Published in final edited form as:

Tissue Antigens. 2012 April ; 79(4): 279–286. doi:10.1111/j.1399-0039.2012.01845.x.

Multi-locus HLA class I and II allele and haplotype associations with follicular lymphoma

Christine F. Skibola^{1,6}, Nicholas K. Akers¹, Lucia Conde¹, Martha Ladner², Sharon K. Hawbecker², Franziska Cohen², Fernanda Ribas², Henry A. Erlich^{2,3}, Damian Goodridge⁴, Elizabeth A. Trachtenberg², Martyn T. Smith¹, and Paige M. Bracci⁵

¹Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA, USA

²Center for Genetics, Children's Hospital Oakland Research Institute, Oakland, CA, USA

³Conexio Genomics, Perth, Australia

⁴Roche Molecular Systems, Alameda CA

⁵Department of Epidemiology and Biostatistics, School of Medicine, University of California San Francisco, San Francisco, CA, USA

Abstract

Follicular lymphoma (FL) is an indolent, sometimes fatal disease characterized by recurrence at progressively shorter intervals and is frequently refractive to therapy. Genome-wide association studies have identified SNPs in the human leukocyte antigen (HLA) region on chromosome 6p21.32–33 that are statistically significantly associated with FL risk. Low to medium resolution typing of single or multiple HLA genes has provided an incomplete picture of the total genetic risk imparted by this highly variable region. To gain further insight into the role of HLA alleles in lymphomagenesis and to investigate the independence of validated SNPs and HLA alleles with FL risk, high-resolution HLA typing was conducted using next-generation sequencing in 222 non-Hispanic white FL cases and 220 matched controls from a larger San Francisco Bay Area population-based case-control study of lymphoma. A novel protective association was found between the *DPB1**03:01 allele and FL risk (OR=0.39, 95% CI 0.21–0.68). Extended haplotypes *DRB1**01:01-*DQA1**01:01-*DQB1**05:01 (OR=2.01, 95% CI 1.22–3.38) and *DRB1**15-*DQA1**01-*DQB1**06 (OR=0.55, 95% CI 0.36–0.82) also influenced FL risk. Moreover, *DRB1**15-*DQA1**01-*DQB1**06 was highly correlated with an established FL risk locus, rs2647012. These results provide further insight into the critical roles of HLA alleles and SNPs in FL pathogenesis that involve multi-locus effects across the HLA region.

Keywords

follicular lymphoma; HLA; genetic risk factors; next-generation sequencing

⁶Correspondence to: School of Public Health, Division of Environmental Health Sciences, 237A Hildebrand Hall, University of California, Berkeley, California 94720-7356 (510) 643-5041 tel / (510) 642-0427 fax / chrisfs@berkeley.edu.

CONFLICTS OF INTEREST

There are no conflicts of interest.

AUTHORSHIP CONTRIBUTIONS

C.F.S., M.T.S. and P.M.B. designed the study. C.F.S. drafted the manuscript with significant contributions from N.K.A. and E.A.T. M.L., S.K.H., F.C., F.R., D.G., H.E. and E.A.T. performed high-resolution HLA sequencing analysis. N.K.A. and L.C. analyzed the data. All authors reviewed and approved the final manuscript.

INTRODUCTION

Follicular lymphoma (FL) is an indolent B-cell malignancy characterized by a highly variable clinical course and multiple relapses (1). Approximately one-third of FL cases transform to a more aggressive histology, usually diffuse large B-cell lymphoma (DLBCL), which is associated with a poor clinical outcome (2, 3). The molecular basis of FL has been fairly well characterized (4–6), although, its root causes remain less clear. In recent genome-wide association studies (GWAS) of non-Hodgkin lymphoma (NHL) and validation within the large InterLymph consortium, we identified three independent susceptibility loci for FL on chromosome 6p21.3 in the human leukocyte antigen (HLA) class I and II regions (7–9). Located in the HLA class I region at 6p21.33 near psoriasis susceptibility region 1, rs6457327 was inversely associated with risk of FL (p -value= 4.7×10^{-11}) (8). In the HLA class II region at 6p21.32, two SNPs, rs10484561 and rs7755224, were associated with two-fold increased risks of FL (p -values= 1.12×10^{-29} and 2.0×10^{-19} , respectively) (7). rs10484561 and rs7755224 are in total linkage disequilibrium (LD) and are located, respectively, 29 and 16 kb centromeric of *HLA-DQB1*. Based on a tag SNP analysis, we inferred that rs10484561 may be part of a high-risk extended haplotype, *DRB1*01:01-DQA1*01:01-DQB1*05:01* (7). Another class II locus in the *HLA-DQB1* region, rs2647012, was inversely associated with FL risk after adjusting for rs10484561 (OR=0.70, p -value= 4×10^{-12}) (9). In subsequent studies, we confirmed a positive association between FL risk and the *DQB1*05* allele group (p -value=0.013) and identified the *DQB1*06* allele group as protective for FL (p -value= 4.5×10^{-5}) (10). An independent study further supported *DRB1*01:01* as a risk locus for FL (11). Taken together, these studies suggest that genetic variation in the HLA region plays an important role in the etiology of FL.

