *07, or *09 of *HLA-DRB1* is, and *HLA-DRB5* is present whenever any of alleles *15 and *16 of *HLA-DRB1* is. When we obtained haplotypes of 4-digit classical alleles of the HLA-DRB genes, we were also able to observe known haplotype structures (*HLA-DRB1* alone, *HLA-DRB1+HLA-DRB3*, *HLA-DRB1+HLA-DRB1+HLA-DRB4*, and *HLA-DRB1+HLA-DRB5*). The frequencies and disease association results of each haplotype in the RA and SLE case-control cohorts are shown in <u>S1</u> and <u>S2</u> Tables. However, we could not evaluate the modifying effects of *HLA-DRB3*, *HLA-DRB4*, or *HLA-DRB5* on



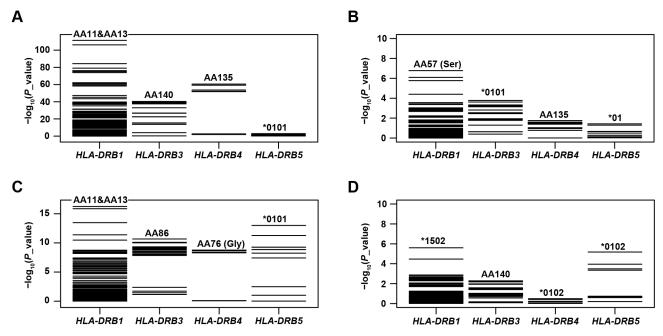


Fig 1. Distributions of p values for disease associations of HLA-DRB variants. P values for all variants of *HLA-DRB1*, *HLA-DRB3*, *HLA-DRB4* and *HLA-DRB5* were calculated by unconditional analyses testing associations with RA and SLE. The most significant association in unconditional analyses was identified at HLA-DR β 1 amino-acid position (AA) 11 and AA13 in both RA (**A**) and SLE (**C**). After conditioning on all reported disease-associated amino-acid positions of HLA-DR β 1 (11, 13, 71, and 74 in RA and 11, 13, and 26 in SLE), no additional association with p < 5 × 10⁻⁸ was identified in RA (**B**) or SLE (**D**). Variants with the lowest p value are shown for each HLA-DRB gene.

doi:10.1371/journal.pone.0150283.g001

the disease susceptibility effect of HLA-DRB1, because in most cases no classical allele of HLA-DRB1 was tightly linked with more than two classical alleles of the other HLA-DRB genes ($\underline{S1}$ and $\underline{S2}$ Tables).

The exclusive association of HLA-DRB1 among the HLA-DRB genes with RA and SLE may reflect an important role of structural variations in the HLA-DRB1 epitope-binding site in the recognition of autoantigens in RA and SLE. Alternatively, pathogenic effects of HLA-DRB3, HLA-DRB4, or HLA-DRB5 could be very small, perhaps due to their relatively low expression or weak function (for example, low affinity to HLA-DRA), which might result in poor representation of disease-risk alleles in patients with RA or SLE and thus low statistical power to detect disease association of such HLA-DRB genes. HLA-DRB3 and HLA-DRB4 are in fact expressed to a much lower degree than is HLA-DRB1, although HLA-DRB5 is highly expressed [14-16].

Conclusion

Previously, comprehensive research on all the functional HLA-DRB genes was lacking, or limited by the high cost of genotyping (resulting in small study sizes) [13,14] and the lack of imputation methods for *HLA-DRB3*, *HLA-DRB4*, and *HLA-DRB5* [2–4], despite interest in the association of HLA-DR with RA and SLE. In the present study, we constructed an HLA reference panel to impute all functional HLA-DRB genes as well as five other MHC class I and II HLA genes. By applying this panel to large case-control cohorts for RA and SLE, we revealed that the risk effects at *HLA-DRB3*, *HLA-DRB4*, and *HLA-DRB5* were neither superior to nor independent of the HLA-DRβ1 amino-acid model in RA and SLE.



In summary, our results support the association of *HLA-DRB1* with RA and SLE, and provide a more complete picture to better understand a source of disease association in the HLA-DR locus.

