Author/s:
Vijai, J;Wang, Z;Berndt, SI;Skibola, CF;Slager, SL;De Sanjose, S;Melbye, M;Glimelius, B;Bracci, PM;Conde, L;Birmann, BM;Wang, SS;Brooks-Wilson, AR;Lan, Q;De Bakker, PIW;Vermeulen, RCH;Portlock, C;Ansell, SM;Link, BK;Riby, J;North, KE;Gu, J;Hjalgrim, H;Cozen, W;Becker, N;Teras, LR;Spinelli, JJ;Turner, J;Zhang, Y;Purdue, MP;Giles, GG;Kelly, RS;Zeleniuch-Jacquotte, A;Ennas, MG;Monnereau, A;Bertrand, KA;Albanes, D;Lightfoot, T;Yeager, M;Chung, CC;Burdett, L;Hutchinson, A;Lawrence, C;Montalvan, R;Liang, L;Huang, J;Ma, B;Villano, DJ;Maria, A;Corines, M;Thomas, T;Novak, AJ;Dogan, A;Liebow, M;Thompson, CA;Witzig, TE;Habermann, TM;Weiner, GJ;Smith, MT;Holly, EA;Jackson, RD;Tinker, LF;Ye, Y;Adami, HO;Smedby, KE;De Roos, AJ;Hartge, P;Morton, LM;Severson, RK;Benavente, Y;Boffetta, P;Brennan, P;Foretova, L;Maynadie, M;McKay, J;Staines, A;DIver, WR;Vajdic, CM;Armstrong, BK;Kricker, A;Zheng, T;Holford, TR;Severi, G;Vineis, P;Ferri, GM;Ricco, R;Miligi, L;Clavel, J;Giovannucci, E;Kraft, P;Virtamo, J;Smith, A;Kane, E;Roman, E;Chiu, BCH;Fraumeni, JF;Wu, X;Cerhan, JR;Offit, K;Chanock, SJ

Title:
A genome-wide association study of marginal zone lymphoma shows association to the HLA region

Date:
2015-01-08

Citation:
Vijai, J., Wang, Z., Berndt, S. I., Skibola, C. F., Slager, S. L., De Sanjose, S., Melbye, M., Glimelius, B., Bracci, P. M., Conde, L., Birmann, B. M., Wang, S. S., Brooks-Wilson, A. R., Lan, Q., De Bakker, P. I. W., Vermeulen, R. C. H., Portlock, C., Ansell, S. M., Link, B. K. ,... Chanock, S. J. (2015). A genome-wide association study of marginal zone lymphoma shows association to the HLA region. Nature Communications, 6 (1), https://doi.org/10.1038/ncomms6751.

Persistent Link:
https://hdl.handle.net/11343/263359

nature COMMUNICATIONS

ARTICLE

Received 16 May 2014 | Accepted 4 Nov 2014 | Published 8 Jan 2015

DOI: 10.1038/ncomms6751 OPEN

A genome-wide association study of marginal zone lymphoma shows association to the HLA region

Joseph Vijai[1,*], Zhaoming Wang[2,*], Sonja I. Berndt[3,*], Christine F. Skibola[4,5,*], Susan L. Slager[6], Silvia de Sanjose[7,8], Mads Melbye[9,10], Bengt Glimelius[11,12], Paige M. Bracci[13], Lucia Conde[4,5], Brenda M. Birmann[14], Sophia S. Wang[15], Angela R. Brooks-Wilson[16,17], Qing Lan[3], Paul I.W. de Bakker[18,19], Roel C.H. Vermeulen[19,20], Carol Portlock[1], Stephen M. Ansell[21], Brian K. Link[22], Jacques Riby[4,5], Kari E. North[23,24], Jian Gu[25], Henrik Hjalgrim[9], Wendy Cozen[26,27], Nikolaus Becker[28], Lauren R. Teras[29], John J. Spinelli[30,31], Jenny Turner[32,33], Yawei Zhang[34], Mark P. Purdue[3], Graham G. Giles[35,36], Rachel S. Kelly[37], Anne Zeleniuch-Jacquotte[38,39], Maria Grazia Ennas[40], Alain Monnereau[41,42,43], Kimberly A. Bertrand[14,44], Demetrius Albanes[3], Tracy Lightfoot[45], Meredith Yeager[2], Charles C. Chung[3], Laurie Burdett[2], Amy Hutchinson[2], Charles Lawrence[46], Rebecca Montalvan[46], Liming Liang[44,47], Jinyan Huang[44], Baoshan Ma[44,48], Danylo J. Villano[1], Ann Maria[1], Marina Corines[1], Tinu Thomas[1], Anne J. Novak[21], Ahmet Dogan[49], Mark Liebow[21], Carrie A. Thompson[21], Thomas E. Witzig[21], Thomas M. Habermann[21], George J. Weiner[22], Martyn T. Smith[5], Elizabeth A. Holly[13], Rebecca D. Jackson[50], Lesley F. Tinker[51], Yuanqing Ye[25], Hans-Olov Adami[44,52], Karin E. Smedby[53], Anneclaire J. De Roos[51,54], Patricia Hartge[3], Lindsay M. Morton[3], Richard K. Severson[55], Yolanda Benavente[7,8], Paolo Boffetta[56], Paul Brennan[57], Lenka Foretova[58], Marc Maynadie[59], James McKay[60], Anthony Staines[61], W. Ryan Diver[29], Claire M. Vajdic[62], Bruce K. Armstrong[63], Anne Kricker[63], Tongzhang Zheng[34], Theodore R. Holford[64], Gianluca Severi[35,36,65], Paolo Vineis[37,65], Giovanni M. Ferri[66], Rosalia Ricco[67], Lucia Miligi[68], Jacqueline Clavel[41,42], Edward Giovannucci[14,44,69], Peter Kraft[44,47], Jarmo Virtamo[70], Alex Smith[45], Eleanor Kane[45], Eve Roman[45], Brian C.H. Chiu[71], Joseph F. Fraumeni[3], Xifeng Wu[25,†], James R. Cerhan[6,†], Kenneth Offit[1,†], Stephen J. Chanock[3,†], Nathaniel Rothman[3,†] & Alexandra Nieters[72,†]

Marginal zone lymphoma (MZL) is the third most common subtype of B-cell non-Hodgkin lymphoma. Here we perform a two-stage GWAS of 1,281 MZL cases and 7,127 controls of European ancestry and identify two independent loci near *BTNL2* (rs9461741, $P = 3.95 \times 10^{-15}$) and *HLA-B* (rs2922994, $P = 2.43 \times 10^{-9}$) in the HLA region significantly associated with MZL risk. This is the first evidence that genetic variation in the major histocompatibility complex influences MZL susceptibility.