HLA class I- and class II-restricted CD8+ and CD4+ T-cell responses are essential for the immune system to mount a successful anti-tumor immune defense or to remove infected cells. A defect in these important processes could allow pathogenic cells to escape host immune recognition that may increase the likelihood of lymphomagenesis. To further pinpoint risk-associated HLA alleles and haplotypes in the pathogenesis of FL, we investigated whether previously validated FL-associated GWAS SNPs (rs6457327, rs10484561, and rs2647012) and HLA alleles were independent risk factors for FL. To this end, we extended the analysis of our NHL case-control study to determine HLA class I (*HLA-A*, *-B*, *-C*) and class II (*DRB1/3/4/5*, *DQA1*, *DQB1*, *DPB1*) alleles using high-resolution HLA typing by next-generation Roche GS FLX 454 sequencing in 222 white, non-Hispanic FL cases and 220 controls frequency-matched by sex and age in 5-year groups.

MATERIALS AND METHODS

Study population

Samples sequenced at HLA included non-Hispanic white FL cases (n=222) and frequency-matched controls (n=200) who were part of a population-based case-control study of NHL conducted in the San Francisco Bay Area that included 2,055 patients newly diagnosed with NHL from 2001 to 2006 frequency-matched to 2,081 control participants. The majority of these FL cases were included in the previously described GWAS (92.8%) (7) and *DQB1* typing study (95.9%) (10). Eligible patients were identified by the Cancer Prevention Institute of California's rapid case ascertainment and by SEER abstract, were 20 to 85 years old at diagnosis, alive at first contact, residents of one of six Bay Area counties, able to complete an interview in English and had no prior history of hematopoietic cancer or physician indicated contra-indications to contact. Eligible controls were identified by random digit dial, and by random sampling of the Centers for Medicare & Medicaid Services lists for individuals aged 65 years or older and were frequency-matched to cases by

5-yr age group, sex and county of residence. A total of 2,055 eligible cases and 2,081 controls were included in the parent study. Blood and/or buccal cells were collected from 85% of eligible study participants. Eligible NHL patients also provided consent (98%) to access their diagnostic materials to confirm diagnosis of NHL and for consistent classification of NHL subtype by the study pathologist using the WHO classification.

HLA genotyping by next-generation sequencing

High-resolution sequencing to obtain HLA genotypes (as in the IMGT/HLA database v3.6.0, <http://www.ebi.ac.uk/imgt/hla/>) was carried out as previously described in detail (12, 13). Briefly, next-generation clonal sequencing of exonic amplicons was performed using the Roche 454 GS FLX massively parallel pyrosequencing system (14). Roche-developed PCR primers to exons 2–4 for HLA class I (*A*, *B*, *C*), exons 2–3 for class II *DQB1*, exon 2 for *DRB1*/3/4/5, *DQA1* and *DPB1*; 11 multiplex identification (MID) tags were used in the 10 ng sample template amplifications. Primary HLA amplicons were purified to remove short artifacts, and then pooled in equimolar concentrations for emulsion PCR, bead recovery and pyrosequencing. Sequence data analysis was accomplished using the Conexio Genomics ATF software (Perth, Australia). In almost all HLA analyses to date, it has been cost-prohibitive to analyze all genomic regions for each gene to determine the unambiguous genotype of each sample; and until most of the genomic region of the genes is sequenced there will always be a level of ambiguity due to the high degree of polymorphism of HLA genes. The Roche GS FLX 454 clonal sequencing of HLA described here consequently results in some residual ambiguity which, although limited compared to other sequencing methods, must still be reduced for analysis. To do this, the alleles analyzed here were called based on the most common, “lowest number” alleles from a list of possible genotypes derived by clonal sequence analysis of particular exons. The allelic genotype calls and the related total possible six digit alleles from resolved genotypes and unresolved ambiguity are listed for each locus in Supplemental Tables 1–8. The nature of the clonal sequencing dramatically reduces the level of possible ambiguity using traditional Sanger sequencing, and this is the first time that a complete ambiguity table has been reported for HLA genotypes in an association study. Genotypes derived from a total of 89 samples (27 for *HLA-A*, 39 for *HLA-B*, 18 for *HLA-C* and 5 for *HLA-DRB1*) which failed at least one exon analysis by the clonal sequencing method were re-sequenced and re-tested using Luminex LABType SSO kits (One Lambda, Inc. Canoga Park, California, USA). This method uses sequence specific oligonucleotide probes bound to fluorescently coded microspheres to identify the HLA alleles in an amplified DNA sample, and alleles were identified using HLA Fusion software, v2.0.0 (One Lambda Inc.). Data from the LABType high resolution bead kits were used in addition to the sequencing data to fill in exon gaps to resolve the genotypes at a level comparable to the 454 genotypes. The HLA nomenclature used for the current data in this manuscript reflects the newest iteration of rules (2010) (<http://hla.alleles.org/announcement.html>) to describe HLA alleles, while all older data is presented as it was originally noted in the earlier publications.

Data analysis

Haplotype frequencies for cases and controls were estimated using the iterative Expectation-Maximization (EM) algorithm implemented in the Pypop software (15). LD between HLA alleles and rs6457327, rs10484561, and rs2647012 was measured in our control population using Pypop that calculates D' and chi-square values based on observed and expected frequencies of haplotypes. Deviations from Hardy-Weinberg equilibrium (HWE) in controls were tested with the Arlequin software v3.5.1.2 (16) using a Markov chain method with exact p-value estimation (17). No significant departure from HWE was observed for any loci at a $p < 0.001$ level.

