REPORT

Coding Variants at Hexa-allelic Amino Acid 13 of HLA-DRB1 Explain Independent SNP Associations with Follicular Lymphoma Risk

Jia Nee Foo,[1,17] Karin E. Smedby,[2,17] Nicholas K. Akers,[3] Mattias Berglund,[4] Ishak D. Irwan,[1] Xiaoming Jia,[5,6] Yi Li,[1] Lucia Conde,[7] Hatef Darabi,[8] Paige M. Bracci,[9] Mads Melbye,[10] Hans-Olov Adami,[8,11] Bengt Glimelius,[4] Chiea Chuen Khor,[1,12,13,14] Henrik Hjalgrim,[10] Leonid Padyukov,[15] Keith Humphreys,[8] Gunilla Enblad,[4] Christine F. Skibola,[7,18] Paul I.W. de Bakker,[5,6,16,18] and Jianjun Liu[1,14,18,*]

Non-Hodgkin lymphoma represents a diverse group of blood malignancies, of which follicular lymphoma (FL) is a common subtype. Previous genome-wide association studies (GWASs) have identified in the human leukocyte antigen (HLA) class II region multiple independent SNPs that are significantly associated with FL risk. To dissect these signals and determine whether coding variants in HLA genes are responsible for the associations, we conducted imputation, HLA typing, and sequencing in three independent populations for a total of 689 cases and 2,446 controls. We identified a hexa-allelic amino acid polymorphism at position 13 of the HLA-DR beta chain that showed the strongest association with FL within the major histocompatibility complex (MHC) region (multiallelic $p = 2.3 \times 10^{-15}$). Out of six possible amino acids that occurred at that position within the population, we classified two as high risk (Tyr and Phe), two as low risk (Ser and Arg), and two as moderate risk (His and Gly). There was a 4.2-fold difference in risk (95% confidence interval = 2.9–6.1) between subjects carrying two alleles encoding high-risk amino acids and those carrying two alleles encoding low-risk amino acids ($p = 1.01 \times 10^{-14}$). This coding variant might explain the complex SNP associations identified by GWASs and suggests a common HLA-DR antigen-driven mechanism for the pathogenesis of FL and rheumatoid arthritis.

Four genome-wide association studies (GWASs) have recently revealed complex associations between genetic variants in the human leukocyte antigen (HLA) region and follicular lymphoma (FL [MIM 613024]) risk,[1–4] particularly two independent associations tagged by rs10484561[1] and rs2647012[2] within the HLA class II region. Further imputation with tag SNPs[1,5] and HLA typing[6] revealed that coding-sequence variation in the molecules encoded by the extended *HLA-DRB1*0101-HLA-DQA1*0101-HLA-DQB1*0501* haplotype might be responsible for the association at rs10484561 and that the *DRB1*15-DQA1*01-DQB1*06* haplotype might partly explain the association at rs2647012.[6] A recent analysis also observed the association between gene-expression changes and rs2647012, but not rs10484561.[7] These previous findings indicate an important role of genetic variation in the HLA class II region in FL pathogenesis, but the underlying causal variants that drive the association are still unknown. Each extended haplotype and classical HLA allele is defined by a precise combination of coding differences at various amino acid (AA) positions in the encoded HLA molecules, and it is possible that changes at the AA level might impact antigen binding and therefore influence disease pathogenesis through altered immune response.

