Title: HLA class I and II diversity contributes to the etiologic heterogeneity of non-Hodgkin lymphoma subtypes

Authors: Sophia S. Wang[1,+], Mary Carrington[2], Sonja I. Berndt[3], Susan L. Slager[4], Paige M. Bracci[5], Jenna Voutsinas[1], James R. Cerhan[4], Karin E. Smedby[6,7], Henrik Hjalgrim[8,9], Joseph Vijai[10], Lindsay M. Morton[3], Roel Vermeulen[11,12], Ora Paltiel[13], Claire M. Vajdic[14], Martha S. Linet[3], Alexandra Nieters[15], Silvia de Sanjose[16,17], Wendy Cozen[18], Elizabeth E. Brown[19] , Jennifer Turner[21,22], John J. Spinelli[23,24], Tongzhang Zheng[25], Brenda M. Birmann[26], Christopher R. Flowers[27], Nikolaus Becker[28], Elizabeth A. Holly[5], Eleanor Kane[20], Dennis Weisenburger[29], Marc Maynadie[30], Pierluigi Cocco[31], Demetrius Albanes[3], Stephanie J. Weinstein[3], Lauren R. Teras[32], W. Ryan Diver[32], Stephanie J. Lax[20], Ruth C. Travis[33], Rudolph Kaaks[28], Elio Riboli[34], Yolanda Benavente[16,17], Paul Brennan[35], James McKay[36], Marie-Hélène Delfau-Larue[36,37], Brian K. Link[38], Corrado Magnani[39], Maria Grazia Ennas[40], Giancarlo Latte[41], Andrew L. Feldman[42], Nicole Wong Doo[43], Graham G. Giles[43,44], Melissa C. Southey[45], Roger L. Milne[43,44], Kenneth Offit[9], Jacob Musinsky[9], Alan A. Arslan[46,47,48], Mark P. Purdue[3], Hans-Olov Adami[50], Mads Melbye[8,51], Bengt Glimelius[52], Lucia Conde[53], Nicola J. Camp[54], Martha Glenn[54], Karen Curtin[54], Jacqueline Clavel[55,56], Alain Monnereau[55,56,57], David G. Cox[58], Hervé Ghesquières[59,60], Gilles Salles[60,61], Paulo Bofetta[62], Lenka Foretova[63], Anthony Staines[64], Scott Davis[65], Richard K. Severson[66], Qing Lan[3], Angela Brooks-Wilson[67,68], Martyn T Smith[69], Eve Roman[20], Anne Kricker[70], Yawei Zhang[71,72], Peter Kraft[73,74], Stephen J. Chanock[3], Nathaniel Rothman[3], Patricia Hartge[3], Christine F. Skibola[27]

Author affiliations:
[1]Department of Population Sciences, Beckman Research Institute and the City of Hope, Duarte, California (Wang, Voutsinas);
[2]Cancer and Inflammation Program, Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, MD (Carrington) and Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA;
[3] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD (Berndt, Morton, Linet, Albanes, Weinstein, Lan, Chanock, Rothman, Hartge, Purdue);
[4]Department of Health Sciences Research, Mayo Clinic, Rochester, MN (Slager, Cerhan);
[5]Department of Epidemiology and Biostatistics, University of California–San Francisco, San Francisco, CA (Bracci, Holly);
[6]Department of Medicine Solna, unit of clinical epidemiology, Karolinska Institutet, Stockholm, Sweden (Smedby);
[7]Hematology Center, Karolinska University Hospital, Stockholm, Sweden (Smedby);
[8]Department of Epidemiology Research, Division of Health Surveillance and Research, Statens Serum Institut, Copenhagen, Denmark (Hjalgrim, Melbye);
[9]Department of Hematology, Rishospitalet, Copenhagen, Denmark (Hjalgrim);
[10]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York City, NY (Vijai, Offit, Musinsky);
[11]Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands (Vermeulen);

[12]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands (Vermeulen);
[13]Braun School of Public Health and Community Medicine, Hadassah-Hebrew University Medical Center, Jerusalem, Israel (Paltiel);
[14]Centre for Big Data Research in Health, The University of New South Wales, Sydney, New South Wales, Australia (Vajdic);
[15]Centre for Chronic Immunodeficiency, University Medical Center Freiburg, Freiburg, Baden-Württemberg, Germany (Nieters);
[16]Unit of Infections and Cancer, Cancer Epidemiology Research Programme, Institut Català d' Oncologia, IDIBELL, 08908 L'Hospitalet de Llobregat, Barcelona, Spain (de Sanjose, Benavente);
[17]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain (de Sanjose, Benavente);
[18]Norris Comprehensive Cancer Center, Keck School of Medicine of USC, Departments of Preventive Medicine and Pathology, University of Southern California, Los Angeles, CA (Cozen);
[19]Department of Pathology, School of Medicine and the UAB Comprehensive Cancer Center, The University of Alabama at Birmingham, Birmingham, AL (Brown);
[20]Department of Health Sciences, University of York, York, United Kingdom (Kane, Lax, Roman);
[21]Department of Histopathology, Douglass Hanly Moir Pathology, Sydney, Australia (Turner);
[22]Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia (Turner);
[23]Cancer Control Research, British Columbia Cancer Agency, Vancouver, British Columbia, Canada (Spinelli);
[24]School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada (Spinelli);
[25]Department of Epidemiology, School of Public Health, Brown University, Providence, RI (Zheng);
[26]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA (Birmann);
[27]Department of Hematology and Medical Oncology, Emory University School of Medicine, Atlanta, GA (Flowers, Skibola);
[28]Division of Clinical Epidemiology, German Cancer Research Centre, Heidelberg, Baden-Württemberg, Germany (Becker, Kaaks);
[29]Department of Pathology, City of Hope, Duarte, California (Weisenburger);
[30]Registry of Hematological Malignancies of Cote d'Or, INSERM UMR1231, University of Burgundy and Dijon University Hospital, Dijon, France (Maynadie);
[31]Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Cagliari, Italy (Cocco);
[32]Epidemiology Research Program, American Cancer Society, Atlanta, GA (Teras, Diver);
[33]Cancer Epidemiology Unit, University of Oxford, Oxford, United Kingdom (Travis, McKay);
[34]School of Public Health, Imperial College London, London, United Kingdom (Riboli);
[35]International Agency for Research on Cancer, Lyon, France (Brennan, McKay);
[36]Department of Immunology, CHU Henri Mondor, Créteil, France (Delfau-Larue)
[37]INSERM U 955, CHU Henri Mondor, Créteil, France (Delfau-Larue);