For each individual allele or haplotype, the independence of the number of observed and unobserved counts in cases and controls was determined using the 'fisher.test' function from the 'stat' package in R (<http://stat.ethz.ch/R-manual/R-patched/library/stats/html/00Index.html>). Odds ratios (ORs) and 95% confidence intervals (CI) were estimated as further measures of the magnitude of the association between alleles or haplotypes and disease status. The 'p.adjust' function from the same package in R was used to adjust the *p*-values for the number of independent statistical tests at each locus using the Bonferroni correction.

Unconditional backward stepwise logistic regression methods in STATA version 11 (StataCorp, College Station, Texas, USA) were used to assess independence of individual risk loci. All established or suspected risk factors in the classic HLA regions (rs6457327, rs2647012, rs10484561-*DRB1**01:01-*DQA1**01:01-*DQB1**05:01, *DRB1**15-*DQA1**01-*DQB1**06, *DRB1**13-*DQA1**01-*DQB1**06, and *DPB1**03:01) were included and the final best fitting model was determined based on a likelihood ratio test. A *p*-value threshold of 0.10 was the criteria used for remaining in the model. For these analyses, one allele of each haplotype was used as a proxy for the haplotype as a whole (*DQB1**05:01, *DRB1**15, and *DRB1**13). Due to collinearity, *DRB1**15 and rs2647012 were assessed as a single variable where 0 indicated presence of neither allele, 1 indicated presence of rs2647012 alone, and 2 indicated the presence of both rs2647012 and *DRB1**15. All other alleles of interest were coded as present vs. absent.

RESULTS

The association results for all HLA class I and II alleles with *p*-values < 0.05 are shown in Table 1. We identified a novel protective allele, *DPB1**03:01, associated with risk of FL (OR=0.39, 95% CI 0.21–0.68, adjusted *p*-value=8.30×10⁻³, Table 2) that was not in significant LD with any HLA alleles previously shown to be associated with FL (*D'*=0.30 with rs10484561, *D'*=0.04 with rs6457327, and *D'*=0.03 with rs2647012, Table 3). *DPB1* is located centromeric to *DRB1*, *DQA1* and *DQB1* and is separated from these genes by a recombination hotspot(18). Using backward stepwise logistic regression methods to analyze HLA alleles and previously identified SNPs of interest, the final best fitting model showed that *DPB1**03:01 was independently associated with FL (Table 4).

As a follow-up to further explore independence between the GWAS SNP, rs10484561, and *DRB1*, *DQA1* and *DQB1* alleles, we confirmed our previous tag SNP analysis (7) implicating the extended haplotype, *DRB1**01:01-*DQA1**01:01-*DQB1**05:01, as a risk factor for FL. Each allele of the haplotype was in strong LD with rs10484561 (*D'*=0.93, 1.0, and 1.0, respectively, Table 3) and was associated with increased risk of FL (Table 5).

*DQB1**06 and *DRB1**13 have been reported as protective alleles for FL (10, 11). Because *DQB1**06 is known to exist in haplotypes with both *DRB1**13 and *DRB1**15 in Caucasian populations, it was unclear which haplotypes may be responsible for these associations. Here, we found that the *DQB1**06 and *DRB1**15 alleles were significantly associated with decreased FL risk (Table 6), and that although non-significant after correction, the frequency of the haplotypes *DRB1**15-*DQA1**01-*DQB1**06 and *DRB1**13-*DQA1**01-*DQB1**06 were similarly decreased in cases (Table 6). Logistic regression analysis revealed that *DRB1**13-*DQA1**01-*DQB1**06 was no longer associated with FL risk after adjustment for other FL associated HLA alleles (OR = 0.92, *p* = 0.83, Table 4). We also found that all carriers of the protective *DRB1**15:01-*DQA1**01:02-*DQB1**06:02 haplotype were carriers of the rs2647012 A allele, although the minor allele frequency of rs2647012 (0.40) was higher than the frequency of the linked *DRB1**15:01-*DQA1**01:02-*DQB1**06:02 haplotype (0.16). LD between rs2647012 and the individual alleles of the haplotype corroborated the high LD

between rs2647012 and *DRB1**15:01-*DQA1**01:02-*DQB1**06:02 (D' =1, 0.86, and 1, respectively, Table 3). Limiting the dataset to those individuals without *DRB1**15:01-*DQA1**01:02-*DQB1**06:02 revealed a modest effect of rs2647012 on FL risk (OR = 0.70, 95% CI 0.45–1.1, p = 0.10).

For HLA class I loci, no significant associations with FL risk were found (Table 1). However, we found that the *C**07:02 and *B**07:02 alleles were linked to rs6457327 'A' carriers (D' =0.93 and 1.0, respectively, Table 3). Restricting the dataset to those individuals without *C**07:02 or *B**07:02 made little change on the estimated risk statistic for rs6457327 (OR = 0.55, 95% CI 0.30–1.00, p = 0.05).

