

Original Article

Genome-wide homozygosity and risk of four non-hodgkin lymphoma subtypes

Amy Moore¹, Mitchell J. Machiela¹, Moara Machado^{1,2}, Sophia S. Wang³, Eleanor Kane⁴, Susan L. Slager⁵, Weiyin Zhou^{6,1}, Mary Carrington^{7,8}, Qing Lan¹, Roger L. Milne^{9,10,11}, Brenda M. Birmann¹², Hans-Olov Adami^{13,14,15}, Demetrius Albanes¹, Alan A. Arslan^{16,17,18}, Nikolaus Becker¹⁹, Yolanda Benavente^{20,21}, Simonetta Bisanzi²², Paolo Boffetta^{23,24}, Paige M. Bracci²⁵, Paul Brennan²⁶, Angela R. Brooks-Wilson^{27,28}, Federico Canzian²⁹, Neil Caporaso¹, Jacqueline Clavel^{30,31}, Pierluigi Cocco³², Lucia Conde³³, David G. Cox³⁴, Wendy Cozen^{35,36}, Karen Curtin³⁷, Immaculata De Vivo^{12,14}, Silvia de Sanjose^{1,21}, Lenka Foretova³⁸, Susan M Gapstur³⁹, Hervé Ghesquières^{40,41}, Graham G. Giles^{9,10,11}, Martha Glenn³⁷, Bengt Glimelius⁴², Chi Gao¹⁴, Thomas M. Habermann⁴³, Henrik Hjalgrim⁴⁴, Rebecca D. Jackson⁴⁵, Mark Liebow⁴³, Brian K. Link⁴⁶, Marc Maynadie⁴⁷, James McKay²⁶, Mads Melbye^{44,48}, Lucia Miligi⁴⁹, Thierry J. Molina⁵⁰, Alain Monnereau^{30,51}, Alexandra Nieters⁵², Kari E. North^{53,54}, Kenneth Offit⁵⁵, Alpa V. Patel³⁹, Sara Piro⁴⁹, Vignesh Ravichandran⁵⁵, Elio Riboli⁵⁶, Gilles Salles^{57,58,41}, Richard K. Severson⁵⁹, Christine F. Skibola⁶⁰, Karin E. Smedby^{61,62}, Melissa C Southey⁶³, John J. Spinelli^{64,65}, Anthony Staines⁶⁶, Carolyn Stewart⁵⁵, Lauren R. Teras³⁹, Lesley F. Tinker⁶⁷, Ruth C. Travis⁶⁸, Claire M. Vajdic⁶⁹, Roel C. H. Vermeulen^{70,71}, Joseph Vijai⁵⁵, Elisabete Weiderpass²⁶, Stephanie Weinstein¹, Nicole Wong Doo⁷², Yawei Zhang⁷³, Tongzhang Zheng⁷⁴, Stephen J. Chanock¹, Nathaniel Rothman¹, James R. Cerhan⁵, Michael Dean¹, Nicola J. Camp³⁷, Meredith Yeager^{6,1}, Sonja I. Berndt¹

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, 20892, USA.

²Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, 31270-901, Brazil.

³Division of Health Analytics, City of Hope Beckman Research Institute, Duarte, California, 91010, USA.

⁴Department of Health Sciences, University of York, York, YO10 5DD, United Kingdom.

⁵Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, 55905, USA.

⁶Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, Maryland, 20877, USA.

⁷Basic Science Program, Frederick National Laboratory for Cancer Research in the Laboratory of Integrative Cancer Immunology, National Cancer Institute, Bethesda, Maryland, 20892, USA.

⁸Ragon Institute of MGH, Cambridge, MA, 02139, USA.

⁹Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, 3004, Australia.

¹⁰Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, 3010, Australia.

¹¹Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, 3800, Australia.

¹²Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, 02115, USA.

¹³Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, 17176, Sweden.

¹⁴Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, 02115, USA.

¹⁵Institute of Health and Society, Clinical Effectiveness Research Group, University of Oslo, Oslo, 0315, Norway.

¹⁶Department of Obstetrics and Gynecology, New York University School of Medicine, New York, New York, 10016, USA.

¹⁷Department of Population Health, New York University School of Medicine, New York, New York, 10016, USA.

¹⁸Perlmutter Comprehensive Cancer Center, NYU Langone Health, New York, New York, 10016, USA.

¹⁹Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Baden-Württemberg, 69120, Germany.

²⁰Cancer Epidemiology Research Programme, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Barcelona, 08908, Spain.

²¹Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Barcelona, 08036, Spain.

²²Regional Cancer Prevention Laboratory, Institute for Cancer Research, Prevention and Clinical Network (ISPRO), Florence, 50139, Italy.

²³Stony Brook Cancer Center, Stony Brook University, Stony Brook, New York, 11794 USA.

²⁴Department of Medical and Surgical Sciences, University of Bologna, Bologna, 41026, Italy.

²⁵Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, 94118, USA.

²⁶International Agency for Research on Cancer (IARC), Lyon, 69372, France.

²⁷Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia, V5Z1L3, Canada.

²⁸Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, V5A1S6, Canada.

²⁹Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany.

³⁰CRESS, UMR1153, INSERM, Villejuif, 75004, France.

³¹Université de Paris, Villejuif, 94800, France.

³²Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Monserrato, Cagliari, 09042, Italy.

³³Bill Lyons Informatics Centre, UCL Cancer Institute, University College London, London, WC1E 6DD, United Kingdom.

³⁴INSERM U1052, Cancer Research Center of Lyon, Centre Léon Bérard, Lyon, 69008, France.

³⁵Department of Preventive Medicine, USC Keck School of Medicine, University of Southern California, Los Angeles, California, 90033, USA.

³⁶Norris Comprehensive Cancer Center, USC Keck School of Medicine, University of Southern California, Los Angeles, California, 90033, USA.

³⁷Department of Internal Medicine and Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah, 84112, USA.

³⁸Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, 656 53, Czech Republic.

³⁹Department of Population Science, American Cancer Society, Atlanta, Georgia, 30303, USA.

⁴⁰Department of Hematology, Centre Léon Bérard, Lyon, 69008, France.

⁴¹INSERM U1052, Cancer Research Center of Lyon, Lyon-1 University, Pierre-Bénite Cedex, 69008, France.

⁴²Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, 75105, Sweden.

⁴³Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota, 55905, USA.

⁴⁴Department of Epidemiology Research, Division of Health Surveillance and Research, Statens Serum Institut, Copenhagen, 2300, Denmark.

⁴⁵Division of Endocrinology, Diabetes and Metabolism, The Ohio State University, Columbus, Ohio, 43210, USA.

⁴⁶Department of Internal Medicine, Carver College of Medicine, The University of Iowa, Iowa City, Iowa, 52242, USA.

⁴⁷U1231, Registre des Hémopathies Malignes de Côte d'Or, University of Burgundy and Dijon University Hospital, Dijon, 21070, France.

⁴⁸Department of Medicine, Stanford University School of Medicine, Stanford, California, 94305, USA.

⁴⁹Environmental and Occupational Epidemiology Branch-Cancer Risk Factors and Lifestyle Epidemiology Unit, Institute for Cancer Research, Prevention and Clinical Network (ISPRO), Florence, 50139, Italy.

⁵⁰Department of Pathology, AP-HP, Necker Enfants Malades, Université Paris Descartes, EA 7324, Sorbonne Paris Cité, 75015, France.

⁵¹Registre des Hémopathies Malignes de la Gironde, Institut Bergonié, Bordeaux Cedex, 33076, France.

⁵²Center for Chronic Immunodeficiency, University Medical Center Freiburg, Freiburg, Baden-Württemberg, 79108, Germany.

⁵³Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA.

⁵⁴Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599, USA.

⁵⁵Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, 10065, USA.

⁵⁶School of Public Health, Imperial College London, London, W2 1PG, United Kingdom.

⁵⁷Department of Hematology, Hospices Civils de Lyon, Pierre Benite Cedex, France.

⁵⁸Department of Hematology, Université Lyon-1, Pierre Benite Cedex, 69495, France.

⁵⁹Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, Michigan, 48201, USA.

⁶⁰Department of Hematology and Medical Oncology, Emory University School of Medicine, Atlanta, Georgia, 30322, USA.

⁶¹Department of Medicine, Solna, Karolinska Institutet, Stockholm, 17176, Sweden.

⁶²Hematology Center, Karolinska University Hospital, Stockholm, 17176, Sweden.

⁶³Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne, Melbourne, Victoria, 3010, Australia.

⁶⁴Cancer Control Research, BC Cancer Agency, Vancouver, British Columbia, V5Z1L3, Canada.

⁶⁵School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, V6T1Z3, Canada.

