Article:

Din, Lennox, Sheikh, Mohammad, Kosaraju, Nikitha et al. (89 more authors) (2019) Genetic overlap between autoimmune diseases and non-Hodgkin lymphoma subtypes. Genetic Edpidemiology. pp. 844-863. ISSN: 0741-0395

https://doi.org/10.1002/gepi.22242

Author Query Form

Journal: GEPI
Article: GEPI22242

Dear Author,

During the copy-editing of your paper, the following queries arose.

Please refer to the query reference callout numbers in the page proofs and respond to each by marking the necessary comments using the PDF annotation tools.

Please remember illegible or unclear comments and corrections may delay publication.

Many thanks for your assistance.

Query Reference	Query	Remarks/Comments	Query Reviewed
Q1	Author: Please confirm that given names (blue) and surnames/family names (vermilion) have been identified correctly.		☐
Q2	Author: Funder details have been incorporated in the funder table using information provided in the article text. Please check that the funder information in the table is correct.		☐
Q3	Author: Please verify that the linked ORCID identifiers are correct for each author.		☐
Q4	Author: Division or department names are required for affiliations 1, 4, 7, 8, 10, 12, 14, 15, 16, 17, 19, 20, 21, 22, 24, 29, 30, 31, 32, 35, 37, 42, 45, 48, 50, 55, 56, 57, 59, 60, 66, 67, 70, 71, 74, 80, 84, and 85.		☐
Q5	Author: Please check and validate all the given affiliations.		☐
Q6	Author: Please provide the publisher's location for "Thun et al., (2017)."		☐

Funding Info Query Form

Please confirm that the funding sponsor list below was correctly extracted from your article: that it includes all funders and that the text has been matched to the correct FundRef Registry organization names. If a name was not found in the FundRef registry, it may not be the canonical name form, it may be a program name rather than an organization name, or it may be an organization not yet included in FundRef Registry. If you know of another name form or a parent organization name for a "not found" item on this list below, please share that information.

FundRef name	**FundRef Organization Name (Country)**
Partial support for all datasets within the UPDB is provided by the Huntsman Cancer Institute (HCI) and the HCI Comprehensive Cancer Center Support grant	

Pierre-Antoine Gourraud is supported by the ATIP-Avenir INSERM program and the Region Pays de Loire ConnecTalent ARSEP Foundation (France) and the Nantes University Foundation.	
Hans-Olov Adami is supported by the Karolinska Institutet Distinguished Professor Award	
Karin E Smedby is supported by the Swedish Cancer Society the Strategic Research Program in Epidemiology at Karolinska Institute and National Institutes of Health	
Nicola Camp reports support from	
John J Spinelli is supported by the Canadian Institutes for Health Research (CIHR); the Canadian Cancer Society; and the Michael Smith Foundation for Health Research	
The UCR is supported in part by NIH contract	
James R Cerhan is supported by NIH grants	
Lenka Foretova is supported by a grant MH CZ - DRO	
David Conti is funded by the following grants	
Karolinska Institutet Distinguished Professor	
NIH	
Huntsman Cancer Institute (HCI) and the HCI Comprehensive Cancer Center	
National Cancer Institute SEER Program	
NIH	
MH CZ - DRO	
Swedish Cancer Society	
Strategic Research Program in Epidemiology at Karolinska Institute and National Institutes of Health	
European Commission	European Commission
Spanish Ministry of Health	
Marató de TV3 Foundation	
Agència de Gestiód'AjutsUniversitarisi de Recerca – Generalitat de Catalunya	
Federal Office for Radiation Protection	
José Carreras Leukemia Foundation	
German Federal Ministry for Education and Research	
National Institutes of Health	Foundation for the National Institutes of Health
National Center for Advancing Translational Science	

Wellcome Trust	
US National Multiple Sclerosis Society	
Neuropromise EU	
Italian Foundation for Multiple Sclerosis	
Association of British Neurologists Spanish Ministry of Health	
HYPERGENES	
Molecular Epidemiology of Non-Hodgkin Lymphoma Survival	
National Cancer Institute	Institut National Du Cancer
Italian Ministry for Education University and Research	
Italian Association for Cancer Research	
Regional Administration of Sardinia	
National Center for Advancing Translational Science	
Australian National Health and Medical Research Council	
Instituto de Salud Carlos III	
Agencia de Gestio d'Ajuts Universitaris i de Recerca (AGAUR) – Generalitat de Catalunya	
Lymphoma Foundation	Lymphoma Foundation
Lymphoma Research Fund	
Robert and Kate Niehaus Clinical Cancer Genetics Research Initiative	
National Natural Science Foundation of China	
National Institute of Environmental Health Sciences	National Institute of Environmental Health Sciences
NIH/NHLBI	
GENEVA Coordinating Center	
Stockholm Country Council	
Lundbeck Foundation Grant	
Danish Cancer Society Grant	
National Heart Lung and Blood Institute National Institutes of Health U.S. Department of Health and Human Services	

Journal	MSP No.	Dispatch: June 27, 2019	CE:
GEPI	22242	No. of Pages: 21	PE: Srinivas Y. V.