[38]Department of Internal Medicine, Carver College of Medicine, The University of Iowa, Iowa City, IA (Link);
[39]Center of Oncological Prevention (CPO) Piemonte and Unit of Medical Statistics and Epidemiology, Department Translational Medicine, University of Piemonte Orientale, Novara, Italy (Magnani);
[40]Department of Biomedical Science, University of Cagliari, Monserrato, Cagliari, Italy (Ennas);
[41]Division of Hematology, S. Francesco Hospital, Nuoro, Italy (Latte);
[42]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN (Feldman);
[43]Cancer Epidemiology & Intelligence Division, Cancer Council Victoria, Melbourne, Australia (Wong Doo, Giles, Milne);
[44]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Victoria, Australia (Giles, Milne);
[45]Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne, Melbourne, Victoria, Australia (Southey);
[46]Department of Obstetrics and Gynecology, New York University School of Medicine, New York City, NY (Arslan);
[47]Department of Environmental Medicine, New York University School of Medicine, New York City, NY (Arslan);
[48]Perlmutter Cancer Center, NYU Langone Medical Center, New York City, NY (Arslan);
[49]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (Adami);
[50]Department of Medicine, Stanford University School of Medicine, Stanford, CA (Melbye);
[51]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden (Glimelius);
[52]Bill Lyons Informatics Centre, UCL Cancer Institute, University College London, WC1E 6DD, London, United Kingdom (Conde);
[53]Department of Internal Medicine, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT (Camp, Glenn, Curtin);
[54]Epidemiology of Childhood and Adolescent Cancers Group, Inserm, Center of Research in Epidemiology and Statistics Sorbonne Paris Cité (CRESS), Paris, France (Clavel, Monnereau);
[55]Université Paris Descartes, Paris, France (Clavel, Monnereau);
[56]Registre des hémopathies malignes de la Gironde, Institut Bergonié, University of Bordeaux, Inserm, Team EPICENE, UMR 1219, France (Monnereau);
[57]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom. Cancer Research Center of Lyon, INSERM UMR1052, Center Léon Bérard, Lyon, France (Cox);
[58]Department of Hematology, Centre Léon Bérard, Lyon, France (Ghesquières);
[59]Laboratoire de Biologie Moléculaire de la Cellule UMR 5239, Centre National de la Recherche Scientifique, Pierre benite Cedex, France (Ghesquières, Salles);
[60]Department of Hematology, Hospices Civils De Lyon, Centre Hospitalier Lyon-Sud and Université Claude Bernard, Lyon, France (Salles);
[61]The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York City, New York (Bofetta);
[62]Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, MF MU, Brno, Czech Republic (Foretova);

[63]School of Nursing and Human Sciences, Dublin City University, Dublin, Ireland (Staines);
[64]Fred Hutchinson Cancer Research Center & School of Public Health and Community Medicine, University of Washington, Seattle, WA (Davis);
[65]Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, MI (Severson);
[66]Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada (Brooks-Wilson);
[67]Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada (Brooks-Wilson);
[68]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, CA (Smith);
[69]Sydney School of Public Health, The University of Sydney, Sydney, Australia (Kricker);
[70]Department of Surgery, Yale School of Medicine, New Haven, CT (Zhang);
[71]Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT (Zhang);
[72]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA (Kraft);
[73]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA (Kraft)

***Running title**: HLA gene diversity reduces risk for non-Hodgkin lymphoma.*

[+]*To whom correspondence should be addressed:*
Sophia S. Wang, Ph.D.
Division of Cancer Etiology
Department of Population Sciences
Beckman Research Institute of the City of Hope
1500 East Duarte Road
Duarte, CA 91010
Phone: (626) 471-7316
Fax: (626) 471-7308
Email: sowang@coh.org

ABSTRACT

A growing number of loci within the human leukocyte antigen (HLA) region have been implicated in non-Hodgkin lymphoma (NHL) etiology. Here, we test a complementary hypothesis of "heterozygote advantage" regarding the role of HLA and NHL, whereby HLA diversity is beneficial and homozygous HLA loci are associated with increased disease risk. HLA alleles at class I and II loci were imputed from genome-wide association studies (GWAS) using SNP2HLA for: 3,617 diffuse large B-cell lymphomas (DLBCL), 2,686 follicular lymphomas (FL), 2,878 chronic lymphocytic leukemia/small lymphocytic lymphomas (CLL/SLL), 741 marginal zone lymphomas (MZL), and 8,753 controls of European descent. Both DLBCL and MZL risk were elevated with homozygosity at class I HLA-B and -C loci (OR DLBCL=1.31, 95% CI=1.06-1.60; OR MZL=1.45, 95% CI=1.12-1.89) and class II HLA-DRB1 locus (OR DLBCL=2.10, 95% CI=1.24-3.55; OR MZL= 2.10, 95% CI=0.99-4.45). Increased FL risk was observed with the overall increase in number of homozygous HLA class II loci (p-trend<0.0001, FDR=0.0005). These results support a role for HLA zygosity in NHL etiology and suggests that distinct immune pathways may underly the etiology of the different NHL subtypes.

Precis/Statement of Significance: *HLA gene diversity reduces risk for non-Hodgkin lymphoma.*

INTRODUCTION

Genome-wide association studies (GWAS) have identified a growing list of common susceptibility loci modestly associated with risk of non-Hodgkin lymphomas (NHLs) including several *HLA* (human leukocyte antigen) genetic variants on chromosome 6p21, a region that is critical for innate and adaptive immune responses. Putative NHL susceptibility loci either directly implicate genes within the Major Histocompatibility Complex (MHC) or appear in strong linkage disequilibrium (LD) with extended *HLA* haplotypes (1-5). Interestingly, there is little convincing overlap of the identified *HLA* susceptibility loci among the NHL subtypes, diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), marginal zone lymphoma (MZL) and chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL), suggesting that disparate aspects of the MHC and resulting immune responses are involved in the etiology of each NHL subtype.

The *HLA* genes are the most polymorphic in the human genome and specific *HLA* loci determine the antigens that are bound by antigen presenting cells (e.g., B cells and dendritic cells) and presented to T cells to elicit immune responses. Functionally, HLA molecules are critical for the host immune response. HLA class I molecules present foreign antigens primarily to cytotoxic T-cells that in response kill these target cells, while HLA class II molecules stimulate antibody production in response to specific antigens.

Reduced diversity, as defined by homozygosity at each co-dominant *HLA* loci, might adversely affect the host's ability to recognize a more diverse array of foreign antigens and thereby increase subsequent disease burden. This concept is supported by *a priori* research that

has examined effects of *HLA* zygosity on infectious disease, whereby a lack of *HLA* class I and II diversity has been associated with increased risk HIV and hepatitis B virus infection (6-8).

Given the growing evidence that genetic variation within *HLA* genes play in the etiology of NHL subtypes (1-4, 9), we specifically aimed to test whether lack of *HLA* diversity - as measured by *HLA* homozygosity – was associated with increased NHL risk. Specifically, we posit that associations with HLA Class II, which primarily presents peptides derived from extracellular sources, would implicate a role in infectious disease etiology. On the other hand, associations with HLA Class I, which primarily presents peptides derived from intracellular sources, would suggest a role in related conditions, such as autoimmune or atopic conditions. We present here results from a pooled analysis of 25 studies from North America, Europe, and Australia where we measured the associations between *HLA* class I and/or class II zygosity and four main NHL subtypes.

MATERIALS AND METHODS

Study sample. Our study sample comprises the same study participants of European descent that were included in the original GWAS efforts from which 25 studies participated. Specifically, adults diagnosed with incident, non-HIV-related B-cell NHL of mostly European descent, ascertained from cancer registries, clinics, or hospitals or through self-report were included and where diagnoses were verified by medical and pathology reports (1-4). Study designs included prospective cohort studies, population- and hospital-based case-control studies, and clinic-based studies. Original details of design methods for each study and of each GWAS have been described previously (1-4).

This study was approved by the City of Hope Institutional Review Board. Each participating study obtained approval from human subjects review committees and written informed consent from all participants. A de-identified pooled dataset with individual-level data on genotypes, demographic characteristics, and NHL subtypes of cases was provided by the InterLymph Data Coordinating Center (Mayo Clinic, Rochester, MN).