⁶⁶School of Nursing, Psychotherapy and Community Health, Dublin City University, Dublin, 9, Ireland.

⁶⁷Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, 98117, Washington, USA.

⁶⁸Cancer Epidemiology Unit, University of Oxford, Oxford, OX3 7LF, United Kingdom.

⁶⁹Centre for Big Data Research in Health, University of New South Wales, Sydney, New South Wales, 2052, Australia.

⁷⁰Institute for Risk Assessment Sciences, Utrecht University, Utrecht, 3584 CG, The Netherlands.

⁷¹Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, 3584 CX, The Netherlands.

⁷²Concord Clinical School, University of Sydney, Concord, New South Wales, 2139, Australia.

⁷³Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut, 06520, USA.

⁷⁴Department of Epidemiology, Brown University, Providence, Rhode Island, 02903, USA.

Correspondence to: Sonja Berndt, Division of Cancer Epidemiology & Genetics, National Cancer Institute, NIH, 9609 Medical Center Drive, Room 6E610, MSC 9771, Rockville, MD 20850, USA. E-mail: berndts@mail.nih.gov

How to cite this article: Moore A, Machiela MJ, Machado M, Wang SS, Kane E, Slager SL, Zhou W, Carrington MN, Lan Q, Milne RL, Birmann BM, Adami HO, Albanes D, Arslan AA, Becker N, Benavente Y, Bisanzi S, Boffetta P, Bracci PM, Brennan P, Brooks-Wilson AR, Canzian F, Caporaso NE, Clavel J, Cocco P, Conde L, Cox DG, Cozen W, Curtin K, De Vivo I, de Sanjose S, Foretova L, Gapstur SM, Ghesquière H, Giles GG, Glenn M, Glimelius B, Gao C, Habermann TM, Hjalgrim H, Jackson RD, Liebow M, Link BK, Maynadie M, McKay J, Melbye M, Miligi L, Molina TJ, Monnereau A, Nieters A, North KE, Offit K, Patel AV, Piro S, Ravichandran V, Riboli E, Salles G, Severson RK, Skibola CF, Smedby KE, Southey MC, Spinelli JJ, Staines A, Stewart C, Teras LR, Tinker LF, Travis RC, Vajdic CM, Vermeulen RCH, Vijai J, Weiderpass E, Weinstein S, Wong Doo N, Zhang Y, Zheng T, Chanock SJ, Rothman N, Cerhan JR, Dean M, Camp NJ, Yeager M, Berndt SI. Genome-wide homozygosity and risk of four non-hodgkin lymphoma subtypes. *J Transl Genet Genom* 2021;5:[Accept]. <http://dx.doi.org/10.20517/jtgg.2021.08>

Received: 1 Mar 2021 **Revised:** 13 May 2021 **Accepted:** 28 May 2021 **First online:** 17 Jun 2021

ABSTRACT

Aim: Recessive genetic variation is thought to play a role in non-Hodgkin lymphoma (NHL) etiology. Runs of homozygosity (ROH), defined based on long, continuous segments of homozygous SNPs, can be used to estimate both measured and unmeasured recessive genetic variation. We sought to examine genome-wide homozygosity and NHL risk.

Methods: We used data from eight genome-wide association studies (GWAS) of four common NHL subtypes: 3061 chronic lymphocytic leukemia (CLL), 3814 diffuse large B-cell lymphoma (DLBCL), 2784 follicular lymphoma (FL), and 808 marginal zone lymphoma (MZL) cases, as well as 9374 controls. We examined the effect of homozygous variation on risk by: (1) estimating the fraction of the autosome containing runs of homozygosity (FROH); (2) calculating an inbreeding coefficient derived from the correlation among uniting gametes (F3); and (3) examining specific autosomal regions containing ROH. For each, we calculated beta coefficients and standard errors using logistic regression and combined estimates across studies using random-effects meta-analysis.

Results: We discovered positive associations between FROH and CLL ($\beta = 21.1$, $SE = 4.41$, $P = 1.6 \times 10^{-6}$) and FL ($\beta = 11.4$, $SE = 5.82$, $P = 0.02$) but not DLBCL ($P = 1.0$) or MZL ($P = 0.91$). For F3, we observed an association with CLL ($\beta = 27.5$, $SE = 6.51$, $P = 2.4 \times 10^{-5}$). We did not find evidence of associations with specific ROH, suggesting that the associations observed with FROH and F3 for CLL and FL risk were not driven by a single region of homozygosity.

Conclusion: Our findings support the role of recessive genetic variation in the etiology of CLL and FL; Additional research is needed to identify the specific loci associated with NHL risk.

Keywords: Non-Hodgkin lymphoma, homozygosity, chronic lymphocytic leukemia, follicular lymphoma, diffuse large B-cell lymphoma, marginal zone lymphoma

INTRODUCTION

Disentangling the heritable component of non-Hodgkin lymphoma (NHL) and its subtypes is an active area of research. An early study of familial aggregation in NHL reported an increased risk of NHL among siblings, but not parents or offspring, of an index NHL case.¹ Several subsequent studies found an elevated risk of NHL associated with a first-degree family history of NHL with the highest risks for siblings.^{2,3} Studies have also reported higher NHL subtype-specific risks for first-degree relatives of cases affected with a given NHL subtype,^{4,5} suggesting a degree of subtype specificity. In general, these findings suggest genetic factors are important in NHL etiology and, in particular, the potential role of recessively acting genetic risk alleles, but they also underscore the potential for genetic heterogeneity in susceptibility to different NHL subtypes. Part of the difficulty in characterizing risk and inheritance patterns of NHL subtypes is the limited study sample size, especially when examining specific subtypes. Genome-wide association studies (GWAS) have identified multiple susceptibility loci associated with four major subtypes of NHL,⁶⁻¹¹ but a substantial fraction of the disease heritability remains unexplained. Most GWAS performed assume an additive model of genetic risk, which has statistical power to detect allelic associations acting through a variety of mechanisms but may not efficiently detect recessive effects, particularly as minor allele frequency and imputation

quality decrease.¹² Therefore, recessively acting loci, particularly those with low minor allele frequency, could be missed by current genome-wide scans and represent potential novel disease-associated loci.

The widespread use of dense genotyping arrays has led to the identification of sizeable genomic regions consisting of consecutive homozygous SNPs in non-consanguineous populations.¹³ These runs of homozygosity (ROH) vary in length, with short ROH persisting from ancient relatedness and long ROH of several megabases arising from recent parental relatedness.¹⁴ The use of ROH as a measure of the burden of homozygosity has been demonstrated to perform better at identifying rare, recessive mutations than a conventional SNP-by-SNP analysis.¹⁵ Furthermore, studies incorporating whole-exome sequencing have uncovered an enrichment of deleterious variants in ROH.¹⁶ In recent years, studies have examined the association between ROH and various cancers,¹⁷⁻²¹ among other complex common diseases and traits.²² Although no association was observed with the cumulative distribution of ROH, individual ROH were associated with the risk of childhood acute lymphoblastic leukemia.¹⁹ Hodgkin lymphoma has been inconsistently associated with specific ROH and overall homozygosity.^{20,21} To our knowledge, no studies have examined ROH in association with adult NHL.

The goal of the present study was to investigate the association of homozygosity with the risk of four major NHL subtypes: chronic lymphocytic leukemia/small cell lymphoma (CLL), diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and marginal zone lymphoma (MZL). Several measures of homozygosity were tested against NHL risk using data from eight GWAS.

METHODS

We used data from eight previous GWAS of NHL^{6,8,10,11} composed of cases and controls of European ancestry (**Tables S1 and S2**). The National Cancer Institute (NCI) NHL GWAS included cases with one of four common NHL subtypes and controls from twenty-two studies of NHL: nine prospective cohort studies, eight population-based case-control studies, and five hospital- or clinic-based case-control studies or case series. These 22 studies comprising the NCI NHL GWAS were genotyped using the Illumina OmniExpress or Omni2.5 arrays and analyzed

as a single study. The other seven GWAS were the University of California at San Francisco Molecular Epidemiology of Non-Hodgkin Lymphoma study (UCSF2),²³ the University of California at San Francisco Molecular Epidemiology of Non-Hodgkin Lymphoma study (UCSF1) combined with controls from the Nurses' Health Study (NHS),²⁴ the Scandinavian Lymphoma Etiology Study (SCALE),²⁵ the Groupe d'Etude des Lymphomes de l'Adulte (GELA) study combined with controls from the European Prospective Investigation into Cancer, Chronic Diseases, Nutrition, and Lifestyles (EPIC), the Mayo Clinic Case–Control Study of Diffuse Large B-cell Lymphoma (Mayo), the Genetic Epidemiology of CLL Consortium (GEC), and the Utah Chronic Lymphocytic Leukemia Study (Utah). Genotyping was performed on commercially available Illumina and Affymetrix platforms (**Table S1**). Details, including information on quality control and data cleaning, have been previously reported.^{6,8,10,11} All studies obtained informed consent from participants and were approved by their appropriate Institutional Review Boards.