[1] Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [2] Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, Maryland 20877, USA. [3] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland 20892, USA. [4] Department of Epidemiology, School of Public Health and Comprehensive Cancer Center, Birmingham, Alabama 35233, USA. [5] Division of Environmental Health Sciences, University of California Berkeley School of Public Health, Berkeley, California 94720, USA. [6] Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota 55905, USA. [7] Unit of Infections and Cancer (UNIC), Cancer Epidemiology Research Programme, Institut Catala d'Oncologia, IDIBELL, Barcelona 8907, Spain. [8] Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Barcelona 8036, Spain. [9] Department of Epidemiology Research, Division of Health Surveillance and Research, Statens Serum Institut, Copenhagen 2300, Denmark. [10] Department of Medicine, Stanford University School of Medicine, Stanford, California 94305, USA. [11] Department of Oncology and Pathology, Karolinska Institutet, Karolinska University Hospital Solna, Stockholm 17176, Sweden. [12] Department of Radiology, Oncology and Radiation Science, Uppsala University, Uppsala 75105, Sweden. [13] Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, California 94118, USA. [14] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [15] Department of Cancer Etiology, City of Hope Beckman Research Institute, Duarte, California 91030, USA. [16] Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, Canada V5Z1L3. [17] Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada V5A1S6. [18] Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht 3584 CG, The Netherlands. [19] Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands. [20] Institute for Risk Assessment Sciences, Utrecht University, Utrecht 3508 TD, The Netherlands. [21] Department of Medicine, Mayo Clinic, Rochester, Minnesota 55905, USA. [22] Department of Internal Medicine, Carver College of Medicine, The University of Iowa, Iowa City, Iowa 52242, USA. [23] Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [24] Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. [25] Department of Epidemiology, M.D. Anderson Cancer Center, Houston, Texas 77030, USA. [26] Department of Preventive Medicine, USC Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. [27] Norris Comprehensive Cancer Center, USC Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA. [28] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg 69120, Germany. [29] Epidemiology Research Program, American Cancer Society, Atlanta, Georgia 30303, USA. [30] Cancer Control Research, BC Cancer Agency, Vancouver, British Columbia, Canada V5Z1L3. [31] School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada V6T1Z3. [32] Pathology, Australian School of Advanced Medicine, Macquarie University, Sydney, New South Wales 2109, Australia. [33] Department of Histopathology, Douglass Hanly Moir Pathology, Macquarie Park, New South Wales 2113, Australia. [34] Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut 06520, USA. [35] Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria 3053, Australia. [36] Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Carlton, Victoria 3010, Australia. [37] MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London W2 1PG, UK. [38] Department of Population Health, New York University School of Medicine, New York, New York 10016, USA. [39] Cancer Institute, New York University School of Medicine, New York, New York 10016, USA. [40] Department of Biomedical Science, University of Cagliari, Monserrato, Cagliari 09042, Italy. [41] Environmental Epidemiology of Cancer Group, Inserm, Centre for research in Epidemiology and Population Health (CESP), U1018, Villejuif F-94807, France. [42] UMRS 1018, Univ Paris Sud, Villejuif F-94807, France. [43] Registre des hémopathies malignes de la Gironde, Institut Bergonié, Bordeaux 33076, France. [44] Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [45] Department of Health Sciences, University of York, York YO10 5DD, UK. [46] Health Studies Sector, Westat, Rockville, Maryland 20850, USA. [47] Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [48] College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China. [49] Departments of Laboratory Medicine and Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. [50] Division of Endocrinology, Diabetes and Metabolism, The Ohio State University, Columbus, Ohio 43210, USA. [51] Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98117, USA. [52] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden. [53] Department of Medicine Solna, Karolinska Institutet, Stockholm 17176, Sweden. [54] Department of Environmental and Occupational Health, Drexel University School of Public Health, Philadelphia, Pennsylvania 19104, USA. [55] Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, Michigan 48201, USA. [56] The Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [57] Group of Genetic Epidemiology, Section of Genetics, International Agency for Research on Cancer, Lyon 69372, France. [58] Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute and MF MU, Brno 65653, Czech Republic. [59] EA 4184, Registre des Hémopathies Malignes de Côte d'Or, University of Burgundy and Dijon University Hospital, Dijon 21070, France. [60] Genetic Cancer Susceptibility Group, Section of Genetics, International Agency for Research on Cancer, Lyon 69372, France. [61] School of Nursing and Human Sciences, Dublin City University, Dublin 9, Ireland. [62] Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales 2052, Australia. [63] Sydney School of Public Health, The University of Sydney, Sydney, New South Wales 2006, Australia. [64] Department of Biostatistics, Yale School of Public Health, New Haven, Connecticut 06520, USA. [65] Human Genetics Foundation, Turin 10126, Italy. [66] Interdisciplinary Department of Medicine, University of Bari, Bari 70124, Italy. [67] Department of Pathological Anatomy, University of Bari, Bari 70124, Italy. [68] Environmental and Occupational Epidemiology Unit, Cancer Prevention and Research Institute (ISPO), Florence 50139, Italy. [69] Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [70] Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki FI-00271, Finland. [71] Department of Health Studies, University of Chicago, Chicago, Illinois 60637, USA. [72] Center For Chronic Immunodeficiency, University Medical Center Freiburg, Freiburg 79108, Germany. * These authors contributed equally to this work. † These authors jointly supervised this work. Correspondence and requests for materials should be addressed to J.V. (email: josephv@mskcc.org).

Marginal zone lymphoma (MZL) encompasses a group of lymphomas that originate from marginal zone B cells present in extranodal tissue and lymph nodes. Three subtypes of MZL have been defined, extranodal MZL of mucosa-associated lymphoid tissue (MALT), splenic MZL and nodal MZL, which together account for 7–12% of all non-Hodgkin lymphoma (NHL) cases. Geographic differences in incidence have been observed[1], and inflammation, immune system dysregulation and infectious agents, such as *Helicobacter pylori*, have been implicated particularly for the gastric MALT subtype[2], but little else is known of MZL aetiology.

Here we perform the first two-stage, subtype-specific genome-wide association study (GWAS) of MZL and identify two independent single-nucleotide polymorphisms (SNPs) within the HLA region associated with MZL risk. Together with recent studies on other common subtypes of NHL, these results point to shared susceptibility loci for lymphoma in the HLA region.

Results

Stage 1 MZL GWAS. As part of a larger NHL GWAS, 890 MZL cases and 2,854 controls from 22 studies in the United States and Europe (Supplementary Table 1) were genotyped using the Illumina OmniExpress array. Genotype data from the Illumina Omni2.5 was also available for 3,536 controls from three of the 22 studies[3]. After applying rigorous quality control filters (Supplementary Table 2, Methods), data for 611,856 SNPs with minor allele frequency (MAF) > 1% in 825 cases and 6,221 controls of European ancestry (Supplementary Fig. 1) remained for the stage 1 analysis (Supplementary Table 3). To discover variants associated with risk, logistic regression analysis was performed on these SNPs adjusting for age, gender and three significant eigenvectors computed using principal components analysis (Supplementary Fig. 2, Methods). Examination of the quantile–quantile (Q–Q) plot (Supplementary Fig. 3) showed minimal detectable evidence for population substructure ($\lambda = 1.01$) with some excess of small P values. A Manhattan plot revealed association signals at the HLA region (Supplementary Fig. 4; 6p21.33:31,061,211–32,620,572) on chromosome 6 reaching genome-wide significance. Removal of all SNPs in the HLA region resulted in an attenuation of the excess of small P values observed in the Q–Q plot, although some excess still remained. To further explore associations within the HLA region and identify other regions potentially associated with risk, common SNPs available in the 1000 Genomes project data release 3 were imputed (Methods).