Genotyping. GWAS platforms used include the Illumina 317K, Illumina HumanHap 610K, Illumina HumanHap 660W, Illumina Human CNV370-Duo BeadChip, Affymetrix SNP 6.0, and the Illumina OmniExpress (**Table 1**). Quality control metrics employed (e.g., QQ plots and Eigenstrat results) and main results of each GWAS have been previously described in-depth (1-4).

HLA imputation. As reported by Skibola et al (2), classical *HLA* alleles were imputed at *HLA* class I (*HLA-A*, *HLA-B*, *HLA-C*) and class II loci (*HLA-DQA1*, *HLA-DQB1*, *HLA-DRB1*, *HLA-DPA1*, *HLA-DPB1*) using SNP2HLA and a reference panel from the Type 1 Diabetes Genetics Consortium that comprised 5,225 individuals of European descent who were typed for *HLA-A, B, C, DQA1, DQB1, DRB1, DPA1, DPB1* 4 digit alleles. We note that the SNP2HLA reference panel is typed both for a panel of MHC SNPs and using classical *HLA* typing; the imputation algorithms used thus rely on both methodologies particularly when only SNPs are available. A comparison of imputed *HLA* alleles to 4-digit *HLA* sequencing data available for a subset of samples showed high concordance: *HLA-A* (97.3%), *B* (98.5%), *C* (98.1%) and *DRB1* (97.5%). In all, 201 classical *HLA* alleles (two- and four-digit resolution) were successfully imputed (info score r^2>0.3 for alleles) and available for analysis. Because of the strong LD between the *HLA* class II A1 and B1 loci (e.g., *HLA-DQA1* and *DQB1*), we present results for

each of the B1 loci (*HLA-DQB1*, *HLA-DRB1*, *HLA-DPB1)* since there were fewer homozygous B1 loci than A1 loci. For each *HLA* locus, individuals were coded as homozygote (for any allele) or heterozygote, as determined from the imputed alleles. All results presented are based on four-digit resolution.

NHL Classification. NHL subtypes were harmonized at the InterLymph Data Coordinating Center using the InterLymph Pathology Working Group guidelines (10,11), which are based on the World Health Organization classification (12).

Final analytic sample. Data for *HLA* loci were directly imputed from the original GWAS SNP panels and evaluated for the 3,617 DLBCL, 2,686 FL, 2,878 CLL/SLL, 741 MZL, and 8,753 controls. We note that, as with the original GWAS manuscripts, the specific numbers of controls differed by NHL subtype, due to different study inclusion and control selection criteria for each NHL subtype analyses, as described by the original GWAS publications (enumerated in **Table 2**).

Statistical analysis. Heterozygosity and homozygosity at each individual *HLA* locus and the number of homozygous loci for class I loci (*A, B, C*) and class II loci (*DQB1, DRB1, DPB1*) were determined; odds ratios (ORs) and 95% confidence intervals (CIs) were calculated as estimates of NHL risk with heterozygotes as the referent category, adjusted for sex, age, study, GWAS platform, and ancestry (with principal components as conducted for each subtype-specific GWAS and previously published (1-4). For analyses of MZL, adjustment by geographic region was conducted due to sample size restrictions (instead of by individual study). In addition to calculating the risk estimates for each additional number of homozygous loci, we further calculated the p-trend.

To further describe associations of zygosity by loci, we conducted joint effects analyses for *HLA* class I loci and class II loci. Each *HLA* loci (class I or II) was conducted in a stratified manner whereby heterozygotes for all loci were the referent groups and all combinations of homozygosity among the loci were evaluated. For example, to pinpoint whether *HLA* class I associations were attributable to *HLA* Class I B or C loci, we modeled as one covariate, 4 levels/combinations for *HLA-B* and -*C* (e.g., homozygous for both *HLA-B* and –*C*, homozygous only for *HLA-B*, homozygous for only *HLA-C*, and heterozygous for both), with heterozygote for both *HLA-B* and -*C* as reference (**Table 3**). For the associated p-trends reported in **Table 3**, each category is modeled based on ordinal variable in the order listed in the table, with heterozygosity at all loci as the referent group in a logistic regression model. For each p-trend, we also present the linearized additive relative-risk-per-locus, reflecting the slope of the trend-line.

Platform-specific results are shown in a **Supplemental Table 1**. Additional sensitivity analysis included evaluation of potential confounders, including evaluation of associations by previously implicated autoimmune conditions and *HLA* loci associated with specific NHL subtypes. We conducted stratified analysis to evaluate whether *HLA* zygosity associations were present among participants with and without autoimmune conditions (generally, and by specific conditions); similarly stratified analyses were conducted among participants with and without previously identified SNPs associated with NHL subtypes. We further calculated the risks, adjusting for autoimmune conditions and for all reported genetic susceptibility loci (for each NHL subtype). As neither variable altered the odds ratio >10%, those data are not presented. Analyses that restricted studies to population-based controls only also did not have measurable effect on the results. Finally, to evaluate the probability that some of our results could be due to chance, we used the Benjamini-Hochberg method to calculate the false discovery rate (FDR) and

applied it to the p-trends as this allows for the fewest number of comparisons and thus degrees of freedom to assess the additive model.

Unconditional logistic regression models were applied using SAS 9.4 (SAS Institute). All tests of statistical significance were 2-sided.

RESULTS

The numbers of European cases and controls from each of the 25 studies in North America, Europe, and Australia for which *HLA* class I and II loci were evaluated are detailed in **Table 1**.

DLBCL. Elevated DLBCL risks of 20-50% were observed for homozygosity for individual *HLA* class I (*B* and *C*) and/or class II loci (*DRB1* and *DQB1*) (**Table 2**). DLBCL risk also increased with increasing number of homozygous class I loci (p-trend=0.0008; FDR p=0.003) and class II loci (p-trend<0.0001; FDR p=0.0005) (**Table 2**). Although homozygosity for *HLA-A* had a borderline non-significant effect for increasing DLBCL risk, joint analyses suggested that the 30% risk increase observed with two or more homozygote loci (**Table 2**) was attributable to homozygosity at the *HLA-B* and *-C* locus (OR=1.31, 95% CI=1.06-1.60, **Table 3**). Similarly, for class II loci, joint analysis showed statistically significant associations for homozygosity specifically at the *HLA-DRB1* locus (OR=2.10, 95% CI=1.24-3.55) as significantly increased risk was observed only in combination with homozygous *HLA-DRB1* locus (**Table 3**).

FL. There were no significant associations between zygosity at *HLA* class I loci and FL risk (**Table 2**). Statistically significant 24-54% increases, however, were observed for FL risk for

each of the three *HLA* class II loci. Further, FL risk increased with the total number of homozygous *HLA* class II loci (p-trend<0.0001; FDR p=0.0005), with an odds ratio of 1.89 (95% CI=1.37-2.61) for those fully homozygous compared with those fully heterozygous at all three *HLA* class II loci. Joint analyses additionally supported a statistically significant increased risk for FL with overall homozygosity at the *HLA* class II loci (p-trend<0.0001; FDR p=0.0005, **Table 3**).

MZL. Homozygosity at *HLA* class I loci *HLA-B* (OR=1.34, 95% CI=1.01-1.78) and *–C* (OR=1.33, 95% CI=1.04-1.70) but not *-A* (OR=1.06, 95% CI=0.82-1.38) increased MZL risk (**Table 2**). Stratified analysis supported independent associations for both *HLA-B* and *–C* and MZL (**Table 3**). Homozygosity at *HLA* class II loci increased MZL risk (**Table 2**), but only the association with *HLA-DRB1* reached statistical significance (OR=1.45, 95% CI-1.12-1.89, **Table 2**). Analyses considering single locus homozygosity provided evidence of a role for *HLA-DRB1* in increasing MZL risk (**Table 3**).