To determine whether specific coding variants within HLA genes contribute to the diverse association signals, we imputed dense SNPs (by using 1000 Genomes Project data[8]) and classical HLA alleles and coding variants across the HLA region (b36, chr6: 20–40 Mb) in our GWAS discovery sample of 379 cases and 791 controls[2] from Sweden and Denmark (the Scandinavian Lymphoma Etiology [SCALE] study). Imputation of classical HLA alleles and their constituent single-nucleotide variants and corresponding AAs was performed with BEAGLE as previously described.[9,10] The reference panel was constructed with the use of genotype data from the Major Histocompatibility Complex (MHC) Working Group of the Type 1 Diabetes

[1]Human Genetics, Genome Institute of Singapore, Agency for Science, Technology, and Research, Singapore 138672, Singapore; [2]Clinical Epidemiology Unit, Department of Medicine, Karolinska Institutet, Stockholm SE-171 77, Sweden; [3]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, CA 94720, USA; [4]Department of Radiology, Oncology, and Radiation Science, Uppsala University, Uppsala SE-751 85, Sweden; [5]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; [6]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; [7]Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL 35294, USA; [8]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-171 77, Sweden; [9]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94107, USA; [10]Department of Epidemiology Research, Statens Serum Institut, Copenhagen DK-2300, Denmark; [11]Department of Epidemiology, Harvard University, Boston, MA 02115, USA; [12]Singapore Eye Research Institute, Singapore 168751, Singapore; [13]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119074, Singapore; [14]Saw Swee Hock School of Public Health, National University Health Systems, National University of Singapore, Singapore 117597, Singapore; [15]Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm SE-171 76, Sweden; [16]Department of Medical Genetics and of Epidemiology, University Medical Center Utrecht, Utrecht 3584 CX, the Netherlands
[17]These authors contributed equally to this work
[18]These authors contributed equally to this work and are co-senior authors
*Correspondence: liuj3@gis.a-star.edu.sg
http://dx.doi.org/10.1016/j.ajhg.2013.05.020.

Table 1. Association Statistics from Testing Protective Alleles versus All Others and Risk Alleles versus All Others at AA 13 of HLA-DRB1

	No. of Cases (Frequency)	No. of Controls (Frequency)	OR (95% CI)	Trend Test p Value
Ser + Arg at AA 13				
Discovery	379 (0.379)	791 (0.509)	0.591 (0.495–0.707)	7.83×10^{-9}
				dosage: 8.26×10^{-9}
Replication 1	222 (0.439)	220 (0.573)	0.584 (0.446–0.766)	1.01×10^{-4}
Replication 2	88 (0.403)	1,435 (0.515)	0.640 (0.469–0.872)	4.69×10^{-3}
Meta-analysis	**p**	**OR**	**P_{het}**	**I^2**
	6.51×10^{-14}	0.598	0.893	0
Tyr + Phe at AA 13				
Discovery	379 (0.326)	791 (0.228)	1.660 (1.363–2.020)	4.87×10^{-7}
				dosage: 4.69×10^{-7}
Replication 1	222 (0.365)	220 (0.239)	1.822 (1.355–2.450)	7.15×10^{-5}
Replication 2	88 (0.375)	1,435 (0.226)	1.954 (1.437–2.660)	1.90×10^{-5}
Meta-analysis	**p**	**OR**	**P_{het}**	**I^2**
	2.00×10^{-14}	1.760	0.659	0

Abbreviations are as follows: OR, odds ratio; CI, confidence interval; AA, amino acid; P_{het}, Cochran's Q test p value; and I^2, inconsistency measure.

Genetics Consortium; these data consist of 2,537 SNPs genotyped in the MHC region and classical types at 4-digit resolution from 11,173 individuals. On the basis of the EMBL-EBI Immunogenetics HLA Database, AA variants (based on codons) were coded as binary markers (present or absent) in the reference panel. The final reference panel for imputation contained 2,767 unrelated founder individuals of European descent from collections across Europe, the United Kingdom, and North America[9,10] and data for 263 classical HLA alleles and 372 AA positions. We used default parameters for BEAGLE (ten iterations of phasing and imputation and testing four pairs of haplotype pairs for each individual at each iteration)[9,10] (average BEAGLE $r^2 = 0.96$). SNPs were imputed with IMPUTE2 (v.2.2.2)[11] on the basis of phased genotype data from the 1000 Genomes Project Phase I integrated variant set v.3[8] (average info score for SNPs with MAF $> 0.01 = 0.90$). We set all SNP genotypes imputed with genotype probability lower than 0.9 to missing and analyzed all imputed HLA genotypes. SNPs imputed with an info score < 0.8 in IMPUTE2, >10% missing data, or MAF < 0.01 were excluded from further analyses. A total of 71,954 SNPs (average info score = 0.97), 263 classical HLA alleles (two- and four-digit resolution), and 372 AA positions were imputed. Of these, 86 AA positions were "multiallelic," i.e., had three to six different residues encoded at each position across all our samples.