DISCUSSION

Previous GWAS and low-to-medium resolution HLA typing studies have identified major FL-susceptibility loci in the HLA class I and II region. As a follow-up, we conducted next generation, high-throughput HLA sequencing of class I (*HLA-A*, *-B*, *-C*) and class II (*DRB1*/3/4/5, *DQA1*, *DQB1*, *DPB1*) alleles to determine the independent role of HLA alleles and SNPs as susceptibility factors for FL. This study provides the first examination of *DPB1* alleles in FL cases, as well as the highest resolution and most complete characterization of HLA class I and II alleles to date. Here, we found that *DPB1**03:01, *DQB1**05:01, rs6457327 and *DRB1**15 all independently influence FL risk. Specifically, we identified a novel, inverse association between the *DPB1**03:01 allele and risk of FL that was independent of other HLA class II alleles based on LD and logistic regression analyses (Table 4). The low LD between *DPB1* and other class II loci is likely a result of the high level of recombination in the region (18). Interestingly, previous studies found that the *DPB1**03:01 allele was positively associated with risk of nodular sclerosing Hodgkin lymphoma (NSHL) (19, 20). Opposite effects with the same HLA alleles on the risk of FL and NSHL also was observed for the *DRB1**15:01 - *DQA1**01:02 - *DQB1**06:02 haplotype (high risk for NSHL, low risk for FL) (21). These findings suggest that HLA class II alleles may modulate risk for NSHL and FL in a divergent manner.

Although non-significant, an inverse association with FL risk was found for *DPB1**20:01, an allele closely related to *DPB1**03:01, and positive associations were found with *DPB1**06:01 and *DPB1**13:01 (Table 2). Examining these alleles at the amino acid level reveals that the *DPB1**03:01 and *DPB1**20:01 alleles that are overrepresented in controls possess a glutamic acid rather than a lysine residue at position 69. These amino acids are oppositely charged, and reside in binding pocket 4, suggesting this change may impact *DPB1* binding. Serological groupings may also be relevant at this locus (22). Characterizing each allele by *DPB1* serological group revealed the DP3 group, containing the 56E and 85–87EAV sequence, represents only 11.8% of case alleles compared to 19.6% of control alleles. If validated, this may indicate a role for anti-DP serological activity in the etiology of FL.

The present study also confirmed our previous report based on a tag SNP analysis (7) that the *DRB1**01:01-*DQA1**01:01-*DQB1**05:01 haplotype was associated with a two-fold increased risk of FL, with *DQB1**05:01 being the most significantly associated allele in the risk haplotype. There is some indication that the risk haplotype includes *DRB1**01:02 and *01:03 (Table 5), though this finding will require replication in independent studies.

We further investigated the inverse associations between the *DQB1**06 and *DRB1**13 alleles and FL risk. As previously described in Caucasians (23), we found that *DRB1**13 was in strong LD with *DQB1**06:03, *06:04 and *06:09 alleles, whereas *DQB1**06:02 (the most common *DQB1**06 allele) was in high LD with *HLA-DRB1**15 (Table 3). Thus, we

observed a decreased risk of FL with all alleles and haplotypes containing *DRB1**13 or *15 and *DQB1**06, with *DQA1**01:02 or *01:03 (Table 6). Due to the extensive LD across *DRB1*, *DQA1*, and *DQB1*, it is unclear which loci drive these haplotype-disease associations. However, *DRB1**13 did not affect FL risk after adjustment for other HLA alleles in logistic regression analyses suggesting that this association may be the result of confounding by other HLA alleles.

We also demonstrated that carriers of the *DRB1**15:01-*DQA1**01:02-*DQB1**06:02 haplotype harbor the rs2647012 variant, which was previously reported as a protective allele for FL (9). This haplotype may be a causal variant driving the observed rs2647012 association with FL. Because there remained a modest reduction in FL risk for rs2647012 after adjusting for *DRB1**15:01-*DQA1**01:02-*DQB1**06:02, larger studies will be needed to determine the independent role of rs2647012 and the haplotype in disease risk. We further investigated LD between the HLA Class I GWAS SNP, rs6457327 (8), and HLA class I alleles. Here, we found that the *C**07:02 and *B**07:02 alleles were in LD with the protective rs6457327 A allele. However, individuals with rs6457327 A had approximately the same risk regardless of *C**07:02 and *B**07:02 status, suggesting the role for a yet unidentified causal locus that is in LD with rs6457327.

HLA class II alleles may influence FL risk through several modes of action including effects on of T-cell activation, antigen presentation of infectious or tumor-associated peptides, and HLA protein/gene expression. FL and Burkitt lymphoma disrupt normal HLA class II-mediated antigen presentation by B-cells and dendritic cells to CD4+ T-cells as a mechanism to hinder their recognition by the immune system (24). Under-expression of HLA class II on HL Reed-Sternberg cells is an independent adverse prognostic factor in classical HL (25), and loss of HLA class II expression on DLBCL tumor cells has been associated with poor survival (26). Further studies will be needed to clarify the functional role of HLA alleles in lymphomagenesis, which will likely expand our knowledge of the deregulated cellular processes that drive FL and its progression.

Use of cancer registry rapid case ascertainment and SEER abstracts to identify newly diagnosed NHL patients helped to diminish selection and participation bias in our study population, although patients with aggressive disease and poor prognosis are likely under-represented. However, as FL in general is a more indolent lymphoma, effects of survival bias on case participation should not have affected these analyses. Further, bias effects were diminished by the high participation rate for biospecimen collection in participants (~87%). The small number of non-white participants precluded analyses by race and ethnicity. Despite evidence of internal consistency in the magnitude and direction of many of our results, we had low power to test associations for low frequency variants and results from analyses with few 'exposed' should be interpreted conservatively and require validation in further studies.