Prior to analysis, additional quality control and filtering were applied to each GWAS separately, including removal of SNPs with a minor allele frequency < 0.05, >3% missing, or Hardy–Weinberg p -value < 1×10^{-6} among controls, and removal of subjects with call rates < 97%. After quality control metrics, genotype data were available for 10,467 NHL cases, including 3061 CLL, 3814 DLBCL, 2784 FL, and 808 MZL cases, as well as 9374 controls (**Table S2**).

We used PLINK1.9^{26,27} to identify runs of homozygosity (ROH); specifically, we used the two-step command *--homozyg*. In the first step, PLINK1.9 identifies directly genotyped SNPs that are possibly within an ROH by looking at 50-SNP sliding windows across the genome and flagging all SNPs that are encompassed by at least 5% of fully homozygous windows. For this step, we allowed one heterozygous SNP and up to five SNPs with no calls within each window to account for a small amount of possible genotyping error and loss. In the second step, ROH are identified from these sliding windows by requiring a minimum number of consecutive homozygous SNPs. We required at least 100 consecutive homozygous SNPs for each ROH and that these SNPs span at least 1500 kilobases (kb), with at least one SNP every 50 kb and the maximum gap between SNPs of 5000 kb. These parameters were selected with reference to the “ROH_1.5Mb” ROH calling parameters used by Gazal et al.²⁸ We restricted analyses to the autosomal chromosomes.

To estimate the extent of homozygosity across the genome, we calculated the fraction of the autosome covered by ROH (FROH) by summing the lengths of ROH and dividing by 3×10^9 base pairs as the approximate size of the autosome for all GWAS. As another measure to assess homozygosity, we also quantified and tested differences in relatedness across the genome in our study using a variant of the inbreeding coefficient, F3.²⁹ F3, which estimates the correlation between uniting gametes, is an alternative to ROH-based estimates with potentially reduced bias and standard errors.³⁰ We estimated F3 using the *-ibc* command in PLINK1.9. To estimate the association of FROH and F3 with NHL, we then estimated beta coefficients and standard errors for each GWAS using logistic regression, adjusting for age, sex (except in the UCSF1/NHS study, where all controls were female), fraction of missing SNPs, and the ten principal components of ancestry to account for population stratification. The fraction of missing SNPs was calculated for each participant as the number of SNPs without calls divided by the total number of SNPs genotyped on the array that passed quality control metrics. Associations were combined across GWAS for each subtype of NHL using random-effects meta-analysis implemented with the command “*metan*” in STATA v15.

After determining ROH as described above, we also tested whether specific genomic regions encompassed by ROH were associated with risk of each of the four NHL subtypes. We divided each autosomal chromosome into “bins” of 500 kb in length. We then calculated the midpoint of each identified ROH and assigned it to the corresponding bin. Each study participant in the analysis was therefore categorized as either homozygous (exposed) or heterozygous (unexposed) at each bin across the autosome. We calculated beta coefficients and standard errors for the association between presence of an ROH in each bin and risk of NHL subtype within each GWAS using logistic regression, adjusting for age, sex (except in the UCSF1/NHS study), fraction of missing SNPs, and ten principal components of ancestry. Results were combined across GWAS using METAL,³¹ and multiple-testing adjustment was performed using a Bonferroni correction.

RESULTS

Table 1 presents summary statistics for the ROH, FROH, and F3 by GWAS. Among participants, the median total length of ROH ranged from 11,535 to 23,014 kb depending on the GWAS. The median number of ROH per individual ranged from 4 in the UCSF2 GWAS, which used an older and less dense GWAS chip, to 8 in the GEC GWAS, which included familial CLL cases. Median FROH ranged from 0.38% to 0.77% of the autosome. Median F3 ranged from -6.99×10^{-4} to 3.02×10^{-3} (**Table 1**).

We discovered a positive association between the risk of CLL and increased homozygosity as measured by FROH ($\beta=21.1$, 95% SE=4.41, $p=1.6 \times 10^{-6}$) and F3 ($\beta=27.5$, SE=6.51, $p=2.4 \times 10^{-5}$) (**Table 2**) with limited evidence of between-study heterogeneity ($p_{\text{het}}=0.42$ and 0.11, respectively) (**Table S3**). As CLL is an indolent lymphoma and there is a potential for tumor DNA contamination in the blood drawn for genotyping, we performed a sensitivity analysis using only CLL cases and controls from prospective nested case-control studies (ATBC, CPS-II, EPIC, HPFS, MCCS, NHS, NYU-WHS, PLCO, and WHI) from the NCI NHL GWAS, where the DNA was often collected many years prior to diagnosis. Despite the reduction in the number of CLL cases ($n=2140$ to 889), the estimated association parameters for FROH and F3 were similar (FROH: $\beta=21.3$, SE=7.76, $p=6.04 \times 10^{-3}$; F3: $\beta=22.5$, SE=5.45, $p=3.57 \times 10^{-5}$).

Figure 1A shows the p-values ($-\log_{10}$) from the meta-analysis assessing the associations of CLL with ROH centered in each 500 kb bin in the autosome. No bins reached statistical significance after correction for multiple testing (CLL Bonferroni alpha level = $0.05/45,590$ bins = 1.1×10^{-5}); the most significantly associated bin was located at chromosome 22q12.2 ($P = 5.92 \times 10^{-4}$). However, one of the top ten associated bins overlapped with the chromosome 13q14 region, a region where a somatic deletion is often seen in CLL.³² To test whether the association between FROH and CLL was due to mosaicism at 13q14, we excluded the 45 CLL cases identified to have this deletion from the NCI NHL GWAS. After removal of these 45 cases, the association between FROH and CLL was slightly attenuated but remained statistically significant ($\beta=16.34$, SE=4.74, $p=5.71 \times 10^{-4}$), similar to the results from the entire NCI NHL GWAS for CLL ($\beta=18.9$, SE=4.76, $p=6.73 \times 10^{-5}$).

As structural alterations in chromosomes, such as trisomy 12, are a hallmark of CLL and could potentially be mistakenly called as ROH using this method, we conducted a sensitivity analysis excluding chromosomes 12 and 13, which are frequently altered in CLL, from the meta-analysis.³³ After excluding these two chromosomes, the results for the association between FROH and CLL were found to be similar ($\beta=19.3$, $SE=4.4$, $p=9.0 \times 10^{-6}$) to previous results for all 22 autosomal chromosomes. Similarly, for F3, exclusion of chromosomes 12 and 13 led to results that are similar in magnitude and remain statistically significant ($\beta=17.6$, $SE=7.4$, $p=0.02$). Between-study heterogeneity ($I^2=62.2\%$, $p_{het}=0.05$) was detected in this sensitivity analysis, largely attributable to the GEC GWAS.

We also identified positive associations between FL risk and increased FROH ($\beta=11.4$, $SE=5.82$, $p=0.02$) but not F3 ($\beta=13.2$, $SE=8.01$, $p=0.10$) (**Table 2**). There was evidence of between-study heterogeneity in the meta-analysis of F3 and FL ($I^2 = 64.2\%$, $p=0.04$) due to the UCSF1/NHS GWAS, but not in the meta-analysis of FROH and FL ($I^2 = 5.3\%$, $p=0.37$) (**Table S3**). We performed a sensitivity analysis using only cases and controls collected from prospective nested case-control studies (ATBC, CPS-II, EPIC, HPFS, MCCS, NHS, NYU-WHS, PLCO, and WHI) as part of the NCI NHL GWAS. This resulted in a reduction of the number of cases ($n=2085$ to 521) and statistical power, but the association between FROH and risk of FL was qualitatively similar ($\beta =14.1$, $SE=11.4$, $p=0.21$). Given the importance of the HLA region in FL risk, we conducted a sensitivity analysis removing chromosome 6 from estimations of FROH and F3 and repeating the meta-analysis of the four GWAS. The results for both FROH ($\beta=12.5$, $SE=6.29$, $p=0.047$) and F3 ($\beta=11.1$, $SE=9.26$, $p=0.23$) were found to be similar to the between-study heterogeneity observed for F3 ($I^2=73.1\%$, $p=0.01$) due to the UCSF1/NHS GWAS.