We conducted trend tests on all the imputed biallelic variants while adjusting for PC1-3 as covariates[2] (genomic inflation in the GWAS = 1.028). For multiallelic AA sites (triallelic, quadrallelic, penta-allelic, or hexa-allelic), we performed the global multiallelic test at each site and further tested every possible combination of one, two, or three encoded AAs against the rest. Association tests were performed on both the best-guess genotypes and the allelic dosages as determined from the imputed genotype probabilities, which accounted for any uncertainties in the imputed genotypes. The results were checked for consistency between the two methods, and the results from best-guess genotypes were presented. We also performed global multiallelic (genotypic) tests while taking into account all alleles at each multiallelic position. The multiallelic test was performed as follows: convert k-alleles to k-1 bialleles, invoke the glm function in R to estimate the multivariate model, and use the likelihood ratio test to compute the global multiallelic test p values. We performed all conditional logistic regression analyses with PLINK[12] by entering any five out of six alleles encoding AA 13 as covariates. Pairwise linkage disequilibrium (LD; r^2) between SNPs and alleles were measured in PLINK.

Among all the variants tested, the top signal of association came from a combination of two possible genotypes, coding for either Ser or Arg at the hexa-allelic AA 13 encoded by exon 2 of *HLA-DRB1* (MIM 142857). This variant showed stronger association (Table 1; $p_{Ser+Arg13} = 7.8 \times 10^{-9}$, odds ratio [OR] = 0.59; 95% confidence interval [CI] = 0.50–0.70) than any other HLA variant or SNP (genotyped or imputed) tested in our study and the strongest association in global multiallelic tests across all AA positions (multiallelic $p = 7.4 \times 10^{-8}$). AA 13 genotypes accounted for slightly more variance in FL risk (3.1%) than both rs2647012 and rs10484561 combined (3.0%). The strongest-associated imputed classical HLA allele across all two- and four-digit alleles was *HLA-DQB1*06* (OR = 0.63; 95% CI = 0.51–0.78; $p = 2.05 \times 10^{-5}$), which could be accounted for by rs2647012 alone