CLL/SLL. Modest CLL/SLL risk increases were observed for *HLA-A* (OR=1.19, 95% CI=1.02-1.38), *HLA-DRB1* (OR=1.19, 95% CI=1.00-1.42) and *HLA-DQB1* (OR=1.20, 95% CI=1.03-1.39) (**Table 2**). Increasing CLL/SLL risk was not observed with increasing number of homozygote class I or class II loci, though when evaluating total numbers of class I and II loci altogether, a borderline significant increased risk was observed for those with all five homozygote class I and II loci (OR=1.57, 95% CI=1.04-2.38, p-trend = 0.029; FDR=0.055) (**Table 2**). We were unable to isolate CLL/SLL associations with *HLA* zygosity to any singular locus (**Table 3**).

DISCUSSION

Based on the largest number of NHL subtypes to date for whom imputed HLA data is available, we demonstrate that *HLA* homozygosity plays a role in four B-cell NHL subtypes, and that the associations between homozygosity at *HLA* Class I and/or Class II loci are distinct by these subtypes. Specifically, FL risk was associated with homozygosity at *HLA* class II loci, but not Class I loci. CLL/SLL risk appeared to be associated (borderline) with homozygosity at either *HLA* Class I or Class II loci. In contrast, while both DLBCL and MZL were associated with zygosity at *HLA* Class I and Class II loci, the associations appeared specific to Class I *HLA-B* and *–C* loci and to the Class II *HLA-DRB1* locus. We note that the p-trends evaluated for each additional homozygous loci remained statistically significant after adjust for multiple comparisons, with exception of that for CLL/SLL. Our results add to the growing body of literature implicating different roles for *HLA* class I and II loci, key modulators of human immune response, in the heterogeneous etiologies of B-NHL subtypes (1-4). Our results also add to the current literature which points to similarities in the etiologic profiles of DLBCL and MZL (13) . Overall, these data support the importance of *HLA* diversity in NHL etiology, with the type of *HLA* diversity potentially varying by NHL subtype.

The underlying hypothesis regarding the role of *HLA* zygosity and disease is that homozygosity at *HLA* loci reduces the diversity of peptides that can be presented, with the hypothesis that these peptides can reflect etiologic agents such as infectious diseases, self-antigens for atopic or autoimmune conditions, and even cancerous cells. At present, there is a growing body of literature supporting that *HLA* heterozygotes are more resistant to infectious diseases, and the corollary, that *HLA* homozygotes are more susceptible to infectious diseases. Specifically, *HLA* class I heterozygote advantage (e.g., presenting greater diversity of antigenic

peptides to CD8+ cytotoxic T lymphocytes) has been demonstrated for slowing progression to AIDS (6), whereas heterozygotes at *HLA* class II loci appear to have greater ability in clearing HBV infection (8) and HCV infection (14) than homozygotes. *HLA-DRB1* heterozygosity has also been reported to confer favorable outcome (e.g., against end-stage liver disease) among HCV-infected liver transplant recipients (15). There are also reports evaluating *HLA* zygosity as a key contributor in autoimmune conditions. For example, reports of heterozygote advantage for class II loci and inflammatory bowel disease (16) and for class I loci and psoriatic arthritis (17) have both been published. Specific associations between *HLA* zygosity and NHL have been limited to reports of CLL. Evidence of the importance of *HLA* zygosity include reports that homozygosity at *HLA-A, -B,* and *-DRB1* are associated with CLL (18) and with CLL disease progression (19-20), with the hypothesis that limited *HLA* diversity provided an advantage of the tumor to escape the immune response.

HLA heterozygote advantage is posited to work in concert with specific allele associations (as opposed to exclusively) (21); our results thus complement ongoing efforts that have identified the most role that specific *HLA* alleles have on NHL subtype risk. In sensitivity analysis, we evaluated the effect of known HLA associations and, in stratified and adjusted analysis, did not find that these associations diminish the reported association between *HLA* zygosity and NHL subtypes. Further evaluation into how these complementary associations act in concert are thus warranted and inclusion of *HLA* zygosity in the construct of genetic risk scores for each NHL subtype should be considered.

Further research to understand the association – or independence - between *HLA* zygosity with infections and autoimmune conditions and NHL risk are also needed (21-24). For

example, efforts to evaluate autoimmune conditions linked to class II alleles (e.g., Sjögren syndrome, systemic lupus erythromatosus, and rheumatoid arthritis) (23) with class II zygosity in relation to FL risk could provide potential insight regarding immune mechanisms modulating FL risk. A particularly pressing research question is understanding what are the underlying mechanisms of individual allele-associations and how are they distinct from *HLA* zygosity associations. Similar efforts to identify commonalities between autoimmune and infectious disease associations with *HLA* loci and zygosity among other NHL subtypes are also warranted. Finally, extension of these efforts towards understanding the genetic and structural variants and HLA expression are also required to fully understand the implication of *HLA*-allelic associations in the context of overall class I or II zygosity.

Study strengths include the large sample size available to evaluate individual NHL subtypes which no studies have been able to do adequately to date (25). Potential study limitations include possible misclassification of *HLA* alleles due to imputation, although direct comparison of a subset with genotyped *HLA* alleles showed >97% concordance (2). While the present analysis leverages the available GWAS data through imputation of *HLA* alleles, we recognize that confirmation with direct *HLA* allelotyping may provide additional levels of information not ascertained in imputed data.

Our study's restriction to individuals of European ancestry requires our results to be replicated for other racial or ethnic groups, as the associations may not apply universally to all ethnic groups. However, as demonstrated for HLA associations in autoimmune conditions, fine-mapping studies show that the same amino acid changes contribute to disease in both European and Asian populations (26), implicating similar underlying biologic mechanisms for disease

etiology. Studies limitations also include our inability to evaluate heterogeneity within NHL subtypes, either defined molecularly, by infectious etiology, or by organ site.

In summary, our results add to the growing evidence of *HLA* alleles as susceptibility loci in the etiology of B-cell NHL subtypes. In addition to ongoing fine-mapping studies being conducted as follow-up to GWAS, our results here suggest that functional studies aiming to understand the underlying biology of zygosity and NHL subtype risk will also be important. Additional efforts to evaluate larger-scale zygosity, such as of immune genes and perhaps the entire genome may prove important in understanding the full extent of the role diversity of the immune response plays in lymphoma etiology.

ACKNOWLEDGEMENTS

This analysis was initiated and conducted by the InterLymph Consortium Immunology and Infections Working Group and the Genome-Wide Association Study. We are grateful to the members of the working group, study investigators from contributing InterLymph case-control studies, and the InterLymph Consortium members for their contributions to this international collaboration. We are also thankful for the contributions from Aaron Norman and Dennis Robinson of the InterLymph Data Coordinating Center at the Mayo Clinic and to Michelle Dich at the City of Hope.

FUNDING ACKNOWLEDGEMENTS

This project was supported in part with funding from the National Institutes of Health (R01CA179558 and R01CA33572).

ATBC - The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, and by U.S. Public Health Service contract HHSN261201500005C from the National Cancer Institute, Department of Health and Human Services.

BC – Canadian Institutes for Health Research (CIHR); Canadian Cancer Society; Michael Smith Foundation for Health Research.