Figure 1B shows the p-values ($-\log_{10}$) from the meta-analysis of logistic regression models estimating the associations of FL with ROH centered in each 500-kb bin in the autosome. No bins reached statistical significance assuming a Bonferroni-corrected alpha level. The most significant association was for a bin overlapping the HLA region 6p21.33 ($p=2.78 \times 10^{-3}$).

No associations were observed with either FROH or F3 and the risk of DLBCL or MZL (Table 2); however, between-study heterogeneity in DLBCL risk was present for both FROH ($I^2 =$

72.5%, $p_{\text{het}} = 0.01$) and F3 ($I^2 = 82.4\%$, $p_{\text{het}}=0.001$) (**Table S3**). In a leave-one-out sensitivity analysis (**Table S4**), removal of either the NCI NHL GWAS or the GELA/EPIC GWAS, which had association results in opposite directions, reduced the I^2 among the remaining three GWAS to 0%. Without the GELA/EPIC GWAS, there were significant associations between FROH and DLBCL ($\beta=13.1$, $SE=4.62$, $p=0.004$) and between F3 and DLBCL ($\beta=13.4$, $SE=3.18$, $p=2.4 \times 10^{-5}$).

DISCUSSION

To our knowledge, the present study represents the first analysis of the association between homozygosity and risk of four major NHL subtypes. We assessed genome-wide homozygosity using F3 and FROH, where the beta coefficient represents the log odds for a one-unit change in the fraction of ROH. We discovered that increased genome-wide homozygosity was associated with an increased risk of CLL and FL. We estimated that a 0.1% increase in the fraction of ROH was associated with a 2% increase in CLL risk and 1% increase in FL risk compared to those without any ROH. While we did not identify specific regions containing novel associations with these four subtypes of NHL, our analyses suggest that recessive genetic variation is likely to play a role in the risk of these two NHL subtypes and may account for a fraction of the unexplained heritability of these diseases. Our study did not find convincing evidence that recessive variation is a significant contributor to the risk of DLBCL and MZL, further supporting the hypothesis that each NHL subtype is likely to have its own genetic architecture of disease susceptibility.

Evidence suggests that CLL has high heritability with first-degree relatives having a 6–8.5-fold increase in risk of CLL.^{5,34} Approximately 17–25% of the familial relative risk of CLL is explained by currently known risk loci,^{6,9} leaving a sizable fraction that remains undiscovered. We discovered strong associations between CLL and genome-wide homozygosity as measured by FROH and F3. These findings suggest that recessive genetic variants are likely to contribute to risk.

CLL is a hematologic malignancy characterized by large-scale chromosomal alterations³⁵ and is often preceded by a long-lasting pre-malignant stage of monoclonal B-cell lymphocytosis (MBL).³⁶ It is possible that cases with diagnosed CLL at the time of enrollment or those with unidentified MBL had tumor DNA in their blood samples, and that some mosaic events may

have been erroneously picked up as ROH or homozygous genotypes due to the calling method utilized. This could have inflated our association results, but our sensitivity analyses suggested that any influence of mosaicism on our findings was likely to be small. We excluded CLL cases with DNA collected after diagnosis in the NCI NHL GWAS and saw similar results to our primary analysis, suggesting that tumor contamination is unlikely to be responsible for the observed association. Further sensitivity analysis excluding individuals with known 13q14 deletion in NCI NHL GWAS showed that the associations between FROH and F3 and CLL were only slightly attenuated. We also removed chromosomes 12 and 13 from the calculation of FROH and F3, as these two chromosomes are frequently altered in newly diagnosed CLL,³³ and again saw results similar to the primary analysis. While we cannot exclude the possibility that some of the association could be due to erroneous homozygosity calling of structural variation, the robustness of the associations indicates that potential bias from these sources probably does not explain the observed association.

We did not identify any specific homozygous regions that were associated with CLL risk after adjustment for multiple comparisons, but we observed nominal associations with ROH at chromosomes 22q12.2 and 13q14.2. The chromosome 22q12.2 region includes POZ/BTB and AT hook containing zinc finger 1 (*PATZI*), which is a transcription factor that interacts with p53 and may modulate p53-mediated apoptosis.³⁷ The region also contains PI3K interacting protein 1 (*PIK3IP1*), which inhibits antitumor immunity.³⁸ The chromosome 13q14.2 region contains several genes of potential interest, including RB transcriptional corepressor 1 (*RBI*) and cysteinyl-leukotriene receptors 2 (*CYSLTR2*). *RBI* is a known tumor suppressor, which is frequently somatically deleted in CLL.³² *CYSLTR2* is part of the leukotriene pathway and thought to play a role in lymphocyte proliferation and maturation.³⁹

For FL, we discovered a positive association between FROH and risk of FL, suggesting recessive genetic variation is a contributor to risk. Previous epidemiologic work has shown that family history of NHL is a risk factor for FL,⁴⁰ and first-degree relatives have a 4–6-fold increase risk of FL.^{4,5} The largest GWAS of FL to date detected multiple genetic loci associated with risk, with the strongest associations seen in the HLA region.¹⁰ Consistent with our findings, a previous study of first-degree relatives suggested that FL may follow an autosomal recessive mode of

inheritance.⁵ In sensitivity analyses, we showed that results for FROH are similar after excluding the entirety of chromosome 6, suggesting that our findings were not due to the HLA region. This finding supports the role of non-HLA variation in the etiology of FL and provides evidence that there may be additional rare, recessive loci that are associated with the risk of FL. We observed similar results after limiting to cases with prospectively collected DNA, suggesting that tumor contamination is unlikely. Although we did not find any individual regions (based on 500-kb bins of ROH) that were significantly associated with the risk of FL after adjustment for multiple testing, we observed nominal evidence for homozygosity overlapping the HLA region at chromosome 6p21.33. We previously reported that homozygosity for classical HLA class II alleles, such as *HLA-DRB1*, was associated with an increased risk of FL compared to individuals with heterozygosity.⁴¹ These findings are consistent with the hypothesis that HLA homozygosity may increase risk by reducing an individual's ability to recognize a diverse set of foreign antigens.

DLBCL shows a strong association with family history of NHL in epidemiologic studies,⁴² and it has been previously estimated that 16% of the variation in DLBCL is due to common genetic variants.⁸ Overall, we did not observe an association between DLBCL and homozygosity, suggesting the recessive variation may not have a strong role in DLBCL susceptibility, a finding consistent with a previous study of DLBCL among first-degree relatives.⁵ The meta-analysis for DLBCL was affected by substantial between-study heterogeneity ($p_{\text{het}}=0.01$ for FROH and $p_{\text{het}}=0.001$ for F3). In the leave-one-out meta-analysis, removing either the NCI NHL GWAS or the GELA/EPIC GWAS reduced the I^2 among the remaining three GWAS to 0%, and significant associations were observed with FROH and F3 after removing GELA/EPIC. The GELA/EPIC GWAS was different from the other studies in that it included patients from clinical trials, who were slightly younger and less likely to be female than the other studies (**Table S2**). In addition, this GWAS combined cases from French clinical trials with controls from various European countries in the EPIC cohort. It is possible that the degree of population matching, while adequate for a GWAS, is not sufficient for analyses of F3 and FROH, which are known to be especially sensitive to population stratification.⁴³ Even after adjustment using principal components, some population substructure may have affected the GELA/EPIC study, leading to results that differ from the other DLBCL studies. Our analysis of FL similarly included one

GWAS that combined the cases and controls from two different U.S. studies. Although we did observe some heterogeneity among the FL studies for F3, we did not observe such heterogeneity for FROH and FL, suggesting the F3 may be more sensitive to population substructure than FROH. The patients in the UCSF1/NHS study were slightly younger and less likely to be female than the cases in the NCI GWAS, which could have also contributed to the heterogeneity in results.

DLBCL is known to have substantial disease heterogeneity.^{44,45} It is possible that a greater understanding of the heterogeneity of DLBCL and a further molecular subtype-specific analysis, which was not possible for the present study, may allow a better elucidation of any association between ROH and disease risk. While the analysis of F3 and DLBCL was similarly affected by between-study heterogeneity (**Table 3**), a potential advantage of F3 over FROH as a measure of homozygosity is its expected smaller variance and bias.³⁰ In addition, the method of calculating FROH used in this study does not count homozygosity below a length threshold of 1500 kilobases.²⁸ It is possible that a portion of the recessive genetic variation in DLBCL resides in short, common ROH of ancient ancestry¹⁴ that was unmeasured by our approach.