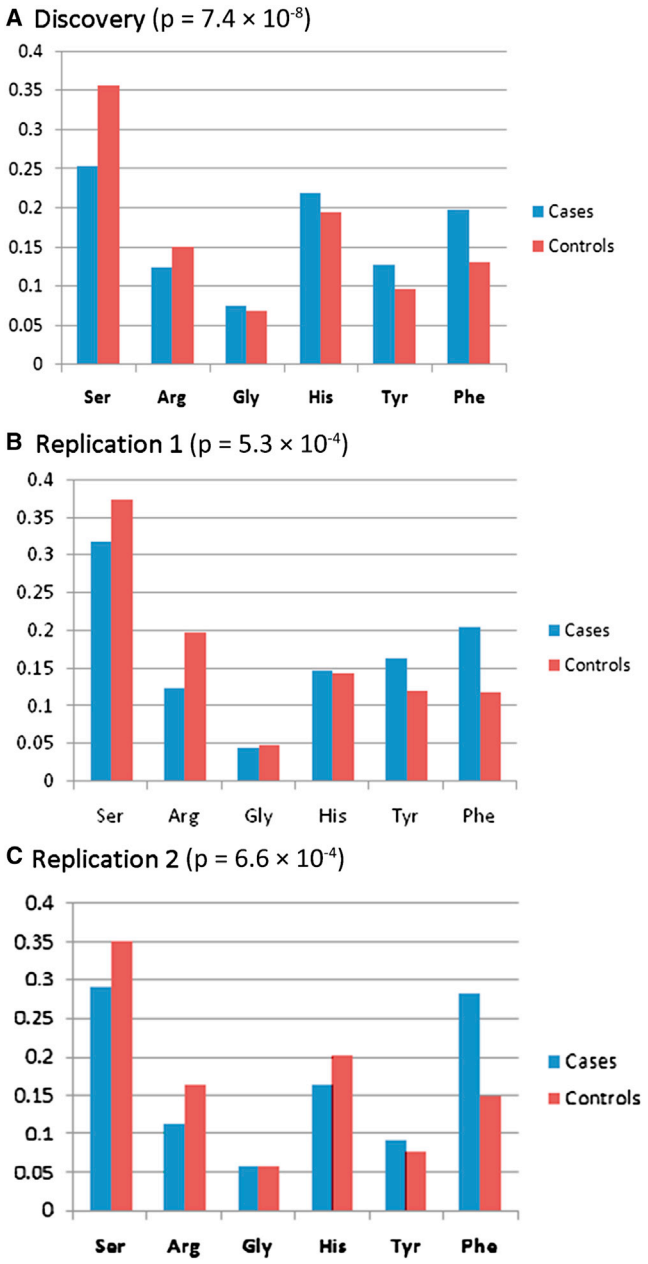

Figure 1. Allele Frequencies at Hexa-Allelic AA 13
Allele frequencies of each AA at position 13 in (A) the SCALE GWAS discovery data set (379 cases and 791 controls), (B) the San Francisco data set (222 cases and 220 controls), and (C) the Swedish validation data set (88 cases and 1,435 controls) and association p values from the global multiallelic test.

($OR_{adj\ rs2647012} = 0.86$; 95% CI = 0.65–1.15; $p_{adj\ rs2647012} = 0.32$). The top imputed SNP was rs9268839, in complete LD ($r^2 = 1$; crude OR = 1.64; 95% CI = 1.37–1.95; $p = 4.57 \times 10^{-8}$) with rs9378212 (which we previously reported and validated technically by Taqman genotyping[2]) and partially correlated with rs2647012 ($r^2 = 0.56$) and rs10484561 ($r^2 = 0.15$) ($OR_{adj\ 2snps} = 1.25$; 95% CI = 0.94–1.64; $p_{adj\ 2snps} = 0.12$). The FL-associated HLA SNPs reported in a recent GWAS[4] were also not independent of these previously reported SNPs in our data set (Table S1, available online).

Six alleles were present at AA position 13 within our samples and in the reference panel;[9,10] those encoding Ser and Arg showed low risk of FL, those encoding Tyr and Phe showed high risk, and those encoding His and Gly showed moderate risk (Figure 1, Table S2, and Figure S2). Conditioning upon the alleles at position 13 was sufficient to eliminate all the association signals within the vicinity in our study (chr6: 32–33 Mb; Table S3 and Figure S1). Dosage analyses (Table 1), HLA typing,[10] and sequencing analyses[13,14] confirmed high accuracy of the imputation (BEAGLE $r^2 = 0.99$) at this position and high genotype concordance across these platforms (~98%; see below).