CPS-II - The Cancer Prevention Study-II (CPS-II) Nutrition Cohort is supported by the American Cancer Society. Genotyping for all CPS-II samples were supported by the Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. The authors would also like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention National Program of Cancer Registries, and cancer registries supported by the National Cancer Institute Surveillance Epidemiology and End Results program.

ELCCS – Bloodwise (formerly Leukaemia & Lymphoma Research), UK.

ENGELA – Association pour la Recherche contre le Cancer (ARC), Institut National du Cancer (INCa), Fondation de France, Fondation contre la Leucémie, Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES)

EPIC – Coordinated Action (Contract #006438, SP23-CT-2005-006438); HuGeF (Human Genetics Foundation), Torino, Italy; Cancer Research UK.

EpiLymph – European Commission (grant references QLK4-CT-2000-00422 and FOOD-CT-2006-023103); the Spanish Ministry of Health (grant references CIBERESP, PI11/01810, PI14/01219, RCESP C03/09, RTICESP C03/10 and RTIC RD06/0020/0095), the Marató de TV3 Foundation (grant reference 051210), the Agència de Gestió d'Ajuts Universitarisi de Recerca – Generalitat de Catalunya (grant reference 2014SRG756) who had no role in the data collection, analysis or interpretation of the results; the NIH (contract NO1-CO-12400); the Compagnia di San Paolo—Programma Oncologia; the Federal Office for Radiation Protection grants StSch4261 and StSch4420, the José Carreras Leukemia Foundation grant DJCLS-R12/23, the German Federal Ministry for Education and Research (BMBF-01-EO-1303); the Health Research Board, Ireland and Cancer Research Ireland; Czech Republic supported by MH CZ – DRO (MMCI, 00209805) and MEYS – NPS I – LO1413; Fondation de France and Association de Recherche Contre le Cancer.

FNLCR - This project has been funded in whole or in part with federal funds from the Frederick National Laboratory for Cancer Research, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This Research was supported in part by the Intramural Research Program of the NIH, Frederick National Lab, Center for Cancer Research.

GEC/Mayo GWAS - National Institutes of Health (CA118444, CA148690, CA92153). Intramural Research Program of the NIH, National Cancer Institute. Veterans Affairs Research Service. Data collection for Duke University was supported by a Leukemia & Lymphoma Society Career Development Award, the Bernstein Family Fund for Leukemia and Lymphoma Research, and the National Institutes of Health (K08CA134919), National Center for Advancing Translational Science (UL1 TR000135).

HPFS – The HPFS was supported in part by National Institutes of Health grants CA167552, CA149445, and CA098122. We would like to thank the participants and staff of the Health Professionals Follow-up Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

Iowa-Mayo SPORE – NCI Specialized Programs of Research Excellence (SPORE) in Human Cancer (P50 CA97274); National Cancer Institute (P30 CA086862, P30 CA15083); Henry J. Predolin Foundation.

Italian GxE - Italian Association for Cancer Research (AIRC, Investigator Grant 11855) (PC); Fondazione Banco di Sardegna 2010-2012, and Regione Autonoma della Sardegna (LR7 CRP-59812/2012) (MGE).

Mayo Clinic Case-Control – National Institutes of Health (R01 CA92153); National Cancer Institute (P30 CA015083).

MCCS – The Melbourne Collaborative Cohort Study recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 209057, 251553 and 504711 and by recurrent funding and infrastructure provided by Cancer Council Victoria. The incidence of malignancy and their participants' vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database.

MSKCC – Geoffrey Beene Cancer Research Grant, Lymphoma Foundation (LF5541); Barbara K. Lipman Lymphoma Research Fund (74419); Robert and Kate Niehaus Clinical Cancer Genetics Research Initiative (57470); U01 HG007033; ENCODE; U01 HG007033.

NCI-SEER – Intramural Research Program of the National Cancer Institute, National Institutes of Health, and Public Health Service (N01-PC-65064, N01-PC-67008, N01-PC-67009, N01-PC-67010, N02-PC-71105).

NHS –The NHS was supported in part by National Institutes of Health grants CA186107, CA87969, CA49449, CA149445, and CA098122. We would like to thank the participants and staff of the Nurses' Health Study for their valuable contributions as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

NSW - NSW was supported by grants from the Australian National Health and Medical Research Council (ID990920), the Cancer Council NSW, and the University of Sydney Faculty of Medicine.

NYU-WHS - National Cancer Institute (R01 CA098661, P30 CA016087); National Institute of Environmental Health Sciences (ES000260).

PLCO - This research was supported by the Intramural Research Program of the National Cancer Institute and by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS.

SCALE – Swedish Cancer Society (2009/659). Stockholm County Council (20110209) and the Strategic Research Program in Epidemiology at Karolinska Institutet. Swedish Cancer Society grant (02 6661). National Institutes of Health (5R01 CA69669-02); Plan Denmark.

UCSF2 – The UCSF studies were supported by the NCI, National Institutes of Health, CA1046282 and CA154643. The collection of cancer incidence data used in this study was supported by the California Department of Health Services as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology, and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement #1U58 DP000807-01 awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the authors, and endorsement by the State of California, the California Department of Health Services, the National Cancer Institute, or the Centers for Disease Control and Prevention or their contractors and subcontractors is not intended nor should be inferred.

UTAH - National Institutes of Health CA134674. Partial support for data collection at the Utah site was made possible by the Utah Population Database (UPDB) and the Utah Cancer Registry (UCR). Partial support for all datasets within the UPDB is provided by the Huntsman Cancer Institute (HCI) and the HCI Cancer Center Support grant, P30 CA42014. The UCR is supported in part by NIH contract HHSN261201000026C from the National Cancer Institute SEER Program with additional support from the Utah State Department of Health and the University of Utah.

YALE – National Cancer Institute (CA62006); National Cancer Institute (CA165923).

REFERENCES

1. Cerhan JR, Berndt SI, Vijai J, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nat Genet.* Nov 2014;46(11):1233-1238.
2. Skibola CF, Berndt SI, Vijai J, et al. Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the HLA region. *Am J Hum Genet.* Oct 02 2014;95(4):462-471.
3. Vijai J, Wang Z, Berndt SI, et al. A genome-wide association study of marginal zone lymphoma shows association to the HLA region. *Nat Commun.* Jan 08 2015;6:5751.
4. Berndt SI, Camp NJ, Skibola CF, et al. Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat Commun.* Mar 09 2016;7:10933.
5. 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.* Apr 2011;7(4):e1001378.
6. Carrington M, Nelson GW, Martin MP, et al. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science.* Mar 12 1999;283(5408):1748-1752.
7. Thursz MR, Thomas HC, Greenwood BM, Hill AV. Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nat Genet.* Sep 1997;17(1):11-12.
8. Thio CL, Thomas DL, Karacki P, et al. Comprehensive analysis of class I and class II HLA antigens and chronic hepatitis B virus infection. *J Virol.* Nov 2003;77(22):12083-12087.
9. Wang SS, Abdou AM, Morton LM, et al. Human leukocyte antigen class I and II alleles in non-Hodgkin lymphoma etiology. *Blood.* Jun 10 2010;115(23):4820-4823.
10. Morton LM, Turner JJ, Cerhan JR, et al. Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph). *Blood.* Jul 15 2007;110(2):695-708.
11. Turner JJ, Morton LM, Linet MS, et al. InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. *Blood.* Nov 18 2010;116(20):e90-98.
12. Swerdlow S, Cancer. IAfRo, Organization. WH. *WHO classification of tumours of haematopoietic and lymphoid tissues*. 4th ed. Lyon: International Agency for Research on Cancer; 2008.
13. Morton LM, Wang SS, Cozen W, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. *Blood.* Dec 15 2008;112(13):5150-5160.
14. Hraber P, Kuiken C, Yusim K. Evidence for human leukocyte antigen heterozygote advantage against hepatitis C virus infection. *Hepatology.* Dec 2007;46(6):1713-1721.
15. Hraber P, Kuiken C, Yusim K. Evidence for human leukocyte antigen heterozygote advantage against hepatitis C virus infection. Hepatology. 2007;46(6):1713-21.
16. Goyette P, Boucher G, Mallon D, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet.* Feb 2015;47(2):172-179.