Stage 2 genotyping. Ten SNPs in promising loci with $P \leq 7.5 \times 10^{-6}$ in the stage 1 discovery were selected for replication (stage 2) in an additional 456 cases and 906 controls of European ancestry (Supplementary Tables 1 and 3). Of the SNPs selected for replication, two SNPs were directly genotyped on the OmniExpress, while the remaining eight were imputed with high accuracy (median info score = 0.99) in stage 1 (Supplementary Table 4). Replication was carried out using Taqman genotyping. In the combined meta-analysis of 1,281 cases and 7,127 controls, we identified two distinct loci (Table 1, Fig. 1, Supplementary Table 4) at chromosomes 6p21.32 and 6p21.33 that reached the threshold of genome-wide statistical significance ($P < 5 \times 10^{-8}$). These are rs9461741 in the butyrophilin-like 2 (MHC class II associated) (*BTNL2*) gene at 6p21.32 in HLA class II ($P = 3.95 \times 10^{-15}$, odds ratio (OR) = 2.66, confidence interval (CI) = 2.08–3.39) and rs2922994 at 6p21.33 in HLA class I ($P = 2.43 \times 10^{-9}$, OR = 1.64, CI = 1.39–1.92). These two SNPs were weakly correlated ($r^2 = 0.008$ in 1000 Genomes CEU population), and when both were included in the same statistical model, both SNPs remained strongly associated with MZL risk (rs9461741, $P = 2.09 \times 10^{-15}$; rs2922994, $P = 6.03 \times 10^{-10}$), suggesting that the two SNPs are independent. Both SNPs were weakly correlated with other SNPs in the HLA region previously reported to be associated with other NHL subtypes or Hodgkin lymphoma ($r^2 < 0.14$ for all pairwise comparisons). None of the previously reported SNPs were significantly associated with MZL risk after adjustment for multiple testing ($P < 0.0025$) in our study, suggesting the associations are subtype-specific (Supplementary Table 5). Another SNP rs7750641 ($P = 3.34 \times 10^{-8}$; Supplementary Table 4) in strong linkage disequilibrium (LD) with rs2922994 ($r^2 = 0.85$) also showed promising association with MZL risk. rs7750641 is a missense variant in transcription factor 19 (*TCF19*), which encodes a DNA-binding protein implicated in the transcription of genes during the G1–S transition in the cell cycle[4]. The non-HLA SNPs genotyped in stage 2 were not associated with MZL risk (Supplementary Table 4).

HLA alleles. To obtain additional insight into plausible functional variants, we imputed the classical HLA alleles and amino acid residues using SNP2HLA[5] (Methods). No imputed HLA alleles or amino acid positions reached genome-wide significance (Supplementary Table 6). However, for HLA class I, the most promising associations were observed with *HLA-B*08* ($P = 7.94 \times 10^{-8}$), *HLA-B*08:01* ($P = 7.79 \times 10^{-8}$) and the *HLA-B* allele encoding an aspartic acid residue at position 9 (Asp9) ($P = 7.94 \times 10^{-8}$), located in the peptide binding groove of the protein. *HLA-B*08:01* and Asp9 are highly correlated ($r^2 \geq 0.99$), and thus their effect sizes were identical (OR = 1.67, 95% CI: 1.38–2.01). They are both also in strong LD with rs2922994 ($r^2 = 0.97$). Due to the fact that they are collinear, the effects of the SNPs and alleles were indistinguishable from one another in conditional modelling. For HLA class II, a suggestive association was observed with *HLA-DRB1*01:02* (OR = 2.24, 95% CI: 1.64–3.07, $P = 5.08 \times 10^{-7}$; Supplementary Table 6), which is moderately correlated with rs9461741 ($r^2 = 0.69$). Conditional analysis revealed that the effects of rs9461741 (the intragenic SNP in *BTNL2*) and *HLA-DRB1*01:02* were indistinguishable statistically (stage 1: rs9461741, $P_{adjusted} = 0.064$ and *HLA-DRB1*01:02*, $P_{adjusted} = 0.29$). A model containing both *HLA-B*08:01* and *HLA-DRB1*01:02* showed that the two alleles were independent (*HLA-B*08:01*: $P_{adjusted} = 4.65 \times 10^{-8}$ and *HLA-DRB1*01:02*: $P_{adjusted} = 2.97 \times 10^{-7}$), further supporting independent associations in HLA class I and II loci.

MALT versus non-MALT. Heterogeneity between the largest subtype of MZL, namely MALT and other subtypes grouped as non-MALT, was evaluated for the MZL associated SNPs (Supplementary Table 7). The effects were slightly stronger for MALT, but no evidence for substantial heterogeneity was observed ($P_{heterogeneity} \geq 0.05$). Studies have suggested that *H. pylori* infection is a risk factor for gastric MZL[2]. An examination of SNPs previously suggested to be associated with *H. pylori* infection in independent studies[6] did not reveal any significant association with the combined MZL or the MALT subtype in this study (Supplementary Table 8). Toll-like receptors (TLR) are considered strong candidates in mediating inflammatory immune response to pathogenic insults. A previous study reported[7] a nominally significant association with rs4833103 in the *TLR10–TLR1–TLR6* region with MZL risk. After excluding the cases and controls in the previous report[7], we found little additional support for this locus (MZL: $P = 0.006$, OR = 1.18 and MALT: $P = 0.38$, OR = 1.08).

Table 1 | Association results for two new independent SNPs with MZL in a two-stage GWAS.

Chr	Nearest gene(s)	SNP	Position*	Risk allele†	Other allele	RAF‡	Stage	No. of cases/ no. of controls	OR	95% CI	*P* value§	$P_{heterogeneity}$	I^2
6p21.32	*BTNL2*	rs9461741	32370587	C	G	0.018	Stage 1	824/6,220	2.40	(1.74–3.31)	9.11E-08		
						0.030	Stage 2	453/877	3.06	(2.10–4.46)	5.24E-09		
							Combined	1,277/7,097	2.66	(2.08–3.39)	3.95E-15	0.216	34.69
6p21.33	*HLA-B*	rs2922994	31335901	G	A	0.113	Stage 1	825/6,221	1.74	(1.43–2.12)	2.89E-08		
						0.094	Stage 2	405/832	1.43	(1.08–1.90)	0.01		
							Combined	1,230/7,053	1.64	(1.39–1.92)	2.43E-09	0.507	0

CI, confidence interval; GWAS, genome-wide association study; MZL, marginal zone lymphoma; OR, odds ratio; RAF, risk allele frequency; SNP, single-nucleotide polymorphism.
*Position according to human reference NCBI37/hg19.
†Allele associated with an increased risk of MZL.
‡Risk allele frequency in controls.
§For stage 1 and 2, *P* values were generated by using logistic regression. For the combined stage, the odds ratio and *P* values were generated using a fixed effects model. Heterogeneity in the effect estimates was assessed using Cochran's *Q* statistic and estimating the I^2 statistic.