To validate the associations, we evaluated genotypes at AA 13 in two independent data sets of 222 FL cases and 220 controls from San Francisco[6] (replication 1) and 88 FL cases and 1,435 controls from Sweden (replication 2) by using a combination of Sanger sequencing (replication 2 cases),[13,14] canonical HLA typing (replication 2 controls),[10] and GS-FLX sequencing (replication 1).[6] Replication 1 included non-Hispanic white FL cases and controls who were part of a population-based case-control non-Hodgkin lymphoma (NHL) study conducted in the San Francisco Bay Area from 2001 to 2006.[6] Controls were matched according to the number or frequency of cases in 5-year age groups, sex, and county of residence. We carried out HLA typing in these samples by using the Roche high-resolution HLA primer set and subsequently performed GS-FLX sequencing as previously described.[6] Replication 2 included 19- to 83-year old (median age = 58 years) FL cases of Swedish descent and with available fresh frozen tumor material assembled in the Uppsala-Örebro region from 1970 to 2006;[15] these cases were typed by PCR amplification and Sanger sequencing (replication 2). Primer sequences for the amplification of *HLA-DRB1* exon 2 were obtained from previously published work.[13,14] PCR products were purified with the use of AMPure XP (Agencourt) solid-phase-reversible-immobilization beads and run on a Bioanalyzer (Agilent) or 15% polyacrylamide gel for ensuring that excess primers were removed prior to sequencing. Sequence reads were visually inspected at the codon position encoding AAs 11 and 13, and genotypes were manually called (Figure S3) on the basis of expected codons at both positions and flanking ones according to sequences from the EMBL-EBI Immunogenetics HLA Database. For controls in replication 2, we used a second independent set of 1,435 Swedish control subject samples collected within the Epidemiological Investigation of Rheumatoid Arthritis (EIRA) study from 1995 to 2006. HLA typing was carried out in the control samples from the EIRA study with the use of sequence-specific primer PCR (DR low-resolution kit; Olerup SSP) as previously described.[16] All discovery and replication studies were conducted in accordance with the ethical standards of the respective institutional review board of each institution, and informed consent was obtained from study participants.

We first tested the accuracy of the imputed genotypes by amplification and Sanger sequencing[13,14] of DRB1 exon 2 in a subset of our discovery samples. We observed high concordance between imputed genotypes and those inferred from the sequence chromatograms (Figure S3) (n = 92, 98.4% allelic, 97% genotypic concordance) and those inferred from HLA types[8] (Table S6; n = 596, 98.6% allelic, 97.3% genotypic concordance), confirming the high accuracy of the imputed genotypes and the reliability of the association results in our samples. To assess the accuracy of the AA genotypes inferred from canonical HLA typing data, we also Sanger sequenced a subset of 189 EIRA controls and found the alleles to be 97.6% concordant; hence, there are not likely to be any major biases because of the use of different genotyping platforms in replication 2.

We observed consistent direction of associations at all of the high- and low-risk residues at AA 13 across the three study populations (Figure 1, Table S2). Although the association did not reach statistical significance for some of the individual alleles (Table S2), it remained significant for the combined protective (encoding Ser or Arg) and risk (encoding Tyr or Phe) alleles in all the studies (Table 1). In the meta-analysis of all three populations, there was strong evidence of association between the AA 13 polymorphism and FL risk and no evidence of heterogeneity (multiallelic $p = 2.3 \times 10^{-15}$; $p_{Ser+Arg13} = 6.5 \times 10^{-14}$; $OR_{Ser+Arg13} = 0.60$; heterogeneity $I^2 = 0$) (Table 1). Across the three data sets, there was a 4.2-fold difference in risk (95% CI = 2.9–6.1) between subjects carrying two high-risk alleles and those carrying two low-risk alleles (Cochran-Mantel-Haenszel $p = 1.01 \times 10^{-14}$).

Conditional analyses followed by meta-analyses across all three data sets suggested that the AA 13 polymorphism might fully explain associations observed at SNPs rs2647012 ($p = 4.72 \times 10^{-11}$ before adjustment; p = 0.804 after adjustment) and rs10484561 ($p = 2.61 \times 10^{-11}$ before adjustment; p = 0.356 after adjustment) (Table S3). However, there was evidence of heterogeneity in the results of the conditional analysis across the three data sets ($Q < 0.05$; $I^2 > 70\%$). In particular, there was residual association at rs10484561 and rs2647012 after adjustment for the genetic effects at AA 13 in the replication 1 data set (Table S3). Conversely, conditioning on genotypes at rs2647012 (multiallelic $p = 1.68 \times 10^{-7}$), rs10484561 (multiallelic $p = 7.58 \times 10^{-7}$), or both SNPs (multiallelic p = 0.005) did not fully eliminate the association observed at AA 13, suggesting that the genotypes at AA 13 are well tagged, but not fully tagged, by these two SNPs.