Received: 21 January 2019 | Revised: 1 May 2019 | Accepted: 7 May 2019

DOI: 10.1002/gepi.22242

Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

WILEY

RESEARCH ARTICLE

Genetic overlap between autoimmune diseases and non-Hodgkin lymphoma subtypes

Lennox Din[1] | Mohammad Sheikh[1] | Nikitha Kosaraju[2] | Pouya Khankhanian[2]* | Karin Ekstrom Smedby[3] | Sasha Bernatsky[4,5] | Sonja Berndt[6] | Christine F. Skibola[7] | Alexandra Nieters[8] | Sophia Wang[9] | James D. McKay[10] | Pierluigi Cocco[11] | Marc Maynadié[12] | Lenka Foretová[13] | Anthony Staines[14] | Thomas M. Mack[15] | Silvia deSanjosé[16] | Timothy J. Vyse[17] | Leonid Padyukov[18] | Zachary Taub[19] | Morris Din[20] | Alain Monnereau[21,22] | Alain A. Arslan[23] | Amy Moore[6] | Angela R. Brooks-Wilson[24,25] | Anne J. Novak[26] | Bengt Glimelius[27] | Brenda M. Birmann[28] | Brian K. Link[29] | Carolyn Stewart[30] | Claire M. Vajdic[31] | Corinne Haioun[32] | Corrado Magnani[33] | David Conti[34] | David G. Cox[35] | Delphine Casabonne[36,37] | Demetrius Albanes[6] | Eleanor Kane[38] | Eve Roman[38] | Giacomo Muzi[39] | Gilles Salles[40] | Graham G. Giles[41,42] | Hans-Olov Adami[43] | Hervé Ghesquières[40] | Immaculata DeVivo[44,45] | Jacqueline Clavel[46] | James R. Cerhan[47] | John J. Spinelli[48] | Jonathan Hofmann[49] | Joseph Vijai[30] | Karen Curtin[50] | Karen H. Costenbader[51] | Kenan Onel[52] | Kenneth Offit[53,54] | Lauren R. Teras[55] | Lindsay Morton[6] | Lucia Conde[56] | Lucia Miligi[57] | Mads Melbye[58,59] | Maria Grazia Ennas[60] | Mark Liebow[61] | Mark P. Purdue[62] | Martha Glenn[50] | Melissa C. Southey[63,64] | Nathaniel Rothman[6] | Nicola J. Camp[50,65] | Nicole Wong Doo[66,67] | Nikolaus Becker[68] | Nisha Pradhan[30] | Paige M. Bracci[69] | Paolo Boffetta[70] | Paolo Vineis[71] | Paul Brennan[10] | Peter Kraft[72] | Qing Lan[6] | Richard K. Severson[73] | Roel C. H. Vermeulen[74] | Roger L. Milne[42,67] | Rudolph Kaaks[75] | Ruth C. Travis[76] | Stephanie J. Weinstein[77] | Stephen J. Chanock[78] | Stephen M. Ansell[61] | Susan L. Slager[79] | Tongzhang Zheng[80] | Yawei Zhang[81] | Yolanda Benavente[37,82] | Lohith Madireddy[83] | Pierre-Antoine Gourraud[84,85] | Jorge R. Oksenberg[83] | Wendy Cozen[86] | Henrik Hjalgrim[87]*

[1]College of Medicine, California Northstate University, Elk Grove, California

ABSTRACT

Epidemiologic studies show an increased risk of non-Hodgkin lymphoma (NHL) in patients with autoimmune disease (AD), due to a combination of shared

* Henrik Hjalgrim and Pouya Khankhanian contributed equally to this work.

[2]Department of Neurology, Hospital of the University of Pennsylvania, Philadelphia, Pennsylvania

[3]Division of Clinical Epidemiology, Department of Medicine, Karolinska University Hospital, Karolinska Institutet, Solna, Sweden

[4]Faculty of Medicine, McGill University, Montreal, Canada

[5]Clinical Epidemiology, McGill University Health Centre, Research Institute, Montreal, Canada

[6]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

[7]School of Medicine, Emory University, Atlanta, Georgia

[8]Centre of Chronic Immunodeficiency, University Medical Centre Freiburg, Freiburg im Breisgau, Germany

[9]Department of Population Sciences, City of Hope and the Beckman Research Institute, Duarte, California

[10]International Agency for Research on Cancer, Lyon, France

[11]Department of Medical Sciences and Public Health, University of Caligari, Calgary, Canada