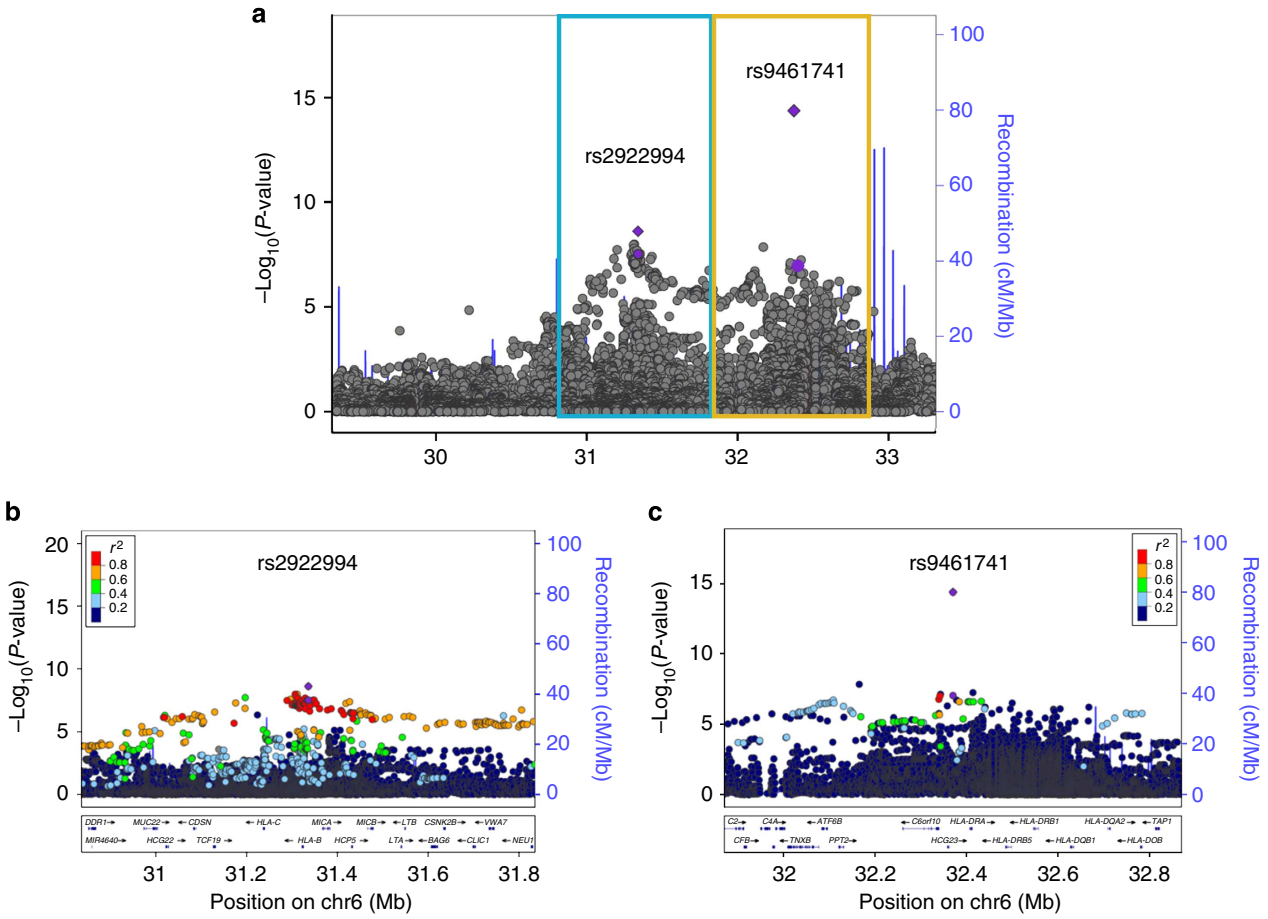

Figure 1 | Regional plot showing the HLA associations with MZL. The figure shows the association $\log_{10}$ *P* values from the log-additive genetic model for all SNPs in the region from stage 1 (dots) ($n=825$ cases, $n=6,221$ controls) and the $\log_{10}$ *P* values from the log-additive genetic model for both stage 1 and 2 combined (purple diamonds) for rs2922994 ($n=1,230$ cases, $n=7,053$ controls) and rs9461741 ($n=1,277$ cases, $n=7,097$ controls). The purple dots show the $\log_{10}$ *P* values of these SNPs in stage 1. Top panel (**a**) shows the region encompassing both SNPs. Bottom panel (**b**) regional plot of the most significant SNP rs2922994 at 6p21.33 (**c**) and rs9461741 at 6p21.32. The colours of the dots reflect the LD (as measured by r^2) with the most significant SNP as shown in the legend box.

Secondary functional analyses. To gain additional insight into potential biological mechanisms, expression quantitative trait loci (eQTL) analyses were performed in two datasets consisting of lymphoblastoid cell lines (Methods). Significant associations were seen for rs2922994 and rs7750641with *HLA-B* and *HLA-C* (Supplementary Table 9) while suggestive associations (false discovery rate, FDR ≤ 0.05) for correlated SNPs of rs2922994 ($r^2 > 0.8$) in HLA class I and *RNF5* (Supplementary Table 10) were observed. No significant eQTL association was observed for rs9461741 or other correlated HLA class II SNPs. Chromatin state analysis (Methods) using ENCODE data revealed correlated SNPs of rs2922994 showed a chromatin state consistent with the prediction for an active promoter (rs3094005) or satisfied the state of a weak promoter (rs2844577) in the lymphoblastoid cell line

GM12878 (Supplementary Fig. 5). GM12878 is the only lymphoblastoid cell line from which high-quality whole-genome annotation data for chromatin state is readily available. Analyses were also performed with HaploReg (Supplementary Table 11) and RegulomeDB (Supplementary Table 12) that showed overlap of the SNPs with functional motifs, suggesting plausible roles in gene regulatory processes.