17. Nelson GW, Martin MP, Gladman D, Wade J, Trowsdale J, Carrington M. Cutting edge: heterozygote advantage in autoimmune disease: hierarchy of protection/susceptibility conferred by HLA and killer Ig-like receptor combinations in psoriatic arthritis. *J Immunol.* Oct 01 2004;173(7):4273-4276.
18. Shah N, Decker WK, Lapushin R, Xing D, Robinson SN, Yang H, Parmar S, Tung SS, O'Brien S, Fernandez-Viña M, Shpall EJ, Wierda WG. HLA homozygosity and haplotype bias among patients with chronic lymphocytic leukemia: implications for disease control by physiological immune surveillance. Leukemia. 2011;25(6):1036-9.
19. Guillaume N, Marolleau JP. Is immune escape via human leukocyte antigen expression clinically relevant in chronic lymphocytic leukemia? Focus on the controversies. Leuk Res. 2013;37(4):473-7.
20. Gragert L, Fingerson S, Albrecht M, Maiers M, Kalaycio M, Hill BT. Fine-mapping of HLA associations with chronic lymphocytic leukemia in US populations. Blood. 2014;124(17):2657-65.
21. Lau Q, Yasukochi Y, Satta Y.A limit to the divergent allele advantage model supported by variable pathogen recognition across HLA-DRB1 allele lineages. Tissue Antigens. 2015;86(5):343-52.
22. Hemminki K, Liu X, Ji J, Forsti A. Origin of B-Cell Neoplasms in Autoimmune Disease. *PLoS One.* 2016;11(6):e0158360.
23. Khankhanian P, Cozen W, Himmelstein DS, et al. Meta-analysis of genome-wide association studies reveals genetic overlap between Hodgkin lymphoma and multiple sclerosis. *Int J Epidemiol.* Jun 2016;45(3):728-740.
24. Engels EA, Parsons R, Besson C, et al. Comprehensive Evaluation of Medical Conditions Associated with Risk of Non-Hodgkin Lymphoma using Medicare Claims ("MedWAS"). *Cancer Epidemiol Biomarkers Prev.* Jul 2016;25(7):1105-1113.
25. McAulay KA, Jarrett RF. Human leukocyte antigens and genetic susceptibility to lymphoma. *Tissue Antigens.* Aug 2015;86(2):98-113.
26. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol.* Apr 27 2017;18(1):76.

TABLES

Table 1. Genome-wide association studies (GWAS) included in the evaluation of human leukocyte antigen (*HLA*) homozygosity and risk of four non-Hodgkin lymphoma (NHL) subtypes: diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL), and marginal zone lymphoma (MZL).

Table 2: Effect of homozygosity at the three *HLA* class I loci -*A*, -*B* and -*C* and three *HLA* class II loci -*DRB1*, *DQB1*, and *DPB1* on susceptibility to four NHL subtypes (DLBCL, FL, CLL/SLL, and MZL) in participants of European-descent within participating lymphoma genome-wide association studies (analyses adjusted for sex, study or region, age, and ancestry/principal components).

Table 3: Effects of zygosity by individual *HLA* class I and class II loci, for DLBCL, MZL, and FL (analyses adjusted for sex, age, and ancestry/principal components).

Table 1. Genome wide association studies (GWAS) included in the evaluation of human leukocyte antigen (HLA) homozygosity and risk of four non-Hodgkin lymphoma (NHL) subtypes: diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL), and marginal zone lymphoma (MZL).

Study Name	Study Abbreviation	GWAS Platform	NHL Cases				Controls
			DLBCL (n=3617)	FL (n=2686)	CLL/SLL (n=2878)	MZL (n=741)	(n=8753)
Alpha-Tocopherol, Beta-Carotene Lung Cancer Prevention Study	ATBC	Illumina OmniExpress	43	17	50	1	238
British Columbia Non-Hodgkin Lymphoma Study	BCCA	Illumina OmniExpress	92	98	26	40	109
American Cancer Society Cancer Prevention Study-II Nutrition Cohort	CPS-II	Illumina OmniExpress	188	141	251	52	220
Treatment program of DLBCL patients from the Groupe d'Etude des Lymphomes de l'Adulte (GELA) consisting in LNH03-1B, 2B, 3B, 39B, 6B and 7B.	GELA	Illumina HumanHap 610K	549	0	0	0	0
Epidemiology & Genetics Unit Lymphoma Case-Control study	ELCCS	Illumina OmniExpress	229	182	0	0	245
Environmental and genetic risks factors study in adult lymphoma	ENGELA	Illumina OmniExpress	56	30	44	5	63
European Prospective Investigation into Cancer, Chronic Diseases, Nutrition and Lifestyles	EPIC	Illumina OmniExpress	46	46	72	8	265
Epilymph case-control study in six European countries	EpiLymph	Illumina OmniExpress	198	123	158	59	211
Genetic Epidemiology of CLL (GEC) Consortium	GEC	Affymetrix 6.0	0	0	391	0	296
Health Professionals Follow-up Study	HPFS	Illumina OmniExpress	12	5	19	5	85
Iowa-Mayo SPORE Molecular Epidemiology Resource	IOWA-MAYO SPORE	Illumina OmniExpress	146	228	242	112	0
Multicenter Italian study on gene-environment interactions in lymphoma etiology: translational aspects	Italian GxE	Illumina OmniExpress	16	16	5	6	45
Mayo Clinic Case-Control Study of NHL and CLL	MAYO-Case-Control	Illumina OmniExpress	25	245	132	75	343
Mayo Clinic Case-Control Study of NHL and CLL	MAYO-Case-Control	Illumina HumanHap 660W	393	0	0	0	172
The Melbourne Collaborative Cohort Study	MCCS	Illumina OmniExpress	71	58	57	8	75
Memorial-Sloan Kettering Lymphoproliferative Disorders Study	MSKCC	Illumina OmniExpress	175	174	36	47	4
National Cancer Institute-Surveillance, Epidemiology, and End Results Interdisciplinary Case-Control Study of Non-Hodgkin's Lymphoma	NCI-SEER	Illumina OmniExpress	251	217	86	62	270
Nurses' Health Study	NHS	Illumina OmniExpress	28	24	18	12	88
New South Wales non-Hodgkin lymphoma study	NSW	Illumina OmniExpress	115	146	13	34	154
New York University Women's Health Study	NYU-WHS	Illumina OmniExpress	8	11	10	6	53
Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial	PLCO	Illumina OmniExpress	153	115	278	26	3076
Scandinavian Lymphoma Epidemiology Study	SCALE	Illumina OmniExpress	405	0	395	64	291
Scandinavian Lymphoma Epidemiology Study	SCALE	Illumina HumanHap 317K	0	376	0	0	791
Molecular Epidemiology of non-Hodgkin lymphoma 1	UCSF1	Illumina OmniExpress	38	7	22	91	10
Molecular Epidemiology of non-Hodgkin lymphoma 1	UCSF1	Illumina HumanCNV370-Duo	254	210	213	0	749
Molecular Epidemiology of non-Hodgkin lymphoma 2	UCSF2	Illumina OmniExpress	0	119	0	0	349
Utah Chronic Lymphocytic Leukemia Study	UTAH	Illumina HumanHap 610K	0	0	321	0	405
Population-based NHL case-control study in Connecticut women	YALE	Illumina OmniExpress	126	98	39	28	146