In conclusion, these studies provide additional evidence that HLA alleles play essential roles in the pathogenesis of FL. As our findings show, this involves complex, multi-locus effects that span the HLA region. Because of the extensive and complex LD patterns within this region, studies in FL case-control populations from non-Caucasian ancestral pedigrees are underway that may help to distinguish between primary (causal) and secondary HLA signals. Because the causative alleles could be in non-coding (nc) regions that effect gene expression, studies are currently underway to test differential allelic gene expression of ncSNPs in high LD with HLA susceptibility alleles. Moreover, the contribution of HLA alleles in the pathogenesis of FL and other subtypes of NHL is a major focus of future studies within InterLymph where the HLA alleles identified here and in other independent case-control studies of NHL will be tested for further validation. Thus, we anticipate that

substantial progress will be made in the near future that will help to elucidate the genetic basis of NHL. Such data will likely highlight pathways and components that may be amenable to therapeutic modulation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institutes of Health grants CA122663, CA154643-01A1 and CA104682 (C.F.S.), and grants CA45614 and CA89745 from the National Cancer Institute, National Institutes of Health (P.M.B).

REFERENCES

1. Johnson PW, Rohatiner AZ, Whelan JS, et al. Patterns of survival in patients with recurrent follicular lymphoma: a 20-year study from a single center. *J Clin Oncol.* 1995; 13:140–147. [PubMed: 7799014]
2. Bastion Y, Sebban C, Berger F, et al. Incidence, predictive factors, and outcome of lymphoma transformation in follicular lymphoma patients. *J Clin Oncol.* 1997; 15:1587–1594. [PubMed: 9193357]
3. Horning SJ, Rosenberg SA. The natural history of initially untreated low-grade non-Hodgkin's lymphomas. *N Engl J Med.* 1984; 311:1471–1475. [PubMed: 6548796]
4. Fitzgibbon J, Iqbal S, Davies A, et al. Genome-wide detection of recurring sites of uniparental disomy in follicular and transformed follicular lymphoma. *Leukemia.* 2007; 21:1514–1520. [PubMed: 17495976]
5. Leich E, Salaverria I, Bea S, et al. Follicular lymphomas with and without translocation t(14;18) differ in gene expression profiles and genetic alterations. *Blood.* 2009; 114:826–834. [PubMed: 19471018]
6. O'Riain C, O'Shea DM, Yang Y, et al. Array-based DNA methylation profiling in follicular lymphoma. *Leukemia.* 2009; 23:1858–1866. [PubMed: 19587707]
7. Conde L, Halperin E, Akers NK, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat Genet.* 2010; 42:661–664. [PubMed: 20639881]
8. Skibola CF, Bracci PM, Halperin E, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. *Nat Genet.* 2009; 41:873–875. [PubMed: 19620980]
9. Smedby KE, Foo JN, Skibola CF, et al. GWAS of Follicular Lymphoma Reveals Allelic Heterogeneity at 6p21.32 and Suggests Shared Genetic Susceptibility with Diffuse Large B-cell Lymphoma. *PLoS Genet.* 2011; 7:e1001378. [PubMed: 21533074]
10. Akers NK, Curry JD, Conde L, Bracci PM, Smith MT, Skibola CF. Association of HLA-DQB1 alleles with risk of follicular lymphoma. *Leuk Lymphoma.* 2011; 52:53–58. [PubMed: 21133715]
11. Wang SS, Abdou AM, Morton LM, et al. Human leukocyte antigen class I and II alleles in non-Hodgkin lymphoma etiology. *Blood.* 2010
12. Bentley G, Higuchi R, Hoglund B, et al. High-resolution, high-throughput HLA genotyping by next-generation sequencing. *Tissue Antigens.* 2009; 74:393–403. [PubMed: 19845894]
13. Holcomb CL, Höglund B, Anderson MW, et al. A Multi-Site Study Employing High Resolution HLA Genotyping by Next Generation Sequencing. *Tissue Antigens.* in press.
14. Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 2005; 437:376–380. [PubMed: 16056220]
15. Lancaster AK, Single RM, Solberg OD, Nelson MP, Thomson G. PyPop update--a software pipeline for large-scale multilocus population genomics. *Tissue Antigens.* 2007; 69 Suppl 1:192–197. [PubMed: 17445199]
16. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 2005; 1:47–50. [PubMed: 19325852]

17. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. 1992; 48:361–372. [PubMed: 1637966]
18. Begovich AB, McClure GR, Suraj VC, et al. Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. *J Immunol*. 1992; 148:249–258. [PubMed: 1727870]
19. Oza AM, Tonks S, Lim J, Fleetwood MA, Lister TA, Bodmer JG. A clinical and epidemiological study of human leukocyte antigen-DPB alleles in Hodgkin's disease. *Cancer Res*. 1994; 54:5101–5105. [PubMed: 7923125]
20. Taylor GM, Gokhale DA, Crowther D, et al. Further investigation of the role of HLA-DPB1 in adult Hodgkin's disease (HD) suggests an influence on susceptibility to different HD subtypes. *Br J Cancer*. 1999; 80:1405–1411. [PubMed: 10424743]
21. Klitz W, Aldrich CL, Fildes N, Horning SJ, Begovich AB. Localization of predisposition to Hodgkin disease in the HLA class II region. *Am J Hum Genet*. 1994; 54:497–505. [PubMed: 8116619]
22. Cano P, Fernandez-Vina M. Two sequence dimorphisms of DPB1 define the immunodominant serologic epitopes of HLA-DP. *Hum Immunol*. 2009; 70:836–843. [PubMed: 19635517]
23. Klitz W, Maiers M, Spellman S, et al. New HLA haplotype frequency reference standards: high-resolution and large sample typing of HLA DR-DQ haplotypes in a sample of European Americans. *Tissue Antigens*. 2003; 62:296–307. [PubMed: 12974796]
24. Amria S, Cameron C, Stuart R, Haque A. Defects in HLA class II antigen presentation in B-cell lymphomas. *Leuk Lymphoma*. 2008; 49:353–355. [PubMed: 18231926]
25. Diepstra A, van Imhoff GW, Karim-Kos HE, et al. HLA class II expression by Hodgkin Reed-Sternberg cells is an independent prognostic factor in classical Hodgkin's lymphoma. *J Clin Oncol*. 2007; 25:3101–3108. [PubMed: 17536082]
26. Rimsza LM, Farinha P, Fuchs DA, Masoudi H, Connors JM, Gascoyne RD. HLA-DR protein status predicts survival in patients with diffuse large B-cell lymphoma treated on the MACOP-B chemotherapy regimen. *Leuk Lymphoma*. 2007; 48:542–546. [PubMed: 17454596]

HLA class I and II allele counts, odds ratios and 95% confidence intervals (with p-values < 0.05) in follicular lymphoma cases and controls (2n =444 cases, 440 controls) from a population-based case-control study of non-Hodgkin lymphoma in the San Francisco Bay Area.

Table 1

Allele	Case alleles (freq)	Control alleles (freq)*	OR (95% CI)	p-value	[†] Bonf. p
C*04:01	70 (0.159)	44 (0.102)	1.67 (1.1–2.56)	1.56E-02	2.81E-01
C*06:02	48 (0.109)	24 (0.056)	2.08 (1.22–3.62)	4.50E-03	8.10E-02
C*07:02	38 (0.086)	61 (0.141)	0.58 (0.36–0.9)	1.38E-02	2.48E-01
C*16:01	7 (0.016)	18 (0.042)	0.37 (0.13–0.95)	2.56E-02	4.60E-01
B*07:02	34 (0.077)	58 (0.133)	0.54 (0.34–0.87)	7.95E-03	2.86E-01
B*35:01	43 (0.097)	26 (0.06)	1.7 (1–2.94)	4.45E-02	1.0
B*50:01	8 (0.018)	1 (0.002)	8 (1.07–356.03)	3.81E-02	1.0
DRB1*01:01	54 (0.122)	31 (0.07)	1.84 (1.13–3.02)	1.17E-02	3.16E-01
DRB1*01:02	17 (0.038)	5 (0.011)	3.48 (1.22–12.15)	1.56E-02	4.21E-01
DRB1*15:01	41 (0.093)	71 (0.161)	0.53 (0.34–0.81)	2.37E-03	6.40E-02
DRB5*01:01	42 (0.095)	71 (0.161)	0.55 (0.35–0.83)	3.43E-03	3.09E-02
DRB3/4/5*ABSENT	99 (0.224)	60 (0.136)	1.83 (1.27–2.65)	8.41E-04	7.57E-03
DQA1*01:01	97 (0.22)	55 (0.125)	1.98 (1.36–2.9)	2.39E-04	1.67E-03
DQA1*01:02	64 (0.145)	101 (0.23)	0.57 (0.4–0.82)	1.82E-03	1.27E-02
DQB1*03:03	29 (0.066)	15 (0.034)	1.98 (1.01–4.03)	4.33E-02	7.37E-01
DQB1*05:01	89 (0.201)	43 (0.098)	2.31 (1.54–3.51)	1.90E-05	3.23E-04
DQB1*06:02	41 (0.093)	69 (0.158)	0.55 (0.35–0.84)	4.21E-03	7.15E-02
DPB1*03:01	19 (0.043)	46 (0.105)	0.39 (0.21–0.68)	4.61E-04	8.30E-03
DPB1*17:01	2 (0.005)	9 (0.02)	0.22 (0.02–1.06)	3.71E-02	6.68E-01
DPB1*20:01	1 (0.002)	8 (0.018)	0.12 (0–0.92)	2.07E-02	3.72E-01

* Odds ratios (OR), 95% confidence interval (95% CI), and p-values were obtained using the 'fisher.test' function from the 'stat' package in R.

[†] p-values were adjusted for the number of alleles tested using a Bonferroni correction.

HLA-DPB1 allele counts, odds ratios and 95% confidence intervals in follicular lymphoma cases and controls (2n =444 cases, 440 controls) from a population-based case-control study of non-Hodgkin lymphoma in the San Francisco Bay Area. Only alleles with p-values < 0.05 are shown.