Discussion

The most statistically significant SNP associated with MZL, rs9461741, is located in HLA class II in the intron between exons 3 and 4 of the *BTNL2* gene. *BTNL2* is highly expressed in lymphoid tissues[8] and has close homology to the B7 co-stimulatory molecules, which initiate lymphocyte activation as part of antigen presentation. Evidence is consistent with *BTNL2* acting as a negative regulator of T-cell proliferation and cytokine production[8,9] and attenuating T-cell-mediated responses in the gut[10]. We were unable to statistically differentiate the effects of rs9461741 from *HLA-DRB1*01:02* and, thus, our observed association could be due to *HLA-DRB1*. *HLA-DRB1* has been shown to be associated with other autoimmune diseases, including rheumatoid arthritis[11] and selective IgA deficiency[12]. Similarly, rs2922994 is located 11 kb upstream of *HLA-B*, which is known to play a critical role in the immune system by presenting peptides derived from the endoplasmic reticulum lumen. rs7750641, a missense variant in *TCF19*, was previously associated with pleiotropic effects on blood-based phenotypes[13] and is highly expressed in germinal center cells and up-regulated in human pro-B and pre-B cells[14]. Autoimmune diseases, such as Sjögren's syndrome and systemic lupus erythematosus, are established risk factors for developing MZL, with the strongest associations seen between Sjögren's syndrome and the MALT subtype[15]. Of note, SNPs in *HLA-B* and the classical alleles *HLA-DRB1**01:02 are strongly associated with Sjögren's syndrome[16], while *HLA-DRB1**03 has been associated with rheumatoid arthritis[17]. The multiple independent associations in the HLA region and their localization to known functional autoimmune and B-cell genes suggest possible shared genetic effects that span both lymphoid cancers and autoimmune diseases. Chronic autoimmune stimulation leading to over-activity and defective apoptosis of B cells, and secondary inflammation events triggered by genetic and environmental factors are biological mechanisms that may contribute to the pathogenesis of MZL.

We have performed the largest GWAS of MZL to date and identified two independent SNPs within the HLA region that are robustly associated with the risk of MZL. In addition to the known diversity in etiology and pathology, there is mounting evidence of genetic heterogeneity across the NHL subtypes of lymphoma. However, the HLA region appears to be commonly associated with multiple major subtypes, such as MZL, CLL[18], DLBCL[19] and FL[20–23]. Further studies are needed to identify biological mechanisms underlying these relationships and advance our knowledge regarding their interactions with associated environmental factors that may modulate disease risks.

Methods

Stage 1 MZL GWAS study subjects and ethics. As part of a larger NHL GWAS initiative, we conducted a GWAS of MZL using 890 cases and 2,854 controls of European descent from 22 studies of NHL (Supplementary Table 1 and Supplementary Table 2), including nine prospective cohort studies, eight population-based case–control studies, and five clinic or hospital-based case–control studies. All studies were approved by the respective Institutional Review Boards as listed. These are ATBC:(NCI Special Studies Institutional Review Board), BCCA: UBC BC Cancer Agency Research Ethics Board, CPS-II: American Cancer Society, ELCCS: Northern and Yorkshire Research Ethics Committee, ENGELA: IRB00003888—Comite d' Evaluation Ethique de l'Inserm IRB # 1, EPIC: Imperial College London, EpiLymph: International Agency for Research on Cancer, HPFS: Harvard School of Public Health (HSPH) Institutional Review Board, Iowa-Mayo SPORE: University of Iowa Institutional Review Board, Italian GxE: Comitato Etico Azienda Ospedaliero Universitaria di Cagliari, Mayo Clinic Case–Control: Mayo Clinic Institutional Review Board, MCCS: Cancer Council Victoria's Human Research Ethics Committee, MD Anderson: University of Texas MD Anderson Cancer Center Institutional Review Board, MSKCC: Memorial Sloan-Kettering Cancer Center Institutional Review Board, NCI-SEER (NCI Special Studies Institutional Review Board), NHS: Partners Human Research Committee, Brigham and Women's Hospital, NSW: NSW Cancer Council Ethics Committee, NYU-WHS: New York University School of Medicine Institutional Review Board, PLCO: (NCI Special Studies Institutional Review Board), SCALE: Scientific Ethics Committee for the Capital Region of Denmark, SCALE: Regional Ethical Review Board in Stockholm (Section 4) IRB#5, UCSF2: University of California San Francisco Committee on Human Research, WHI: Fred Hutchinson Cancer Research Center, Yale: Human Investigation Committee, Yale University School of Medicine. Informed consent was obtained from all participants.

Cases were ascertained from cancer registries, clinics or hospitals or through self-report verified by medical and pathology reports. To determine the NHL subtype, phenotype data for all NHL cases were reviewed centrally at the International Lymphoma Epidemiology Consortium (InterLymph) Data Coordinating Center and harmonized using the hierarchical classification proposed by the InterLymph Pathology Working Group[24,25] based on the World Health Organization (WHO) classification[26].

Genotyping and quality control. All MZL cases with sufficient DNA ($n=890$) and a subset of controls ($n=2,854$) frequency matched by age, sex and study to the entire group of NHL cases, along with 4% quality control duplicates, were genotyped on the Illumina OmniExpress at the NCI Core Genotyping Resource (CGR). Genotypes were called using Illumina GenomeStudio software, and quality control duplicates showed >99% concordance. Monomorphic SNPs and SNPs with a call rate of <95% were excluded. Samples with a call rate of ≤93%, mean heterozygosity <0.25 or >0.33 based on the autosomal SNPs or gender discordance (>5% heterozygosity on the X chromosome for males and <20% heterozygosity on the X chromosome for females) were excluded. Furthermore, unexpected duplicates (>99.9% concordance) and first-degree relatives based on identity by descent sharing with Pi-hat >0.40 were excluded. Ancestry was assessed using the Genotyping Library and Utilities (GLU-http://code.google.com/p/glu-genetics/) struct.admix module based on the method by Pritchard *et al.*[27] and participants with <80% European ancestry were excluded (Supplementary Fig. 1). After exclusions, 825 cases and 2,685 controls remained (Supplementary Table 2). Genotype data previously generated on the Illumina Omni2.5 from an additional 3,536 controls from three of the 22 studies (ATBC, CPS-II and PLCO) were also included[3], resulting in a total of 825 cases and 6,221 controls for the stage 1 analysis (Supplementary Table 3). Of these additional 3,536 controls, 703 (~235 from each study) were selected to be representative of their cohort and cancer free[3], while the remainder were cancer-free controls from an unpublished study of prostate cancer in the PLCO. SNPs with call rate <95%, with Hardy–Weinberg equilibrium $P<1\times10^{-6}$, or with a MAF <1% were excluded from analysis, leaving 611,856 SNPs for analysis. To evaluate population substructure, a principal components analysis was performed using the Genotyping Library and Utilities (GLU), version 1.0, struct.pca module, which is similar to EIGENSTRAT[28] -http://genepath.med.harvard.edu/~reich/Software.htm. Plots of the first five principal components are shown in Supplementary Fig. 2. Genomic inflation factor was computed prior ($\lambda=1.014$) and after removal of SNPs in the HLA loci ($\lambda=1.010$). Association testing was conducted assuming a log-additive genetic model, adjusting for age, sex and three significant principal components. All data analyses and management were conducted using GLU.

Imputation of variants. To more comprehensively evaluate the genome for SNPs associated with MZL, SNPs in the stage 1 discovery GWAS were imputed using IMPUTE2 (ref. 29)-http://mathgen.stats.ox.ac.uk/impute/impute_v2.html and the 1000 Genomes Project (1kGP-http://www.1000genomes.org/) version 3 data[29,30]. SNPs with a MAF <1% or information quality score (info) <0.3 were excluded from analysis, leaving 8,478,065 SNPs for association testing. Association testing on the imputed data was conducted using SNPTEST[31]—https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html version 2, assuming dosages for the genotypes and adjusting for age, sex and three significant principal components. In a null model for MZL risk, the three eigenvectors EV1, EV3 and EV8 were nominally associated with MZL risk and hence were included to account for potential population stratification. Heterogeneity between MZL subtypes was assessed using a case–case comparison, adjusting for age, sex and significant principal components.