Although family history of NHL or other hematologic cancers has been associated with increased risk of MZL,⁴⁶ we did not find evidence linking FROH or F3 to MZL. Our published GWAS of MZL found two independent loci conferring risk, both in the HLA region of the genome,¹¹ and a follow-up study reported that homozygosity at class I HLA-B and -C and class II HLA-DRB1 loci was associated with increased risk of MZL.⁴¹ Although our study did not detect an association with homozygosity more broadly across the genome, local homozygosity, at least in select regions, is still likely to play a role in risk. Of the four subtypes of NHL examined in the present study, MZL is the least common. Our power to detect an association with MZL was limited due to the small sample size, which was possibly compounded by the known heterogeneity within MZL, with multiple recognized subtypes.⁴⁷ Given that we expect ROH to capture the effects of small, scattered, recessive genotypes, a larger sample size is likely needed to elucidate the genetic etiology of MZL. As other highly heritable traits, such as certain autoimmune diseases and atopy, are associated with the risk of MZL,⁴⁶ identifying shared genetic

architecture with these more common phenotypes may be a complementary strategy to further elucidate the underlying genetic architecture.

This study had both limitations and strengths. The use of FROH as a measure of homozygosity is known to require large sample sizes.³⁰ We were able to combine eight GWAS across four subtypes of NHL to increase our sample size and provide a fairly comprehensive analysis of recessive inheritance. However, as discussed, we may have had insufficient sample size to detect associations for MZL. Further, we could not separately examine clinically relevant molecular subtypes of DLBCL. ROH patterns are exquisitely sensitive to population history.⁴⁸ Thus, we restricted our study population to individuals of European ancestry and adjusted for principal components of ancestry. This limited the potential for problematic population stratification in our study sample, but it also reduced the generalizability of the findings to individuals of other ancestries. To date, the optimal approach to analyze specific ROH to identify genomic regions with novel, recessively acting risk alleles has not been established, but both the ROH calling procedure and the use of F3 are accepted as means of identifying recent relatedness in outbred European populations^{28,30} and can provide evidence for the presence of recessive loci. Although it is possible that the associations between FROH and F3 and CLL are due in part to DNA contamination by tumor cells, our sensitivity analyses suggest that germline homozygosity is likely an independent risk factor for CLL. Finally, although our study provides clues as to the genetic etiology of NHL, the clinical value of these findings is uncertain. Additional studies are needed to further elucidate the role of recessive variation in NHL risk.

In conclusion, we provide new evidence for the role of recessive genetic variation in the risk of CLL and FL in outbred European-ancestry populations. The knowledge that recessive variation in disease susceptibility to NHL is likely to be present suggests that further studies should be undertaken to identify the specific loci responsible for the associations reported here. As GWAS increase in sample size, they will have greater statistical power to identify recessive genetic variants-associated risk and further characterize the underlying genetic architecture of specific NHL subtypes.

DECLARATIONS

Disclaimer: This project was funded by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH and 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. Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

DECLARATIONS

Acknowledgments

ATBC – The ATBC Study is supported by the Intramural Research Program of the U.S. National Cancer Institute, National Institutes of Health, *Department of Health and Human Services*.

BCCA – 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 - Blood Cancer UK, United Kingdom.

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.

The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by the Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Centre (DKFZ), Federal Ministry of Education and Research (BMBF), Deutsche Krebshilfe, Deutsches Krebsforschungszentrum and Federal Ministry of Education and Research (Germany); the Hellenic Health Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); ERC-2009-AdG 232997 and Nordforsk, Nordic Centre of Excellence programme on Food, Nutrition and Health (Norway); Health Research Fund (FIS), PI131/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia (no. 6236) and Navarra, ISCIII RETIC ((RD06/0020) (Spain); Swedish Cancer Society, Swedish Research Council and Country Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14136 to K.T. Khaw, N.J. Wareham; C570/A16491 to RCT and C8221/A19170 to T. Key (EPIC-Oxford), Medical Research Council (10001433 to K.T. Khaw, N.J. Wareham, MR/M012190/1 to T. Key (EPIC-Oxford)) (United Kingdom)

EpiLymph – European Commission (grant references QLK4-CT-2000-00422 and FOOD-CT-2006-023103); We thank CERCA programme / Generalitat de Catalunya for institutional support.

This work was supported by Spanish Ministry of Economy and Competitiveness - Carlos III Institute of Health cofunded by FEDER funds/European Regional Development Fund (ERDF) - a way to build Europe (grant references CIBERESP, PI17/01280, PI20/00288 with the support of the Secretariat for Universities and Research of the Ministry of Business and Knowledge of the Government of Catalonia (2017SGR1085) 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.

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 (Walter C. Willet) – The HPFS was supported in part by National Institutes of Health grants U01 CA167552, R01 CA149445, and R01 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. The study protocol was approved by the institutional review boards of the Brigham and Women’s Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required.

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 and CA200703); 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, 396414 and 1074383* and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), 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 (Meir J. Stampfer) – The NHS was supported in part by National Institutes of Health grants CA186107, CA87969, CA49449, CA149445, CA098122 and CA134958. 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. 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/Sheffield - 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. Partial support for data collection in Sheffield, UK was made possible by funds from Yorkshire Cancer Research and the Sheffield Experimental Cancer Medicine Centre. We thank the NCRI Haemato-oncology Clinical Studies Group, colleagues in the North Trent Cancer Network the North Trent Haemato-oncology Database.

UTAH-MM - National Institutes of Health CA134674, the Leukemia Lymphoma Society 6067-09, and pilot funds from the Huntsman Cancer Institute (HCI). 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 HCI Comprehensive Cancer Center Support grant, P30 CA42014 (M Beckerle). 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.

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

Authors' contributions

Oversaw or contributed to the study design and analysis: Moore A, Machiela MJ, Machado M, Wang SS, Kane E, Slager SL, Carrington MN, Lan Q, Rothman N, Cerhan JR, Dean M, Camp NJ, Yeager M, Berndt SI

Performed or assisted in analyses or writing the initial draft of the manuscript: Moore A, Machiela MJ, Machado M, Wang SS, Kane E, Slager SL, Zhou W, Birmann BM, Milne RL, Cerhan JR, Dean M, Camp NJ, Yeager M, Berndt SI

Collected/contributed study data or samples: Wang SS, Kane E, Slager SL, Lan Q, Adami HO, Albanes D, Arslan AA, Becker N, Benavente Y, Birmann BM, Bisanzi S, Boffetta P, Bracci PM, Brennan P, Brooks-Wilson AR, Canzian F, Caporaso NE, Clavel J, Cocco P, Conde L, Cox DG, Cozen W, Curtin K, De Vivo I, de Sanjose S, Foretova L, Gapstur SM, Ghesquières H, Giles GG, Glenn M, Glimelius B, Gao C, Habermann TM, Hjalgrim H, Jackson RD, Liebow M, Link BK, Maynadie M, McKay J, Melbye M, Miligi L, Milne RL, Molina TJ, Monnereau A, Nieters A, North KE, Offit K, Patel, AV, Piro S, Ravichandran V, Riboli E, Salles G, Severson RK, Skibola CF, Smedby KE, Southey MC, Spinelli JJ, Staines A, Stewart C, Teras LR, Tinker LF, Travis RC, Vajdic CM, Vermeulen RCH, Joseph V, Weiderpass E, Weinstein S, Wong Doo N, Zhang Y, Zheng T, Chanock SJ, Rothman N, Cerhan JR, Camp NJ, Berndt SI

Read, revised and/or provided feedback on the manuscript: Moore A, Machiela, MJ, Machado M, Wang SS, Kane E, Slager SL, Zhou W, Carrington M, Lan Q, Milne RL, Birmann BM, Adami HO, Albanes D, Arslan, AA, Becker N, Benavente Y , Bisanzi S, Boffetta P, Bracci PM,

Brennan P, Brooks-Wilson AR, Canzian F, Caporaso N, Clavel J, Cocco P, Conde L, Cox DG, Cozen W, Curtin K, De Vivo I, de Sanjose S, Foretova L, Gapstur SM, Ghesquières H, Giles GG, Glenn M, Glimelius B, Gao C, Habermann TM, Hjalgrim H, Jackson RD, Liebow M, Link BK, Maynadie M, McKay J, Melbye M, Miligi L, Molina TJ, Monnereau A, Nieters A, North KE, Offit K, Patel AV, Piro S, Ravichandran V, Riboli E, Salles G, Severson RK, Skibola CF, Smedby KE, Southey MC, Spinelli JJ, Staines A, Stewart C, Teras LR, Tinker LF, Travis RC, Vajdic CM, Vermeulen RCH, Joseph V, Weiderpass E, Weinstein S, Wong Doo N, Zhang Y, Zheng T, Chanock SJ, Rothman N, Cerhan JR, Dean M, Camp NJ, Yeager M, Berndt SI

Availability of data: NCI NHL GWAS data used in this manuscript has been deposited in the dbGaP repository with accession code phs000801.v2.p1. Other relevant data used in this study are available upon request.