Table 2: Effect of homozygosity at the three HLA class I loci *-A, -B* and *-C* and three HLA class I loci *-DRB1, DQB1,* and *DPB1* on susceptibility to four NHL subtypes (DLBCL, FL, CLL/SLL, and MZL) in Caucasian participants within participating lymphoma genome-wide association studies (analyses adjusted for sex, study or region, age, and ancestry/principal components).

		Controls (n=6912)		DLBCL (n=3617)			Controls (n=7880)		FL (n=2686)		
		n	%	n	%	OR (95% CI)	n	%	n	%	OR (95% CI)
Class I locus											
HLA-A	Heterozygote	6039	89	3096	86	1.00 (ref)	6843	88	2330	88	1.00 (ref)
	Homozygote	756	11	484	14	1.14 (0.98-1.34)	923	12	313	12	1.03 (0.88-1.21)
HLA-B	Heterozygote	6430	93	3297	91	1.00	7330	93	2469	92	1.00 (ref)
	Homozygote	476	7	318	9	**1.22 (1.01-1.47)**	544	7	216	8	1.14 (0.94-1.38)
HLA-C	Heterozygote	6238	90	3182	88	1.00	7112	90	2383	89	1.00 (ref)
	Homozygote	674	10	435	12	**1.20 (1.02-1.41)**	768	10	302	11	1.13 (0.96-1.34)
Total # of homozygous Class I loci	0	5535	80	2792	77	1.00	6266	80	2121	79	1.00 (ref)
	1	950	14	524	14	1.05 (0.90-1.21)	1120	14	361	64	0.98 (0.84-1.13)
	2	297	4	187	5	**1.33 (1.05-1.69)**	342	4	132	5	1.18 (0.93-1.51)
	3	130	2	114	3	1.31 (0.95-1.81)	152	2	72	3	1.29 (0.93-1.79)
	p-trend					***0.0008***					*0.12*
	OR per locus					1.11 (1.03-1.19)					1.06 (0.98-1.15)
Class II locus											
HLA-DRB1	Heterozygote	6331	92	3173	88	1.00 (ref)	7212	92	2339	87	1.00 (ref)
	Homozygote	561	8	435	12	**1.51 (1.27-1.78)**	648	8	338	13	**1.54 (1.31-1.82)**
HLA-DQB1	Heterozygote	6137	89	3055	84	1.00 (ref)	6999	89	2255	84	1.00 (ref)
	Homozygote	773	11	561	16	**1.30 (1.12-1.51)**	879	11	431	16	**1.42 (1.23-1.65)**
HLA-DPB1	Heterozygote	5544	80	2817	78	1.00 (ref)	6292	80	2064	77	1.00 (ref)
	Homozygote	1356	20	798	22	1.05 (0.93-1.19)	1576	20	620	23	**1.24 (1.10-1.40)**
Total # of homozygous Class II loci	0	4889	71	2341	65	1.00 (ref)	5545	71	1694	63	1.00 (ref)
	1	1428	21	830	23	1.08 (0.95-1.22)	1656	21	660	25	**1.28 (1.14-1.45)**
	2	426	6	344	10	**1.51 (1.25-1.83)**	493	6	239	9	**1.47 (1.21-1.78)**
	3	136	2	91	3	1.30 (0.92-1.82)	153	2	82	3	**1.89 (1.37-2.61)**
	p-trend					***<0.0001***					***<0.0001***
	OR per locus					1.15 (1.07-1.23)					1.24 (1.15-1.32)
Total # of homozygous Class I or Class II loci	0	3972	59	1866	52	1.00 (ref)	4486	58	1390	53	1.00 (ref)
	1	1750	26	992	28	1.11 (0.98-1.25)	2029	26	710	27	**1.13 (1.00-1.28)**
	2	625	9	407	11	**1.32 (1.11-1.58)**	741	10	293	11	**1.22 (1.03-1.45)**
	3	232	3	128	4	0.96 (0.73-1.27)	259	3	130	5	**1.55 (1.19-2.00)**
	4	98	1	82	2	**1.92 (1.30-2.81)**	114	1	58	2	**1.94 (1.32-2.85)**
	5+	84	1	92	3	**1.72 (1.19-2.49)**	100	1	50	2	**1.50 (1.02-2.22)**
	p-trend					**<0.0001**					**<0.0001**
	OR per locus					1.10 (1.06-1.16)					1.13 (1.08-1.18)

*adjusted by geographic region (continent) rather than study

Table 2 (*continued*): Effect of homozygosity at the three HLA class I loci *-A, -B* and *-C* and three HLA class I loci *-DRB1, DQB1,* and *DPB1* on susceptibility to four NHL subtypes (DLBCL, FL, CLL/SLL, and MZL) in Caucasian participants within participating lymphoma genome-wide association studies (analyses adjusted for sex, study or region, age, and ancestry/principal components).