Stage 2 replication of SNPs from the GWAS. After ranking the SNPs by P value and LD filtering ($r^2<0.05$), 10 SNPs from the most promising loci identified from stage 1 after imputation with $P<7.5\times10^{-6}$ were taken forward for *de novo* replication in an additional 456 cases and 906 controls (Supplementary Tables 1 and 4). Wherever possible, we selected either the best directly genotyped SNP or

the most significant imputed SNP for the locus. Only imputed SNPs with an information score >0.8 were considered for replication. Only SNPs with MAF >1% were selected for replication, and no SNPs were taken forward for replication in regions where they appeared as singletons or obvious artifacts. For the HLA region, we selected one additional SNP (rs7750641) that was highly correlated with rs2922994 for additional confirmation. Genotyping was conducted using custom TaqMan genotyping assays (Applied Biosystems) validated at the NCI Core Genotyping Resource. Genotyping was done at four centres. HapMap control samples genotyped across two centres yielded 100% concordance as did blind duplicates (~5% of total samples). Due to the small number of samples, the MD Anderson, Mayo and NCI replication studies were pooled together for association testing; however, MSKCC samples were analysed separately to account for the available information on Ashkenazi ancestry. Association results were adjusted for age and gender and study site in the pooled analysis. The results from the stage 1 and stage 2 studies were then combined using a fixed effect meta-analysis method with inverse variance weighting based on the estimates and s.e. from each study. Heterogeneity in the effect estimates across studies was assessed using Cochran's Q statistic and estimating the I^2 statistic. For all SNPs that reached genome-wide significance in Table 1, no substantial heterogeneity was observed among the studies ($P_{\text{heterogeneity}} \geq 0.1$ for all SNPs, Supplementary Table 4).

Technical validation of imputed SNPs. Genotyping was conducted using custom TaqMan genotyping assays (Applied Biosystems) at the NCI Cancer Genomics Research Laboratory on a set of 470 individuals included in the stage 1 MZL GWAS. The allelic dosage r^2 was calculated between the imputed genotypes and the technical validation done using assayed genotypes which showed that both SNPs were imputed with high accuracy (INFO ≥0.99) and a high correlation ($r^2 \geq 0.99$) between dosage imputation and genotypes obtained by Taqman assays.

HLA imputation and analysis. To determine if specific coding variants within HLA genes contributed to the diverse association signals, we imputed the classical HLA alleles (*A*, *B*, *C*, *DQA1*, *DQB1*, *DRB1*) and coding variants across the HLA region (chr6:20–40 Mb) using SNP2HLA[5]-http://www.broadinstitute.org/mpg/snp2hla/. The imputation was based on a reference panel from the Type 1 Diabetes Genetics Consortium (T1DGC) consisting of genotype data from 5,225 individuals of European descent who were typed for *HLA-A*, *B*, *C*, *DRB1*, *DQA1*, *DQB1*, *DPB1*, *DPA1* 4-digit alleles. Imputation accuracy of HLA alleles was assessed by comparing HLA alleles to the HLA sequencing data on a subset of samples from the NCI[32]. The concordance rates obtained were 97.32, 98.5, 98.14 and 97.49% for *HLA-A*, *B*, *C* and *DRB1*, respectively, in the NCI GWAS suggesting robust performance of the imputation method. Due to the limited number of SNPs (7,253) in the T1DGC reference set, imputation of HLA SNPs was conducted with IMPUTE2 and the 1kGP reference set as described above. A total of 68,488 SNPs, 201 classical HLA alleles (two- and four-digit resolution) and 1,038 AA markers including 103 AA positions that were 'multi-allelic' with three to six different residues present at each position, were successfully imputed (info score >0.3 for SNPs or $r^2 > 0.3$ for alleles and AAs) and available for downstream analysis. Multi-allelic markers were analysed as binary markers (for example, allele present or absent) and a meta-analysis was conducted where we tested SNPs, HLA alleles and AAs across the HLA region for association with MZL using PLINK[33] or SNPTEST[31] as described above.

eQTL analysis. We conducted an eQTL analysis using two independent datasets: childhood asthma[34] and HapMap[35]. As described previously[34] for the childhood asthma data set[35], peripheral blood lymphocytes were transformed into lymphoblastoid cell lines for 830 parents and offspring from 206 families of European ancestry. Data from 405 children were used for the analysis as follows: using extracted RNA, gene expression was assessed with the Affymetrix HG-U133 Plus 2.0 chip. Genotyping was conducted using the Illumina Human-1 Beadchip and Illumina HumanHap300K Beadchip, and imputation performed using data from 1kGP. All SNPs selected for replication were tested for *cis* associations (defined as gene transcripts within 1 Mb), assuming an additive genetic model, adjusting for non-genetic effects in the gene expression value. Association testing was conducted using a variance component-based score test[36] in MERLIN[37], which accounts for the correlation between siblings. To gain insight into the relative importance of associations with our SNPs compared with other SNPs in the region, we also conducted conditional analyses, in which both the MZL SNP and the most significant SNP for the particular gene transcript (that is, peak SNP) were included in the same model. Only *cis* associations that reached $P < 6.8 \times 10^{-5}$, which corresponds to a FDR of 1% are reported (Supplementary Table 9).

The HapMap data set consisted of a publicly available RNAseq data set[35] from transformed lymphoblastoid cell lines from 41 CEPH Utah residents with ancestry from northern and western Europe (HapMap-CEU) samples available from the Gene Expression Omnibus repository (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE16921. In this data set, we examined the association between the two reported SNPs in the HLA region, rs2922994 and rs9461741, as well as all SNPs in LD ($r^2 > 0.8$ in HapMap-CEU release 28) and expression levels of probes within 1 Mb of the SNPs. As rs9461741 was not genotyped in HapMap, we selected rs7742033 as a proxy as it was the strongest linked SNP available in HapMap ($r^2 = 0.49$ in 1kGP-CEU). Genotyping data for these HapMap-CEU individuals were directly downloaded from HapMap (www.hapmap.org). Correlation between expression and genotype for each SNP-probe pair was tested using the Spearman's rank correlation test with *t*-distribution approximation and estimated with respect to the minor allele in HapMap-CEU. *P* values were adjusted using the Benjamini–Hochberg FDR correction and eQTLs were considered significant at an FDR<0.05 (Supplementary Table 10).

Bioinformatics ENCODE and chromatin state dynamics. To assess chromatin state dynamics, we used Chromos[38], which has precomputed data from ENCODE on nine cell types using Chip-Seq experiments[39]. These consist of B-lymphoblastoid cells (GM12878), hepatocellular carcinoma cells (HepG2), embryonic stem cells, erythrocytic leukaemia cells (hK562), umbilical vein endothelial cells, skeletal muscle myoblasts, normal lung fibroblasts, normal epidermal keratinocytes and mammary epithelial cells. These precomputed data have genome-segmentation performed using a multivariate hidden Markov model to reduce the combinatorial space to a set of interpretable chromatin states. The output from Chromos lists data into 15 chromatin states corresponding to repressed, poised and active promoters, strong and weak enhancers, putative insulators, transcribed regions and large-scale repressed and inactive domains (Supplementary Fig. 5).