Financial support and sponsorship: This project was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH and in part by the Intramural Research Program of the NIH, Frederick National Lab, Center for Cancer Research. It has also been funded in whole or in part with federal funds from the Frederick National Laboratory for Cancer Research, under Contract No. HHSN261200800001E.

Conflicts of interest

All authors declare that there are no conflicts of interest.

Disclaimer: This project was funded by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH and 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. Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in t

his article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Ethical approval and consent to participate

All studies obtained informed consent from participants and were approved by their 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; Mayo GEC: Mayo Clinic Institutional Review Board, #07-005788-03; 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; UCSF1/UCSF2: University of California San Francisco Committee on Human Research; UTAH: University of Utah; WHI: Fred Hutchinson Cancer Research Center; Yale: Human Investigation Committee, Yale University School of Medicine.

Consent for publication

Not applicable.

Copyright

REFERENCES

1. Chatterjee N, Hartge P, Cerhan JR, et al. Risk of non-Hodgkin's lymphoma and family history of lymphatic, hematologic, and other cancers. *Cancer Epidemiol Biomarkers Prev* 2004;13:1415-21..
2. Chang ET, Smedby KE, Hjalgrim H, et al. Family history of hematopoietic malignancy and risk of lymphoma. *J Natl Cancer Inst* 2005;97:1466-74.
3. Wang SS, Slager SL, Brennan P, et al. Family history of hematopoietic malignancies and risk of non-Hodgkin lymphoma (NHL): a pooled analysis of 10 211 cases and 11 905 controls from the International Lymphoma Epidemiology Consortium (InterLymph). *Blood* 2007;109:3479-88.
4. Goldin LR, Björkholm M, Kristinsson SY, Turesson I, Landgren O. Highly increased familial risks for specific lymphoma subtypes. *Br J Haematol* 2009;146:91-4.
5. Altieri A, Bermejo JL, Hemminki K. Familial risk for non-Hodgkin lymphoma and other lymphoproliferative malignancies by histopathologic subtype: the Swedish Family-Cancer Database. *Blood* 2005;106:668-72.
6. Berndt SI, Camp NJ, Skibola CF, et al. Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat Commun* 2016;7:10933.
7. Berndt SI, Skibola CF, Joseph V, et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat Genet* 2013;45:868-76.
8. Cerhan JR, Berndt SI, Vijai J, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nat Genet* 2014;46:1233-8.
9. Law PJ, Berndt SI, Speedy HE, et al. Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. *Nat Commun* 2017;8:14175.
10. 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* 2014;95:462-71.
11. 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* 2015;6:5751.

12. Lettre G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. *Genet Epidemiol* 2007;31:358-62.
13. Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Hum Mol Genet* 2006;15:789-95.
14. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. *Am J Hum Genet* 2012;91:275-92.
15. Keller MC, Visscher PM, Goddard ME. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 2011;189:237-49.
16. Szpiech ZA, Xu J, Pemberton TJ, et al. Long runs of homozygosity are enriched for deleterious variation. *Am J Hum Genet* 2013;93:90-102.
17. Enciso-Mora V, Hosking FJ, Houlston RS. Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. *Eur J Hum Genet* 2010;18:909-14.
18. Spain SL, Cazier JB, Houlston R, Carvajal-Carmona L, Tomlinson I; CORGI Consortium,. Colorectal cancer risk is not associated with increased levels of homozygosity in a population from the United Kingdom. *Cancer Res* 2009;69:7422-9.
19. Hosking FJ, Papaemmanuil E, Sheridan E, et al. Genome-wide homozygosity signatures and childhood acute lymphoblastic leukemia risk. *Blood* 2010;115:4472-7.
20. Thomsen H, Inacio da Silva Filho M, Fuchs M, et al. Evidence of Inbreeding in Hodgkin Lymphoma. *PLoS One* 2016;11:e0154259.
21. Sud A, Cooke R, Swerdlow AJ, Houlston RS. Genome-wide homozygosity signature and risk of Hodgkin lymphoma. *Sci Rep* 2015;5:14315.
22. Joshi PK, Esko T, Mattsson H, et al. Directional dominance on stature and cognition in diverse human populations. *Nature* 2015;523:459-62.
23. Conde L, Halperin E, Akers NK, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat Genet* 2010;42:661-4.
24. De Vivo I, Prescott J, Setiawan VW, et al; Australian National Endometrial Cancer Study Group. Genome-wide association study of endometrial cancer in E2C2. *Hum Genet* 2014;133:211-24.

25. Smedby KE, Foo JN, Skibola CF, et al. GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. *PLoS Genet* 2011;7:e1001378.
26. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015;4:7.
27. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75.
28. Gazal S, Sahbatou M, Perdry H, Letort S, Génin E, Leutenegger AL. Inbreeding coefficient estimation with dense SNP data: comparison of strategies and application to HapMap III. *Hum Hered* 2014;77:49-62.
29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76-82.
30. Yengo L, Zhu Z, Wray NR, et al. Detection and quantification of inbreeding depression for complex traits from SNP data. *Proc Natl Acad Sci U S A* 2017;114:8602-7.
31. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190-1.
32. Edelmann J, Holzmann K, Miller F, et al. High-resolution genomic profiling of chronic lymphocytic leukemia reveals new recurrent genomic alterations. *Blood* 2012;120:4783-94.
33. Puiggros A, Blanco G, Espinet B. Genetic abnormalities in chronic lymphocytic leukemia: where we are and where we go. *Biomed Res Int* 2014;2014:435983.
34. Goldin LR, Björkholm M, Kristinsson SY, Turesson I, Landgren O. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. *Haematologica* 2009;94:647-53.
35. Malek SN. The biology and clinical significance of acquired genomic copy number aberrations and recurrent gene mutations in chronic lymphocytic leukemia. *Oncogene* 2013;32:2805-17.
36. Rawstron AC, Bennett FL, O'Connor SJ, et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med* 2008;359:575-83.
37. Valentino T, Palmieri D, Vitiello M, Pierantoni GM, Fusco A, Fedele M. PATZ1 interacts with p53 and regulates expression of p53-target genes enhancing apoptosis or cell survival based on the cellular context. *Cell Death Dis* 2013;4:e963.

38. Chen Y, Wang J, Wang X, et al. Pik3ip1 Is a Negative Immune Regulator that Inhibits Antitumor T-Cell Immunity. *Clin Cancer Res* 2019;25:6180-94.
39. Paulucci BP, Pereira J, Picciarelli P, Levy D, di Francesco RC. Expression of CysLTR1 and 2 in Maturing Lymphocytes of Hyperplastic Tonsils Compared to Peripheral Cells in Children. *Inflammation* 2016;39:1216-24.
40. Linet MS, Vajdic CM, Morton LM, et al. Medical history, lifestyle, family history, and occupational risk factors for follicular lymphoma: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. *J Natl Cancer Inst Monogr* 2014;2014:26-40.
41. Wang SS, Carrington M, Berndt SI, et al. HLA Class I and II Diversity Contributes to the Etiologic Heterogeneity of Non-Hodgkin Lymphoma Subtypes. *Cancer Res* 2018;78:4086-96.
42. Cerhan JR, Krickler A, Paltiel O, et al. Medical history, lifestyle, family history, and occupational risk factors for diffuse large B-cell lymphoma: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. *J Natl Cancer Inst Monogr* 2014;2014:15-25.
43. Howrigan DP, Simonson MA, Keller MC. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* 2011;12:460.
44. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503-11.
45. Monti S, Savage KJ, Kutok JL, et al. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* 2005;105:1851-61.
46. Bracci PM, Benavente Y, Turner JJ, et al. Medical history, lifestyle, family history, and occupational risk factors for marginal zone lymphoma: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. *J Natl Cancer Inst Monogr* 2014;2014:52-65.
47. Swerdlow SH, Campo E, Harris NL, et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. IARC Press 2008.
48. McQuillan R, Leutenegger AL, Abdel-Rahman R, et al. Runs of homozygosity in European populations. *Am J Hum Genet* 2008;83:359-72.