Further haplotype analysis indicated that the minor C allele at rs10484561 partially tags the haplotypes carrying the high-risk allele encoding Phe (OR = 1.81; 95% CI = 1.51–2.18), whereas the minor T allele at rs2647012 partially tags the haplotypes carrying the low-risk alleles encoding Ser (OR = 0.69; 95% CI = 0.59–0.81) and Arg (OR = 0.67; 95% CI = 0.55–0.82) (Table S4). For this analysis, genotypes (six imputed alleles at AA 13 and two SNPs) were phased with the PHASE program[17] with default parameters. Phased haplotypes were then tested for association with FL with the use of logistic regression analysis in PLINK. Only haplotypes with minor allele frequencies > 1% in controls were analyzed. It is interesting to note that some rare (allele frequencies ~2%) haplotypes, such as Phe-C-A (Table S4), are not tagged by rs10484561 and rs2647012 but showed diverse associations across the three study populations, and this contributed to the large heterogeneity of the results of conditional association analysis across the three data sets above. Further studies with much larger sample sizes will be needed for evaluating the effects of these rare haplotypes. Taken together, our data suggest that the hexa-allelic AA 13 polymorphism might be the primary driver of the association within the region, whereas the diverse associations at multiple SNPs might be due to their differential tagging effects of various AA risk variants.

AA 13 is located in the middle of the peptide binding groove of the HLA-DR heterodimer molecule (Figure S2) and is thus well positioned to directly interact with bound peptides. This position, together with positions 70, 71, and 74, has been shown to play important roles in the binding-specificity profile of pocket 4, which is one of the most important pockets for antigen interaction and presentation by the HLA-DR molecule.[18] Chemical properties of the side chain at this position might have a direct effect on antigen binding and recognition within the binding groove. The high-risk alleles at AA 13 both encode AAs with bulky hydrophobic side chains (Phe and Tyr), whereas the low-risk alleles encode AAs with polar or charged side chains (Figure S2). The roles of the moderate-risk alleles (encoding His and Gly) will need to be further explored in larger samples and/or functional assays. Although our results seem to suggest that variation at position 13 plays an important role in influencing FL risk, other nearby residues within the extended HLA haplotype might also influence peptide binding, and detailed functional work will be needed for proving the importance of this single position.

Residues at AA 13 are in high LD with nearby residues at AA 11, the second-best-associated coding variant (multiallelic $p = 3.6 \times 10^{-7}$) in discovery GWASs (Table S5). Given the close proximity and tight LD between the two variants, we were unable to distinguish the effects without additional functional investigation. Variants at HLA-DRB1 AAs 11 and 13 have previously been shown to associate with risk of rheumatoid arthritis (RA [MIM 180300]).[10] RA is an autoimmune disorder with a well-established correlation with NHL risk (mostly with the diffuse large B cell NHL subtype, but also with FL).[19] Although the patterns of association at AAs 11 and 13 differ between RA and FL,[10] the shared associations between the variants at HLA-DRB1 AAs 11–13 and FL and RA clearly suggest the involvement of common HLA-DR antigen-driven pathogenesis in the two different diseases.

Further studies on large cohorts with detailed clinical information on both diseases are likely to reveal information on the shared or distinct etiological mechanisms of RA and lymphoma pathogenesis. In vitro biochemical studies will be needed for testing the effects of mutants with different combinations of AAs at these two positions on binding of a relevant autoantigen or tumor antigen to the beta chain of HLA-DR.