References

1. Doglioni, C., Wotherspoon, A. C., Moschini, A., de Boni, M. & Isaacson, P. G. High incidence of primary gastric lymphoma in northeastern Italy. *Lancet* **339,** 834–835 (1992).
2. Parsonnet, J. *et al. Helicobacter pylori* infection and gastric lymphoma. *New Engl. J. Med.* **330,** 1267–1271 (1994).
3. Wang, Z. *et al.* Improved imputation of common and uncommon SNPs with a new reference set. *Nat. Genet.* **44,** 6–7 (2012).
4. Ku, D. H. *et al.* A new growth-regulated complementary DNA with the sequence of a putative trans-activating factor. *Cell Growth Differ.* **2,** 179–186 (1991).
5. Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8,** e64683 (2013).
6. Mayerle, J. *et al.* Identification of genetic loci associated with *Helicobacter pylori* serologic status. *JAMA* **309,** 1912–1920 (2013).
7. Purdue, M. P. *et al.* A pooled investigation of Toll-like receptor gene variants and risk of non-Hodgkin lymphoma. *Carcinogenesis* **30,** 275–281 (2009).
8. Arnett, H. A. *et al.* BTNL2, a butyrophilin/B7-like molecule, is a negative costimulatory molecule modulated in intestinal inflammation. *J. Immunol.* **178,** 1523–1533 (2007).
9. Nguyen, T., Liu, X. K., Zhang, Y. & Dong, C. BTNL2, a butyrophilin-like molecule that functions to inhibit T cell activation. *J. Immunol.* **176,** 7354–7360 (2006).
10. Swanson, R. M. *et al.* Butyrophilin-like 2 modulates B7 costimulation to induce Foxp3 expression and regulatory T cell development in mature T cells. *J. Immunol.* **190,** 2027–2035 (2013).
11. Kallberg, H. *et al.* Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis. *Am. J. Hum. Genet.* **80,** 867–875 (2007).
12. Ferreira, R. C. *et al.* High-density SNP mapping of the HLA region identifies multiple independent susceptibility loci associated with selective IgA deficiency. *PLoS Genet.* **8,** e1002476 (2012).
13. Ferreira, M. A. *et al.* Sequence variants in three loci influence monocyte counts and erythrocyte volume. *Am. J. Hum. Genet.* **85,** 745–749 (2009).
14. Hystad, M. E. *et al.* Characterization of early stages of human B cell development by gene expression profiling. *J. Immunol.* **179,** 3662–3671 (2007).
15. Dias, C. & Isenberg, D. A. Susceptibility of patients with rheumatic diseases to B-cell non-Hodgkin lymphoma. *Nat. Rev. Rheumatol.* **7,** 360–368 (2011).
16. Lessard, C. J. *et al.* Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. *Nat. Genet.* **45,** 1284–1292 (2013).
17. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44,** 291–296 (2012).
18. Berndt, S. I. *et al.* Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat. Genet.* **45,** 868–876 (2013).
19. Cerhan, J. R. *et al.* Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nat. Genet.* **46,** 1233–1238 (2014).
20. Smedby, K. E. *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.* **7,** e1001378 (2011).
21. Conde, L. *et al.* Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat. Genet.* **42,** 661–664 (2010).

22. Vijai, J. *et al.* Susceptibility loci associated with specific and shared subtypes of lymphoid malignancies. *PLoS Genet.* **9,** e1003220 (2013).
23. Skibola, C. F. *et al.* Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the HLA region. *Am. J. Hum. Genet.* **95,** 462–471 (2014).
24. Morton, L. M. *et al.* Proposed classification of lymphoid neoplasms for epidemiologic research from the Pathology Working Group of the International Lymphoma Epidemiology Consortium (InterLymph). *Blood* **110,** 695–708 (2007).
25. Turner, J. J. *et al.* InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. *Blood* **116,** e90–e98 (2010).
26. Swerdlow, S., Campo, E. & Harris, N. *World Health Organization Classification of Tumours of Haematopoietic and Lymphoid Tissues* (IARC Press, 2008).
27. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).
28. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38,** 904–909 (2006).
29. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).
30. Abecasis, G. R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467,** 1061–1073 (2010).
31. Ferreira, T. & Marchini, J. Modeling interactions with known risk loci-a Bayesian model averaging approach. *Ann. Hum. Genet.* **75,** 1–9 (2011).
32. Wang, S. S. *et al.* Human leukocyte antigen class I and II alleles in non-Hodgkin lymphoma etiology. *Blood* **115,** 4820–4823 (2010).
33. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
34. Dixon, A. L. *et al.* A genome-wide association study of global gene expression. *Nat. Genet.* **39,** 1202–1207 (2007).
35. Cheung, V. G. *et al.* Polymorphic cis- and trans-regulation of human gene expression. *PLoS Biol.* **8,** 9 (2010).
36. Chen, W. M. & Abecasis, G. R. Family-based association tests for genomewide association scans. *Am. J. Hum. Genet.* **81,** 913–926 (2007).
37. Abecasis, G. R. & Wigginton, J. E. Handling marker-marker linkage disequilibrium: pedigree analysis with clustered markers. *Am. J. Hum. Genet.* **77,** 754–767 (2005).
38. Barenboim, M. & Manke, T. ChroMoS: an integrated web tool for SNP classification, prioritization and functional interpretation. *Bioinformatics* **29,** 2197–2198 (2013).
39. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).