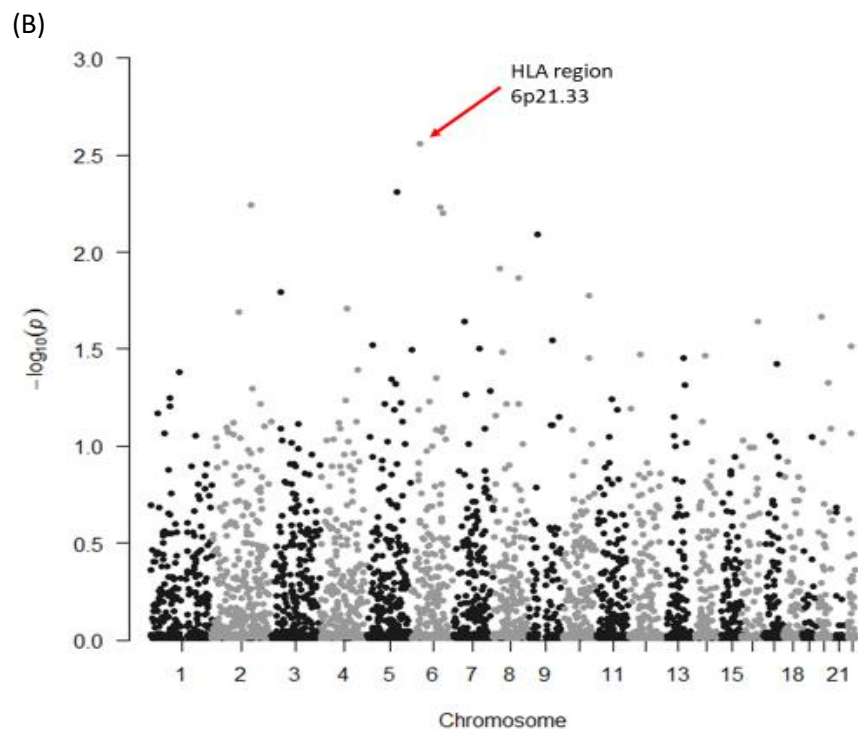
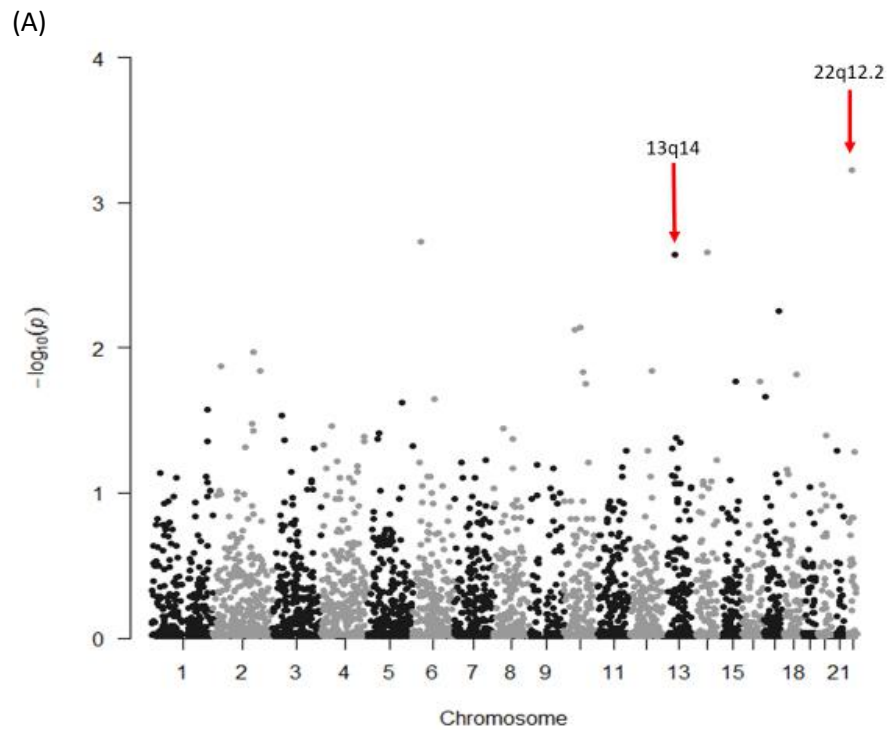


Figure 1. Manhattan plots of $\log_{10}(P)$ values for the association between runs of homozygosity (ROH) and the risk of CLL (A) and FL (B). Here, ROH were divided into 500-kb bins across each chromosome, and each bin was tested for its association with CLL or FL. No bins reached statistical significance after correction for multiple testing.

Table 1. Summary of genome-wide homozygosity measures by GWAS and case-control status *.

Subtype	Study	No. Cases	No. Controls	Median ROH		Median Number of ROH	IQR Number of ROH	Median FROH	IQR FROH	Median F3	IQR F3
				Total Length (KB)	IQR Length (KB)						
CLL	NCI	Cases	2140	18593	13057-25486	7	5-9	0.62%	0.44-0.79%	2.57x10 ⁻³	-1.34x10 ⁻³ -7.49x10 ⁻³
		Controls	6105	18045	12674-24396	7	5-9	0.60%	0.42-0.81%	1.48x10 ⁻³	-2.07x10 ⁻³ -5.36x10 ⁻³
	GEC	Cases	387	23014	16496-29054	8	6-10	0.77%	0.55-0.97%	6.41x10 ⁻⁴	-3.66x10 ⁻³ -4.85x10 ⁻³
		Controls	294	21920	14852-28391	8	6-10	0.73%	0.50-0.95%	-6.99x10 ⁻⁴	-4.90x10 ⁻³ -3.06x10 ⁻³
	UCSF2	Cases	213	12993	7893-19530	4	3-6	0.43%	0.26-0.65%	2.73x10 ⁻³	-0.63x10 ⁻³ -8.83x10 ⁻³
		Controls	746	12173	7523-17111	4	3-6	0.41%	0.25-0.57%	1.51x10 ⁻³	-2.25x10 ⁻³ -5.72x10 ⁻³
	Utah	Cases	321	18806	13535-24431	7	5-9	0.63%	0.45-0.81%	3.69x10 ⁻⁴	-2.79x10 ⁻³ -4.71x10 ⁻³
		Controls	403	17452	12288-22866	6	5-8	0.58%	0.41-0.76%	9.47x10 ⁻⁴	-2.09x10 ⁻³ -3.89x10 ⁻³
DLBCL	NCI	Cases	2621	18046	12671-24990	7	5-9	0.60%	0.42-0.83%	2.11x10 ⁻³	-1.48x10 ⁻³ -6.54x10 ⁻³
		Controls	6105	18199	12753-24519	7	5-9	0.61%	0.43-0.82%	1.46x10 ⁻³	-2.05x10 ⁻³ -5.39x10 ⁻³
	Mayo	Cases	393	16586	12167-21774	6	5-8	0.55%	0.41-0.73%	1.67x10 ⁻⁴	-3.03x10 ⁻³ -3.07x10 ⁻³
		Controls	172	16451	11368-23319	6	4-8	0.55%	0.38-0.78%	2.19x10 ⁻⁴	-2.55x10 ⁻³ -2.94x10 ⁻³
	UCSF2	Cases	253	11535	5957-19153	4	3-6	0.38%	0.20-0.64%	1.69x10 ⁻³	-1.62x10 ⁻³ -6.60x10 ⁻³
		Controls	745	12182	7554-17158	4	3-6	0.41%	0.25-0.57%	1.48x10 ⁻³	-2.20x10 ⁻³ -5.82x10 ⁻³
	GELA/EPIC	Cases	547	15526	10391-21466	6	4-7	0.52%	0.35-0.72%	1.27x10 ⁻³	-2.27x10 ⁻³ -5.00x10 ⁻³
		Controls	525	17327	11782-23437	6	4-8	0.58%	0.39-0.78%	2.56x10 ⁻³	-0.87x10 ⁻³ -6.14x10 ⁻³
FL	NCI	Cases	2085	17838	12387-24319	7	5-9	0.59%	0.41-0.81%	1.41x10 ⁻³	-2.14x10 ⁻³ -6.15x10 ⁻³
		Controls	6105	18110	12753-24512	7	5-9	0.60%	0.43-0.82%	1.49x10 ⁻³	-2.06x10 ⁻³ -5.41x10 ⁻³
	UCSF1/NHS	Cases	119	17058	12645-22150	7	5-8	0.57%	0.42-0.74%	8.51x10 ⁻⁵	-3.57x10 ⁻³ -3.05x10 ⁻³
		Controls	340	17144	12100-24513	6.5	5-9	0.57%	0.40-0.82%	6.35x10 ⁻⁴	-3.05x10 ⁻³ -4.29x10 ⁻³
	UCSF2	Cases	209	13621	8911-19655	5	3-7	0.45%	0.30-0.66%	3.02x10 ⁻³	-1.34x10 ⁻³ -7.58x10 ⁻³
		Controls	745	12071	7517-17115	4	3-6	0.40%	0.25-0.57%	1.50x10 ⁻³	-2.16x10 ⁻³ -5.76x10 ⁻³
	SCALE	Cases	371	12582	7643-19776	5	3-7	0.42%	0.25-0.66%	3.02x10 ⁻³	-0.50x10 ⁻³ -7.52x10 ⁻³
		Controls	790	12132	7506-17654	5	3-6	0.40%	0.25-0.59%	1.22x10 ⁻³	-1.60x10 ⁻³ -5.00x10 ⁻³
MZL	NCI	Cases	808	18097	12831-24555	7	5-9	0.60%	0.43-0.82%	1.83x10 ⁻³	-1.54x10 ⁻³ -7.30x10 ⁻³
		Controls	6102	18268	12811-24623	7	5-9	0.61%	0.43-0.82%	1.46x10 ⁻³	-2.01x10 ⁻³ -5.38x10 ⁻³

* Median and interquartile range (IQR) are provided for runs of homozygosity (ROH), fraction of runs of homozygosity (FROH), and correlation between uniting gametes (F3).