		Controls (n=7441)		CLL/SLL (n=2878)			Controls (n=5991)		MZL* (n=741)		
		n	%	n	%	OR (95% CI)	n	%	n	%	OR (95% CI)
Class I locus											
HLA-A	Heterozygote	6504	89	2460	87	1.00 (ref)	5244	89	649	89	1.00 (ref)
	Homozygote	821	11	378	13	**1.19 (1.02-1.38)**	646	11	78	11	1.06 (0.82-1.38)
HLA-B	Heterozygote	6916	93	2656	92	1.00 (ref)	5576	93	675	91	1.00 (ref)
	Homozygote	519	7	221	8	1.04 (0.87-1.26)	411	7	66	9	**1.34 (1.01-1.78)**
HLA-C	Heterozygote	6719	90	2576	90	1.00 (ref)	5414	90	651	88	1.00 (ref)
	Homozygote	722	10	301	10	1.10 (0.94-1.29)	577	10	90	12	**1.33 (1.04-1.70)**
Total # of homozygous Class I loci	0	5965	80	2225	77	1.00 (ref)	4805	80	586	79	1.00 (ref)
	1	1009	14	457	70	**1.19 (1.03-1.36)**	822	14	94	13	0.97 (0.76-1.23)
	2	323	4	130	5	1.08 (0.85-1.37)	256	4	37	5	1.16 (0.80-1.68)
	3	144	2	66	2	1.16 (0.83-1.62)	108	2	24	3	**2.13 (1.33-3.42)**
	p-trend					*0.0518*					***0.026***
	OR per locus					1.08 (1.00-1.16)					1.08 (1.00-1.16)
Class II locus											
HLA-DRB1	Heterozygote	6810	92	2583	90	1.00 (ref)	5500	92	663	89	1.00 (ref)
	Homozygote	608	8	286	10	**1.19 (1.00-1.42)**	480	8	78	11	**1.45 (1.12-1.89)**
HLA-DQB1	Heterozygote	6603	89	2494	87	1.00 (ref)	5310	89	638	98	1.00 (ref)
	Homozygote	836	11	384	13	**1.20 (1.03-1.39)**	681	11	10	2	1.20 (0.95-1.52)
HLA-DPB1	Heterozygote	5972	80	2320	81	1.00 (ref)	4809	80	582	79	1.00 (ref)
	Homozygote	1455	20	554	19	0.92 (0.81-1.04)	1176	20	158	21	1.13 (0.93-1.38)
Total # of homozygous Class II loci	0	5255	71	1994	70	1.00 (ref)	4247	71	501	68	1.00 (ref)
	1	1543	21	586	20	0.95 (0.84-1.08)	1241	21	159	21	1.08 (0.89-1.31)
	2	462	6	224	8	1.20 (0.99-1.46)	363	6	60	8	**1.42 (1.05-1.91)**
	3	143	2	62	2	1.10 (0.77-1.57)	123	2	20	3	1.48 (0.89-2.43)
	p-trend					*0.1924*					***0.0124***
	OR per locus					1.04 (0.97-1.12)					1.15 (1.03-1.28)
Total # of homozygous Class I or Class II loci	0	4275	59	1569	56	1.00 (ref)	3446	59	407	56	1.00 (ref)
	1	1880	26	767	27	1.07 (0.95-1.20)	1528	26	181	25	1.03 (0.85-1.25)
	2	683	9	284	10	1.11 (0.94-1.32)	538	9	71	10	1.16 (0.87-1.53)
	3	247	3	106	4	1.10 (0.84-1.44)	204	3	39	5	**1.52 (1.04-2.21)**
	4	111	2	49	2	1.06 (0.71-1.57)	83	1	16	2	**1.84 (1.04-3.27)**
	5+	87	1	49	2	**1.57 (1.04-2.38)**	70	1	12	2	1.64 (0.86-3.13)
	p-trend					**0.029**					**0.0024**
	OR per locus					1.05 (1.01-1.10)					1.12 (1.04-1.20)

*adjusted by geographic region (continent) rather than study

Table 3: Effects of zygosity by individual HLA Class I and Class II loci, for DLBCL, MZL, FL, and CLL/SLL (analyses adjusted for sex, age, study/region, and ancestry/principal components).

			Controls (n=6912)		DLBCL (n=3617)			Controls (n=7880)		FL (n=2686)		
			n	%	n	%	OR (95% CI)	n	%	n	%	OR (95% CI)
Class I locus												
HLA-B	***HLA-C***											
Heterogyzote	Heterozygote		6133	89	3127	87	1.00 (ref)	6992	89	2350	88	1.00 (ref)
Heterogyzote	Homozygote		297	4	170	5	1.07 (0.83-1.36)	338	4	119	4	1.01 (0.79-1.29)
Homozygote	Heterozygote		100	1	53	1	0.89 (0.58-1.38)	115	1	33	1	0.81 (0.51-1.28)
Homozygote	Homozygote		376	5	265	7	**1.31 (1.06-1.60)**	429	5	183	7	1.23 (1.00-1.52)
		p-trend					*0.02*					*0.1258*
		p-trend OR					*1.08 (1.01-1.15)*					*1.05 (0.99-1.13)*
Class II locus												
HLA-DPB1	***HLA-DQB1***	***HLA-DRB1***										
Heterozygote	Heterozygote	Heterozygote	4889	72	2341	65	1.00 (ref)	5545	71	1694	63	1.00 (ref)
Heterozygote	Heterozygote	Homozygote	52	1	42	1	**2.10 (1.24-3.55)**	62	1	48	2	**2.60 (1.66-4.06)**
Heterozygote	Homozygote	Heterozygote	239	3	149	4	1.01 (0.77-1.33)	263	3	122	5	**1.33 (1.03-1.73)**
Heterozygote	Homozygote	Homozygote	345	5	277	8	**1.54 (1.25-1.91)**	403	5	195	7	**1.42 (1.14-1.76)**
Homozygote	Heterozygote	Heterozygote	1137	17	639	18	1.05 (0.92-1.21)	1331	17	490	18	**1.21 (1.06-1.39)**
Homozygote	Heterozygote	Homozygote	28	0	24	1	1.44 (0.77-2.71)	30	0	13	0	1.33 (0.66-2.68)
Homozygote	Homozygote	Heterozygote	53	1	43	1	1.38 (0.82-2.33)	60	1	31	1	**1.88 (1.14-3.10)**
Homozygote	Homozygote	Homozygote	136	2	91	3	1.30 (0.92-1.82)	153	2	82	3	**1.89 (1.37-2.61)**
		p-trend					*0.0091*					*<0.0001*
		p-trend OR					*1.04 (1.01-1.06)*					*1.07 (1.05-1.10)*

Table 3 (*continued*): Effects of zygosity by individual HLA Class I and Class II loci, for DLBCL, MZL, FL, and CLL/SLL (analyses adjusted for sex, age, study/region, and ancestry/principal components).

			Controls (n=5991)		MZL* (n=741)			Controls (n=7441)		CLL/SLL (n=2878)		
			n	%	n	%	OR (95% CI)	n	%	n	%	OR (95% CI)
Class I locus												
HLA-B	***HLA-C***											
Heterogyzote	Heterozygote		5321	89	637	86	1.00 (ref)	6608	89	2520	88	1.00 (ref)
Heterogyzote	Homozygote		255	4	38	5	1.28 (0.89-1.85)	308	4	135	5	1.19 (0.94-1.50)
Homozygote	Heterozygote		89	1	14	2	1.27 (0.70-2.30)	106	1	55	2	1.10 (0.75-1.62)
Homozygote	Homozygote		322	5	52	7	**1.38 (1.01-1.90)**	413	6	166	6	1.04 (0.84-1.29)
		p-trend					*0.02*					*0.43*
		p-trend OR					*1.12 (1.02-1.24)*					*1.03 (0.96-1.10)*
Class II locus												
HLA-DPB1	***HLA-DQB1***	***HLA-DRB1***										
Heterozygote	Heterozygote	Heterozygote	4247	71	501	68	1.00 (ref)	5255	71	1994	70	1.00 (ref)
Heterozygote	Heterozygote	Homozygote	42	1	9	1	2.10 (0.99-4.45)	61	1	39	1	**1.76 (1.09-2.86)**
Heterozygote	Homozygote	Heterozygote	213	4	24	3	0.80 (0.51-1.26)	257	3	110	4	1.15 (0.88-1.49)
Heterozygote	Homozygote	Homozygote	297	5	48	6	**1.43 (1.03-2.00)**	378	5	172	6	1.12 (0.90-1.39)
Homozygote	Heterozygote	Heterozygote	986	17	126	17	1.11 (0.89-1.37)	1225	17	437	15	0.88 (0.76-1.01)
Homozygote	Heterozygote	Homozygote	18	0	1	0	0.62 (0.08-4.82)	26	0	13	0	1.15 (0.54-2.48)
Homozygote	Homozygote	Heterozygote	48	1	11	1	1.54 (0.75-3.16)	58	1	39	1	**1.78 (1.11-2.85)**
Homozygote	Homozygote	Homozygote	123	2	20	3	1.48 (0.90-2.44)	143	2	62	2	1.10 (0.77-1.57)
		p-trend					*0.04*					*0.95*
		p-trend OR					*1.04 (1.00-1.09)*					*1.00 (0.97-1.03)*