Acknowledgements

Support for individual studies: ATBC—Intramural Research Program of the National Institutes of Health, NCI, Division of Cancer Epidemiology and Genetics. U.S. Public Health Service contracts (N01-CN-45165, N01-RC-45035, N01-RC-37004); BCCA (J.J.S., A.R.B.-W.)—Canadian Institutes for Health Research (CIHR). Canadian Cancer Society, Michael Smith Foundation for Health Research; CPS-II (L.F.T.)—The American Cancer Society funds the creation, maintenance and updating of the CPS-II cohort. We thank the CPS-II participants and the Study Management Group for their invaluable contributions to this research. We would also like to acknowledge the contribution to this study from the 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 (E.R.)—Leukaemia & Lymphoma Research; ENGELA (J.C.)—Fondation ARC pour la Recherche sur le Cancer. Fondation de France. French Agency for Food, Environmental and Occupational Health & Safety (ANSES), the French National Cancer Institute (INCa); EPIC (E.R.)—Coordinated Action (Contract #006438, SP23-CT-2005-006438). HuGeF (Human Genetics Foundation), Torino, Italy; EpiLymph—European Commission (grant references QLK4-CT-2000-00422 and FOOD-CT-2006-023103); the Spanish Ministry of Health (grant references CIBERESP, PI11/01810, RCESP C03/09, RTICESP C03/10 and RTIC RD06/0020/0095), the Marató de TV3 Foundation (grant reference 051210), the Agència de Gestiód'AjutsUniversitarisi de Recerca—Generalitat de Catalunya (grant reference 2009SGR1465) 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 RECAMO, CZ.1.05/2.1.00/03.0101; Fondation de France and Association de Recherche Contre le Cancer; HPFS (Walter C. Willet)—The HPFS was supported in part by National Institutes of Health grants CA167552, CA149445, CA098122, CA098566 (K.A.B.) and K07 CA115687 (B.M.B.). 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. In addition, this study was approved by the Connecticut Department of Public Health (DPH) Human Investigations Committee. Certain data used in this publication were obtained from the DPH. The authors assume full responsibility for analyses and interpretation of these data; Iowa-Mayo SPORE (G.J.W., J.R.C., T.E.W.)—National Institutes of Health (CA97274). Specialized Programs of Research Excellence (SPORE) in Human Cancer (P50 CA97274). Molecular Epidemiology of Non-Hodgkin Lymphoma Survival (R01 CA129539). Henry J. Predolin Foundation; Italian GxE (P.C.)—Italian Ministry for Education, University and Research (PRIN 2007 prot.2007WEJLZB, PRIN 2009 prot. 20092ZELR2); the Italian Association for Cancer Research (AIRC, Investigator Grant 11855). (M.G.E.)—Regional Law N. 7, 2007: 'Basic research' (Progetti di ricerca fondamentale o di base) by the Regional Administration of Sardinia (CRP-59812/2012), Fondazione Banco di Sardegna 2010–2012; Mayo Clinic Case–Control (J.R.C.)—National Institutes of Health (R01 CA92153). National Center for Advancing Translational Science (UL1 TR000135); MCCS (G.G.G., G.S.)—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 infrastructure provided by Cancer Council Victoria; MD Anderson (X.W.)—Institutional support to the Center for Translational and Public Health Genomics; MSKCC (K.O.)—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; 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 (Meir J. Stampfer)—The NHS was supported in part by National Institutes of Health grants CA87969, CA49449, CA149445, CA098122, CA098566 (K.A.B.), and K07 CA115687 (B.M.B.). 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. In addition, this study was approved by the Connecticut Department of Public Health (DPH) Human Investigations Committee. Certain data used in this publication were obtained from the DPH. The authors assume full responsibility for analyses and interpretation of these data; NSW (C.M.Vajdic)—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 (K.E.S., H.O.A., H.H.)—Swedish Cancer Society (2009/659). Stockholm County Council (20110209) and the Strategic Research Program in Epidemiology at Karolinska Institute. Swedish Cancer Society grant (02 6661). Danish Cancer Research Foundation Grant. Lundbeck Foundation Grant (R19-A2364). Danish Cancer Society Grant (DP 08–155). National Institutes of Health (5R01 CA69669-02). Plan Denmark; UCSF2 (C.F.S.)—National Institutes of Health RO1CA1046282 and RO1CA154643; (E.A.H., P.M.B.)–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; WHI—WHI investigators are: Program Office—(National Heart, Lung, and Blood Institute, Bethesda, Maryland) Jacques Rossouw, Shari Ludlam, Dale Burwen, Joan McGowan, Leslie Ford, and Nancy Geller; Clinical Coordinating Center—(Fred Hutchinson Cancer Research Center, Seattle, WA) Garnet Anderson, Ross Prentice, Andrea LaCroix, and Charles Kooperberg; Investigators and Academic Centers—(Brigham and Women's Hospital, Harvard Medical School, Boston, MA) JoAnn E. Manson; (MedStar Health Research Institute/Howard University, Washington, DC) Barbara V. Howard; (Stanford Prevention Research Center, Stanford, CA) Marcia L. Stefanick; (The Ohio State University, Columbus, OH) Rebecca Jackson; (University of Arizona, Tucson/Phoenix, AZ) Cynthia A. Thomson; (University at Buffalo, Buffalo, NY) Jean Wactawski-Wende; (University of Florida, Gainesville/Jacksonville, FL) Marian Limacher; (University of Iowa, Iowa City/Davenport, IA) Robert Wallace; (University of Pittsburgh, Pittsburgh, PA) Lewis Kuller; (Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker; Women's Health Initiative Memory Study—(Wake Forest University School of Medicine, Winston-Salem, NC) Sally Shumaker. The WHI program is funded by the National Heart, Lung and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C and HHSN271201100004C; YALE (T.Z.)—National Cancer Institute (CA62006).

Author contributions

J.Vijai, S.I.B., C.F.S., S.L.S., B.M.B., S.S.W., A.R.B.-W., Q.L., H.H., W.C., L.R.T., J.J.S., Y.Z., M.P.P., A.Z.-J., C.L., R.M., K.E.S., P.H., J.M., B.K.A., A.K., G.S., P.V., J.F.F., J.R.C., K.O., S.J.C., N.R. and A.N. organized and designed the study. J.Vijai, S.I.B., L.B., A.H., X.W., J.R.C., K.O., S.J.C. and N.R. conducted and supervised the genotyping of samples. J.Vijai, Z.W., S.I.B., C.F.S., S.d.S., L.C., P.I.W.d.B., J.G., M.Y., C.C.C., L.L., J.H., B.M., S.J.C. and N.R. contributed to the design and execution of statistical analysis. J.Vijai, Z.W., S.I.B., C.F.S., J.R.C., K.O., S.J.C., N.R. and A.N. wrote the first draft of the manuscript. J.Vijai, C.F.S., S.L.S., S.d.S., M.Melbye, B.G., P.M.B., L.C., B.M.B., S.S.W., A.R.B.-W., Q.L., R.C.H.V., C.P., S.M.A., B.K.L., J.R., K.E.N., J.G., H.H., W.C., N.B., L.R.T., J.J.S., J.T., Y.Z., M.P.P., G.G.G., R.S.K., A.Z.-J., M.G.E., A.Monnereau, K.A.B., D.A., T.L., D.J.V., A.Maria, M.C., T.T., A.J.N., A.D., M.L., C.A.T., T.E.W., T.M.H., G.J.W., M.T.S., E.A.H., R.D.J., L.F.T., Y.Y., H.-O.A., K.E.S., A.J.D.R., P.H., L.M.M., R.K.S., Y.B., P.Boffetta, P.Brennan, L.F., M.Maynadie, J.M., A.Staines, W.R.D., C.M.V., B.K.A., A.K., T.Z., T.R.H., G.S., P.V., G.M.F., R.R., L.M., J.C., E.G., P.K., J.Virtamo, A.Smith, E.K., E.R., B.C.H.C., J.F.F., X.W., J.R.C., K.O., N.R. and A.N. conducted the epidemiological studies and contributed samples to the GWAS and/or follow-up genotyping. All authors contributed to the writing of the manuscript.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Vijai, J. *et al.* A genome-wide association study of marginal zone lymphoma shows association to the HLA region. *Nat. Commun.* 6:5751 doi: 10.1038/ncomms6751 (2015).