Table 2. Risk of NHL subtypes associated with measures of genome-wide homozygosity, FROH and F3*.

FROH					
Subtype	β	SE	p-value	I²	P_{het}
CLL	21.14	4.41	1.59x10 ⁻⁶	0.0%	0.42
DLBCL	0.04	10.89	1.0	72.5%	0.01
FL	11.39	5.82	0.02	5.3%	0.37
MZL	-0.87	7.88	0.91		
F3					
Subtype	β	SE	p-value	I²	P_{het}
CLL	27.46	6.51	2.44x10 ⁻⁵	49.7%	0.11
DLBCL	1.96	9.57	0.84	82.4%	0.001
FL	13.19	8.01	0.10	64.2%	0.04
MZL	6.40	5.20	0.22		

* Estimates of the log odds (β), standard error (SE), and p-value are provided for the association between FROH and F3 and each subtype, adjusted for age, sex (except UCSF1/NHS), percentage of missing SNPs, and principal components and combined using random effects meta-analysis. The I² statistic provides an estimate of heterogeneity in association estimates across GWAS, and P_{het} is the p-value for heterogeneity among studies.

Table S1. Sample size and genotype arrays used for each NHL GWAS.

Study	No. of Subjects in Analysis				Controls	Genotyping Platform	No. of Autosomal SNPs
	CLL	DLBCL	FL	MZL			
NCI*	2,140	2,621	2,085	808	6,105	OmniExpress/Omni2.5	554,315
GELA/EPIC		547			525	HumanHap 610K/660W	486,277
GEC	387				294	Affymetrix 6.0	604,007
Mayo		393			172	HumanHap 660W	492,020
SCALE			371		790	HumanHap 317K	289,632
UCSF2*	213	253	209		745	HumanCNV370-Duo	289,900
UCSF1/NHS			119		340	OmniExpress	554,430
Utah	321				403	HumanHap 610K	483,331
Total	3061	3814	2784	808	9374		

* The same set of controls was used for the different NHL subtypes in the study; however, quality control metrics were applied separately to each set of cases and controls, resulting in small differences in the number of controls and SNPs in the analysis.

Table S2. Characteristics of cases and controls included in each NHL subtype analysis.

Study	No. of Subjects		% Female		Mean Age (\pm SD)	
	Cases	Controls	Cases	Controls	Cases	Controls
CLL						
NCI NHL	2,140	6,105	44.4%	27.3%	65.6 (\pm 9.7)	66.0 (\pm 10.4)
USCF2	213	746	36.2%	42.4%	63.9 (\pm 11.2)	61.5 (\pm 13.0)
GEC	387	294	34.9%	36.7%	61.3 (\pm 11.1)	62.5 (\pm 11.2)
Utah	321	403	40.2%	43.9%	63.0 (\pm 10.3)	64.1 (\pm 10.8)
FL						
NCI NHL	2,085	6,105	54.2%	27.3%	60.3 (\pm 12.1)	66.0 (\pm 10.4)
USCF2	209	745	47.4%	42.4%	60.9 (\pm 11.9)	61.5 (\pm 13.0)
SCALE	371	790	50.4%	68.2%	57.0 (\pm 9.5)	50.8 (\pm 11.6)
UCSF1/NHS	119	340	48.7%	100.0%	58.8 (\pm 12.9)	62.6 (\pm 8.6)
DLBCL						
NCI NHL	2,621	6,105	51.1%	27.3%	60.7 (\pm 14.0)	66.0 (\pm 10.4)
USCF2	253	745	44.3%	42.4%	62.4 (\pm 13.9)	61.5 (\pm 13.0)
Mayo	393	172	47.3%	61.6%	61.7 (\pm 14.4)	51.1 (\pm 13.8)
GELA/EPIC	547	525	43.0%	55.1%	58.7 (\pm 16.7)	58.5 (\pm 9.5)
MZL						
NCI NHL	808	6,102	59.7%	27.3%	64.0 (\pm 12.1)	66.0 (\pm 10.4)

Table S3. Individual GWAS and combined association results for the risk of each NHL subtype with FROH and F3.

	Study	FROH					F3				
		β	SE	p-value	I ²	p _{het}	β	SE	p-value	I ²	p _{het}
CLL	NCI	18.85	4.73	6.73x10 ⁻⁵			27.45	3.47	2.82x10 ⁻¹⁵		
	GEC	50.61	19.64	0.01			33.84	10.61	0.001		
	UCSF2	20.81	21.81	0.34			43.03	12.16	4.01x10 ⁻⁴		
	Utah	33.78	21.83	0.12			-0.41	13.99	1.0		
	<i>Combined</i>	21.1	4.41	1.59x10 ⁻⁶	0.0%	0.42	27.5	6.51	2.44x10 ⁻⁵	49.7%	0.11
DLBCL	NCI	14.17	4.78	0.003			14.12	3.35	2.55x10 ⁻⁵		
	Mayo	7.96	35.21	0.82			-6.15	19.04	0.75		
	UCSF2	-3.98	20.47	0.85			12.41	11.83	0.29		
	GELA/EPIC	-15.71	7.80	0.04			-15.53	6.53	0.02		
	<i>Combined</i>	0.0	10.89	1.0	72.5%	0.01	2.0	9.57	0.84	82.4%	0.001
FL	NCI	6.72	5.12	0.19			5.42	3.62	0.13		
	UCSF1/NHS	5.15	34.26	0.88			-18.04	22.07	0.41		
	UCSF2	35.13	18.54	0.06			21.75	11.68	0.06		
	SCALE	20.05	10.45	0.06			28.64	9.04	0.002		
	<i>Combined</i>	11.4	5.82	0.02	5.3%	0.37	13.2	8.01	0.10	64.2%	0.04
MZL	NCI	-0.87	7.88	0.91			6.4	5.2	0.22		

* Estimates of the log odds (β), standard error (SE), and p-value are provided for the association between FROH and F3 and each subtype, adjusted for age, sex (except UCSF1/NHS), percentage of missing SNPs, and principal components and combined using random effects meta-analysis. The I² statistic provides an estimate of heterogeneity in association estimates across GWAS, and P_{het} is the p-value for heterogeneity among studies.

Table S4. Sensitivity analysis of the risk of each NHL subtype associated with FROH and F3 after removing individual studies from the meta-analysis*.

	Study Removed [†]	FROH			F3		
		β	SE	p-value	β	SE	p-value
CLL	NCI	36.19	12.13	0.003	26.58	12.14	0.03
	GEC	19.57	4.52	1.49x10 ⁻⁵	25.08	9.32	0.007
	UCSF2	26.29	8.98	0.003	23.83	7.42	0.001
	Utah	23.25	7.62	0.002	29.09	3.19	6.25x10 ⁻²⁰
DLBCL	NCI	-13.31	7.13	0.06	-4.68	9.75	0.63
	Mayo	-0.78	12.08	0.95	3.36	10.89	0.76
	UCSF2	0.83	13.11	0.95	-1.37	12.30	0.91
	GELA/EPIC	13.14	4.62	0.004	13.43	3.18	2.34x10 ⁻⁵
FL	NCI	22.46	8.80	0.01	18.22	10.37	0.08
	UCSF1/NHS	14.14	7.04	0.04	16.67	8.41	0.05
	UCSF2	9.23	4.56	0.04	10.35	10.30	0.32
	SCALE	10.12	6.72	0.13	7.38	6.97	0.29

* Estimates of the log odds (β), standard error (SE), and p-value are provided for the association between FROH and F3 and each subtype, adjusted for age, sex (except UCSF1/NHS), percentage of missing SNPs, and principal components and combined using random effects meta-analysis. [†] This column indicates the specific study removed from the meta-analysis. The estimates presented are from the random effects meta-analysis after excluding this study.