Table 2

Allele	Case count (freq.)	Control Count (freq.)	*OR (95% CI)	*P-value	†Bonf. p
<i>DPB1</i> *01:01	22 (0.050)	25 (0.057)	0.87 (0.46–1.64)	6.56E-01	1.0
<i>DPB1</i> *02:01	79 (0.179)	66 (0.150)	1.23 (0.85–1.79)	2.76E-01	1.0
<i>DPB1</i> *03:01	19 (0.043)	46 (0.105)	0.39 (0.21–0.68)	4.61E-04	8.30E-03
<i>DPB1</i> *04:01	193 (0.437)	181 (0.411)	1.11 (0.84–1.46)	4.54E-01	1.0
<i>DPB1</i> *04:02	51 (0.115)	46 (0.105)	1.12 (0.72–1.75)	6.67E-01	1.0
<i>DPB1</i> *05:01	7 (0.016)	6 (0.014)	1.16 (0.33–4.23)	1.00E+00	1.0
<i>DPB1</i> *06:01	10 (0.023)	3 (0.007)	3.37 (0.86–19.2)	8.99E-02	1.0
<i>DPB1</i> *09:01	1 (0.002)	1 (0.002)	1 (0.01–78.27)	1.00E+00	1.0
<i>DPB1</i> *10:01	9 (0.020)	7 (0.016)	1.29 (0.42–4.1)	8.02E-01	1.0
<i>DPB1</i> *11:01	11 (0.025)	8 (0.018)	1.38 (0.5–3.99)	6.44E-01	1.0
<i>DPB1</i> *13:01	17 (0.038)	7 (0.016)	2.47 (0.96–7.13)	6.04E-02	1.0
<i>DPB1</i> *14:01	7 (0.016)	8 (0.018)	0.87 (0.27–2.77)	8.02E-01	1.0
<i>DPB1</i> *15:01	3 (0.007)	4 (0.009)	0.75 (0.11–4.43)	7.25E-01	1.0
<i>DPB1</i> *16:01	3 (0.007)	4 (0.009)	0.75 (0.11–4.43)	7.25E-01	1.0
<i>DPB1</i> *17:01	2 (0.005)	9 (0.020)	0.22 (0.02–1.06)	3.71E-02	6.68E-01
<i>DPB1</i> *19:01	2 (0.005)	4 (0.009)	0.5 (0.05–3.48)	4.51E-01	1.0
<i>DPB1</i> *20:01	1 (0.002)	8 (0.018)	0.12 (0–0.92)	2.07E-02	3.72E-01
<i>DPB1</i> *23:01	2 (0.005)	6 (0.014)	0.33 (0.03–1.85)	1.77E-01	1.0

* Odds ratios (OR), 95% confidence interval (95% CI), and p-values were obtained using the 'fisher.test' function from the 'stat' package in R.

† p-values were adjusted for the number of alleles tested using a Bonferroni correction.

Table 3

Linkage equilibrium (LD) values in the control population for selected allele combinations spanning the HLA region. LD values were estimated with the Pypop software (<http://www.pypop.org/>).

Allele 1	Allele 2	Haplotype Freq.	Observed	Expected	D'	ChiSq
rs10484561 'G'	<i>DQA1</i> *05:01	0.111	18.0	2.1	1.0	152
	<i>DQA1</i> *01:01	0.111	18.0	2.7	1.0	116
	<i>DRB1</i> *01:01	0.086	14.0	1.7	0.93	113
	<i>DRB3/4/5</i> *ABSENT	0.111	18.0	2.8	1.0	111
rs6457327 'A'	<i>C</i> *07:02	0.134	22.0	7.9	0.93	45
	<i>B</i> *07:02	0.122	20.0	6.8	1.0	44
	<i>C</i> *03:04	0.049	8.0	2.7	1.0	16
	<i>B</i> *44:03	0.049	8.0	2.7	1.0	16
	<i>C</i> *02:02	0.043	7.0	2.4	1.0	14
	<i>C</i> *16:01	0.043	7.0	2.4	1.0	14
rs2647012 'A'	<i>DQA1</i> *01:02	0.191	31.3	15.1	0.86	39
	<i>DQB1</i> *06:02	0.159	26	11.6	1.0	39
	<i>DRB1</i> *15:01	0.159	26	11.6	1.0	39
	<i>DRB5</i> *01:01	0.159	26	11.6	1.0	39
	<i>DRB1</i> *03:01	0.152	25	11.1	1.0	37
	<i>DQB1</i> *02:01	0.159	26	12.0	0.96	35
<i>DPB1</i> *03:01 (46)	rs10484561*T	0.108	17.5	16.9	0.30	0.23
	rs2647012*G	0.069	11.4	11.1	0.03	0.02
	rs6457327*C	0.082	13.5	13.2	0.04	0.02
<i>DQB1</i> *06	<i>DRB1</i> *15	0.174	76.0	22.0	0.95	224
	<i>DRB1</i> *15:01	0.158	69.0	11.2	1.0	423
	<i>DQB1</i> *15:02	0.018	8	0.1	1.0	438
<i>DQB1</i> *06	<i>DRB1</i> *13	0.098	43	13.9	0.8	95
	<i>DRB1</i> *13:02	0.032	14.0	0.7	0.93	281
	<i>DRB1</i> *13:02	0.014	6	0.3	1.0	127

Allele 1	Allele 2	Haplotype Freq.	Observed	Expected	D'	ChiSq
<i>DQB1*06:14</i>	<i>DRB1*13:01</i>	0.005	2	0.1	1.0	38
<i>DQB1*06:03</i>	<i>DRB1*13:05</i>	0.002	1	0.04	1.0	21

Best fitting model from multivariable backward stepwise logistic regression of known HLA risk loci* associated with follicular lymphoma risk.