We have performed comprehensive imputation of SNPs, classical HLA alleles, and HLA coding variants by using a large reference panel of European subjects. Recent evaluation studies have demonstrated high accuracy of the imputations of these coding variants.[20] Although our imputation confidence was high (BEAGLE $r^2 = 0.96$), there might have been minor inaccuracies owing to the nature of the imputation. It is possible that the GWAS samples might have been differentiated (in terms of population structure) relative to the reference samples used for the HLA imputations. We expect that this would have led to a consistent loss of imputation accuracy and therefore of power in both cases and controls and hence do not expect a differential bias in terms of imputation accuracy between cases and controls (a differential bias could inflate type 1 error). Nonetheless, we have confirmed the accuracy of genotypes at the most significant imputed variants by direct Sanger resequencing and additional HLA typing using sequence-specific primer PCR and have demonstrated high concordance. We have also replicated the results in two independent sample collections that were directly genotyped with experimental methods. We anticipate that there might be haplotypes or variants, especially rare ones with allele frequencies $< 1\%$, that are poorly tagged and hence might have been missed in this analysis. In particular, there might be additional rare causal variants that could account for the residual association observed in replication 1. Our current sample size is, however, not large enough for a conclusive assessment of these possibilities, and the analysis of much larger reference panels and GWAS sample sizes will be needed for accurately evaluating the associations of these rare haplotypes.

In summary, through a comprehensive imputation and further experimental validation analysis of the HLA region (Figure S4), we have shown that the variants at a single hexa-allelic AA position (13) of HLA-DRB1 influence FL risk. This AA might account for most of the currently observed independent SNP signals previously identified through GWASs in populations of European descent. There was, however, an indication of residual associations in the replication 1 data set, and confirming these results therefore warrants further study. Nevertheless, here we show strong evidence that coding variants at a single AA position of HLA-DRB1 contribute to multiple association signals observed for FL. Our findings further suggest that this multiallelic AA polymorphism might explain a significant portion of the genetic associations observed for FL within the HLA class II region.

Supplemental Data

Supplemental Data include four figures and six tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

This work was funded by the Agency for Science and Technology and Research of Singapore, the Swedish Cancer Society (090659), the Swedish Research Council (K2008-64X-20737-01-2, 523-2006-972), and the Danish Medical Research Council (FSS 09-63424). The collection of blood samples in the founding case-control study was supported by the National Cancer Institute (NCI) (CA069269-01). Sample collection in the Epidemiological Investigation of Rheumatoid Arthritis study was supported by the Swedish Council for Working life (2008-0567), AFA Insurance, and the COMBINE project, funded by Vinnova. The San Francisco study was supported by the NCI, National Institutes of Health (CA122663, CA104682, CA45614, and CA89745). The Uppsala study was supported by the Lions Cancer Research Foundation, the Selanders Research Foundation, and the Research Foundation of the Department of Oncology of Uppsala University Hospital. P.I.W.d.B. is supported, in part, by a VIDI Award from the Netherlands Organization for Scientific Research (NWO-016.126.354).

Received: April 8, 2013
Revised: May 21, 2013
Accepted: May 23, 2013
Published: June 20, 2013

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes Project, http://www.1000genomes.org/
EMBL-EBI Immunogenetics HLA Database, http://www.ebi.ac.uk/ipd/imgt/hla/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
Type 1 Diabetes Genetics Consortium, http://www.t1dgc.org/