Supporting Information

S1 Fig. Dosage correlation between imputed and genotyped alleles. Imputed dosages (0 to 2) of 2-digit alleles (red), 4-digit alleles (green), and copy number of HLA-DRB3, HLA-DRB4, HLA-DRB5 were compared with the actual dosage (0, 1 or 2). The correlation coefficient between the imputed and actual dosages of each allele with allele frequency \geq 0.01 was plotted according to its allele frequency. (TIF)

S1 Table. Results for association of *HLA-DRB* haplotypes with RA susceptibility. (PDF)

S2 Table. Results for association of *HLA-DRB* haplotypes with SLE susceptibility. (PDF)

Acknowledgments

This study was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1C1A1A02036527 to Dr. Kwangwoo Kim), the US National Institutes of Health (R01MD007909 and R01AR060366 to Dr. Swapan K. Nath), and the Korea Healthcare Technology R&D Project of the Ministry for Health & Welfare (HI13C2124 to Dr. Sang-Cheol Bae). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions

Conceived and designed the experiments: KK SCB. Performed the experiments: KK SCB. Analyzed the data: KK. Contributed reagents/materials/analysis tools: KK SYB DHY SKC CBC YKS THK JBJ YMK CHS SCS SSL JL WTC SKK JYC SKN HSL SCB. Wrote the paper: KK SCB.

References

- Reche PA, Reinherz EL. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. J Mol Biol 2003; 331: 623–641. PMID: 12899833
- Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 2012; 44: 291–296. doi: 10.1038/ng.1076 PMID: 22286218
- Okada Y, Kim K, Han B, Pillai NE, Ong RT, Saw WY, et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. Hum Mol Genet 2014; 23: 6916–6926. doi: 10.1093/hmg/ddu387 PMID: 25070946
- Kim K, Bang SY, Lee HS, Okada Y, Han B, Saw WY, et al. The HLA-DRbeta1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. Nat Commun 2014; 5: 5902. doi: 10.1038/ ncomms6902 PMID: 25533202
- Kim K, Bang SY, Lee HS, Bae SC. Construction and application of a Korean reference panel for imputing classical alleles and amino acids of human leukocyte antigen genes. PLoS One 2014; 9: e112546. doi: 10.1371/journal.pone.0112546 PMID: 25398076
- Robinson J, Halliwell JA, McWilliam H, Lopez R, Parham P, Marsh SG. The IMGT/HLA database. Nucleic Acids Res 2013; 41: D1222–1227. doi: 10.1093/nar/gks949 PMID: 23080122



- Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am J Hum Genet 2009; 84: 210–223. doi: 10.1016/ j.ajhg.2009.01.005 PMID: 19200528
- Jia X, Han B, Onengut-Gumuscu S, Chen WM, Concannon PJ, Rich SS, et al. Imputing amino acid polymorphisms in human leukocyte antigens. PLoS One 2013; 8: e64683. doi: <u>10.1371/journal.pone.</u> 0064683 PMID: 23762245
- Kim K, Bang SY, Lee HS, Cho SK, Choi CB, Sung YK, et al. High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. Ann Rheum Dis 2015; 74: e13. doi: 10.1136/annrheumdis-2013-204749 PMID: 24532676
- 10. Andersson G. Evolution of the human HLA-DR region. Front Biosci 1998; 3: d739-745. PMID: 9675159
- Andersson G, Andersson L, Larhammar D, Rask L, Sigurdardottir S. Simplifying genetic locus assignment of HLA-DRB genes. Immunol Today 1994; 15: 58–62. PMID: 8155263
- Satta Y, Mayer WE, Klein J. HLA-DRB intron 1 sequences: implications for the evolution of HLA-DRB genes and haplotypes. Hum Immunol 1996; 51: 1–12. PMID: 8911992
- Hohjoh H, Ohashi J, Takasu M, Nishioka T, Ishida T, Tokunaga K. Recent divergence of the HLA-DRB1*04 allelic lineage from the DRB1*0701 lineage after the separation of the human and chimpanzee species. Immunogenetics 2003; 54: 856–861. PMID: 12671736
- 14. Cotner T, Charbonneau H, Mellins E, Pious D. mRNA abundance, rather than differences in subunit assembly, determine differential expression of HLA-DR beta 1 and -DR beta 3 molecules. J Biol Chem 1989; 264: 11107–11111. PMID: 2786876
- Czerwony G, Alten R, Gromnica-Ihle E, Hagemann D, Reuter U, Sorensen H, et al. Differential surface expression of HLA-DRB1 and HLA-DRB4 among peripheral blood cells of DR4 positive individuals. Hum Immunol 1999; 60: 1–9. PMID: 9952022
- Prat E, Tomaru U, Sabater L, Park DM, Granger R, Kruse N, et al. HLA-DRB5*0101 and -DRB1*1501 expression in the multiple sclerosis-associated HLA-DR15 haplotype. J Neuroimmunol 2005; 167: 108–119. PMID: 16111772