Table 4

Allele*	Odds Ratio	Std. Error	95% CI	z-score	p-value
<i>DPB1</i> *03:01	0.29	0.11	0.14 – 0.6	–3.32	0.001
<i>DQB1</i> *05:01	1.75	0.56	0.94 – 3.29	1.75	0.08
rs6457327	0.66	0.14	0.44 – 1.01	–1.93	0.054
rs2647012 + <i>DRB1</i> *15	0.53	0.10	0.37 – 0.75	–3.54	<0.001

* HLA alleles and SNPs coded as present/absent (a dominant effect model) included rs6457327, rs2647012, *DQB1**05:01 (as a proxy for the rs10484561-*DRB1**01:01-*DQA1**01:01-*DQB1**05:01 haplotype), *DRB1**15 (as a proxy for *DRB1**15-*DQA1**01-*DQB1**06), *DRB1**13 (as a proxy for *DRB1**13 - *DQA1**01 - *DQB1**03:01), and *DPB1**06). The rs2647012 and *DRB1**15 allele effects were assessed as 0 = carrier of neither allele, 1 = carrier of rs2647012, and 2 = carrier of both rs2647012 and *DRB1**15. Likelihood ratio test p-value cutoff of 0.10 was used as criteria for significant contribution to the model.

*HLA DRB1*01-DQA1*01-DQB1*05* haplotype and allele counts, odds ratios and 95% confidence intervals in follicular lymphoma cases and controls (2n=444 cases, 440 controls) from a population-based case-control study of non-Hodgkin lymphoma in the San Francisco Bay Area.

Table 5

Haplotype and Individual Alleles	Case count (freq.)	Control Count (freq.)	*OR (95% CI)	*P-value	*Bonf. p
<i>DRB1*01-DQA1*01-DQB1*05</i>	78 (0.178)	39 (0.089)	2.22 (1.45–3.43)	1.43E-04	2.23E-03
<i>DRB1*01:01</i>	54 (0.122)	31 (0.070)	1.84 (1.13–3.02)	1.17E-02	3.16E-01
<i>DRB1*01:02</i>	17 (0.038)	5 (0.011)	3.48 (1.22–12.2)	1.56E-02	4.21E-01
<i>DRB1*01:03</i>	10 (0.023)	5 (0.011)	1.99 (0.615–7.50)	2.98E-01	1.0
<i>DQA1*01:01</i>	97 (0.22)	55 (0.125)	1.98 (1.36–2.90)	2.39E-04	1.67E-03
<i>DQB1*05:01</i>	89 (0.201)	43 (0.098)	2.31 (1.54–3.51)	1.90E-05	3.23E-04

* Odd ratios (OR), 95% confidence intervals (CI); p-values are based on a Bonferroni (Bonf) correction for the number of alleles tested at each locus (16 *DRB1-DQA1-DQB1* haplotypes [2 digit], 27 *DRB1* alleles [4 digit], 7 *DQA1* alleles [4 digit], and 17 *DQB1* alleles [4 digit]).

HLA-*DRB1**15 and *DRB1**13 haplotypes and constituent alleles and risk of follicular lymphoma (2n=444 cases, 440 controls).

Table 6

Haplotype and Individual Alleles	Case Count (freq.)	Control Count (freq.)	*OR (95%CI)	*P-value	†Bonf. p
<i>DRB1</i> *15 - <i>DQA1</i> *01 - <i>DQB1</i> *06	45.0 (0.103)	76.0 (0.174)	0.55 (0.36–0.82)	3.18E-03	5.09E-02
<i>DRB1</i> *13 - <i>DQA1</i> *01 - <i>DQB1</i> *06	24.0 (0.055)	43.0 (0.098)	0.53 (0.30–0.92)	2.15E-02	3.44E-01
<i>DRB1</i> *15	45 (0.102)	79 (0.180)	0.52 (0.34–0.78)	9.60E-04	1.25E-02
<i>DRB1</i> *13	31 (0.07)	50 (0.114)	0.59 (0.36–0.96)	2.69E-02	3.50E-01
<i>DQA1</i> *01:02	64 (0.145)	101 (0.23)	0.57 (0.4–0.82)	1.82E-03	1.27E-02
<i>DQA1</i> *01:03	19 (0.043)	32 (0.073)	0.58 (0.3–1.07)	8.25E-02	5.78E-01
<i>DQB1</i> *06	73 (0.165)	122 (0.279)	0.51 (0.36–0.72)	6.52E-05	3.26E-04

* Odd ratios (OR), 95% confidence intervals (CI) and p-values were obtained using the 'fisher.test' function from the 'stat' package in R.

† P-values were adjusted for the number of alleles tested at each locus using a Bonferroni correction (16 *DRB1*-*DQA1*-*DQB1* haplotypes [2 digit], 13 *DRB1* alleles [2 digit], 7 *DQA1* alleles [4 digit], and 5 *DQB1* alleles [2 digit]).