References

1. Conde, L., Halperin, E., Akers, N.K., Brown, K.M., Smedby, K.E., Rothman, N., Nieters, A., Slager, S.L., Brooks-Wilson, A., Agana, L., et al. (2010). Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. Nat. Genet. *42*, 661–664.
2. Smedby, K.E., Foo, J.N., Skibola, C.F., Darabi, H., Conde, L., Hjalgrim, H., Kumar, V., Chang, E.T., Rothman, N., Cerhan, J.R., et al. (2011). GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. PLoS Genet. *7*, e1001378.
3. Skibola, C.F., Bracci, P.M., Halperin, E., Conde, L., Craig, D.W., Agana, L., Iyadurai, K., Becker, N., Brooks-Wilson, A., Curry, J.D., et al. (2009). Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat. Genet. *41*, 873–875.
4. Vijai, J., Kirchhoff, T., Schrader, K.A., Brown, J., Dutra-Clarke, A.V., Manschreck, C., Hansen, N., Rau-Murthy, R., Sarrel, K., Przybylo, J., et al. (2013). Susceptibility loci associated with

specific and shared subtypes of lymphoid malignancies. PLoS Genet. *9*, e1003220.
5. de Bakker, P.I., McVean, G., Sabeti, P.C., Miretti, M.M., Green, T., Marchini, J., Ke, X., Monsuur, A.J., Whittaker, P., Delgado, M., et al. (2006). A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat. Genet. *38*, 1166–1172.
6. Skibola, C.F., Akers, N.K., Conde, L., Ladner, M., Hawbecker, S.K., Cohen, F., Ribas, F., Erlich, H.A., Goodridge, D., Trachtenberg, E.A., et al. (2012). Multi-locus HLA class I and II allele and haplotype associations with follicular lymphoma. Tissue Antigens *79*, 279–286.
7. Conde, L., Bracci, P.M., Richardson, R., Montgomery, S.B., and Skibola, C.F. (2013). Integrating GWAS and expression data for functional characterization of disease-associated SNPs: an application to follicular lymphoma. Am. J. Hum. Genet. *92*, 126–130.
8. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A.; 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.
9. Pereyra, F., Jia, X., McLaren, P.J., Telenti, A., de Bakker, P.I., Walker, B.D., Ripke, S., Brumme, C.J., Pulit, S.L., Carrington, M., et al.; International HIV Controllers Study. (2010). The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. Science *330*, 1551–1557.
10. 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.
11. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.
12. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
13. Sayer, D.C., Whidborne, R., De Santis, D., Rozemuller, E.H., Christiansen, F.T., and Tilanus, M.G. (2004). A multicenter international evaluation of single-tube amplification protocols for sequencing-based typing of HLA-DRB1 and HLA-DRB3,4,5. Tissue Antigens *63*, 412–423.
14. Sayer, D., Whidborne, R., Brestovac, B., Trimboli, F., Witt, C., and Christiansen, F. (2001). HLA-DRB1 DNA sequencing based typing: an approach suitable for high throughput typing including unrelated bone marrow registry donors. Tissue Antigens *57*, 46–54.
15. Thunberg, U., Enblad, G., Turesson, I., and Berglund, M. (2010). Genetic variation in tumor necrosis factor and risk of diffuse large B-cell lymphoma and follicular lymphoma: differences between subgroups in Swedish patients. Leuk. Lymphoma *51*, 1563–1566.
16. Lundström, E., Källberg, H., Smolnikova, M., Ding, B., Rönnelid, J., Alfredsson, L., Klareskog, L., and Padyukov, L. (2009). Opposing effects of HLA-DRB1*13 alleles on the risk of developing anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. Arthritis Rheum. *60*, 924–930.
17. Stephens, M., Smith, N.J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. *68*, 978–989.
18. Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Tuereci, O., Sahin, U., Braxenthaler, M., Gallazzi, F., Protti, M.P., Sinigaglia, F., and Hammer, J. (1999). Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. Nat. Biotechnol. *17*, 555–561.
19. Baecklund, E., Backlin, C., Iliadou, A., Granath, F., Ekbom, A., Amini, R.M., Feltelius, N., Enblad, G., Sundström, C., Klareskog, L., et al. (2006). Characteristics of diffuse large B cell lymphomas in rheumatoid arthritis. Arthritis Rheum. *54*, 3774–3781.
20. 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. http://dx.doi.org/10.1371/journal.pone.0064683.