[12]The University of Burgundy-Franche-Comté and CHU Dijon Bourgogne, Registre des Hémopathies Malignes de Côte d'Or, INSERM U1231, Dijon, France

[13]Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic

[14]School of Nursing and Human Science, Dublin City University, Dublin, Ireland

[15]Norris Comprehensive Cancer Center and Hospital, University of Southern California, Los Angeles, California

[16]Institut Catala d' Oncologia, Barcelona, Spain

[17]Medical Education, King's College London, London, England

[18], Department of Medicine, Karolinska Institutet, Stockholm, Sweden

[19]University of California, Davis, California

[20]California State University, Sacramento, Califonia

[21]Center of Research in Epidemiology and Statistics, Sorbonne (CRESS), Epidemiology of Childhood and Adolescent Cancer Group, INSERM, Paris, France

environmental factors and/or genetic factors, or a causative cascade: chronic inflammation/antigen-stimulation in one disease leads to another. Here we assess shared genetic risk in genome-wide-association-studies (GWAS).

Secondary analysis of GWAS of NHL subtypes (chronic lymphocytic leukemia, diffuse large B-cell lymphoma, follicular lymphoma, and marginal zone lymphoma) and ADs (rheumatoid arthritis, systemic lupus erythematosus, and multiple sclerosis). Shared genetic risk was assessed by (a) description of regional genetic of overlap, (b) polygenic risk score (PRS), (c)"diseasome", (d)meta-analysis.

Descriptive analysis revealed few shared genetic factors between each AD and each NHL subtype. The PRS of ADs were not increased in NHL patients (nor vice versa). In the diseasome, NHLs shared more genetic etiology with ADs than solid cancers ($p = .0041$). A meta-analysis (combing AD with NHL) implicated genes of apoptosis and telomere length.

This GWAS-based analysis four NHL subtypes and three ADs revealed few weakly-associated shared loci, explaining little total risk. This suggests common genetic variation, as assessed by GWAS in these sample sizes, may not be the primary explanation for the link between these ADs and NHLs.

KEYWORDS

autoimmune disease, genome-wide association study, meta-analysis, non-Hodgkin lymphoma

[22]Registre des Hémopathies Malignes de la Gironde, Institut Bergonie, Bordeaux, France

[23]The Department of Obstetrics and Gynecology, NYU School of Medicine, New York, New York

[24]Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, Canada

[25]Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, Canada

[26]Department of Medicine, Mayo Clinic, Rochester, Minnesota

[27]Department of Immunology, Genetics, and Pathology, Uppsala Universitet, Uppsala, Sweden

[28]Channing Division of Network Medicine, Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts

[29]Internal Medicine, University of Iowa Hospitals and Clinics, Iowa City, Iowa

[30]Memorial Sloan Kettering Cancer Center, New York, New York

[31]Centre for Big Data Research in Health, University of New South Wales, Kensington, Australia

[32]Lymphoid Malignancies Unit, Henri Mondor Hospital, University Paris-Est Créteil (UPEC), Paris, France

[33]Medicina Traslazionale, Università del Piemonte Orientale, Vercelli, Italy

[34]Department of Preventive Medicine, Division of Biostatistics, University of Southern California, Los Angeles, Califonia

[35]Centre de Recherche en Cancerologie de Lyon, Lyon, France

[36]Unit of Infections and Cancer, Cancer Epidemiology Research Programme, Catalan Institute of Oncology, IDIBELL, Barcelona, Spain

[37]CIBER Epidemiología y Salud Pública, Madrid, Spain

[38]Department of Health Sciences, University of York, York, England

[39]Department of Occupational Medicine, Universita degli Studi di Perugia, Perugia, Italy

[40]Hematology Department, Centre Hospitalier Universitaire de Lyon, Lyon, France

[41]Division of Cancer Epidemiology & Intelligence, Cancer Council Victoria, Melbourne, Australia

[42]Centre for Epidemiology and Biostatistics, University of Melbourne, Melbourne, Australia

[43]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Solna, Sweden

[44]Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts

[45]Harvard Medical School, Brigham Women's Hospital, Boston, Massachusetts

[46]Department of Environmental Epidemiology of Cancers, INSERM, Paris, France

[47]Department of Health Sciences Research and Clinical Epidemiology, Mayo Clinic, Rochester, Minnesota

[48]Population Oncology, British Columbia Cancer Agency, Vancouver, Canada

[49]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland

[50]Huntsman Cancer Institute, University of Utah, Salt Lake, Utah

[51]Division of Medicine, Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Brookline, Massachusetts

[52]Department of Pediatrics, University of Chicago, Chicago, Illinois

[53]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York

[54]Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, New York

[55]Epidemiology Research Program, American Cancer Society, New York, New York

[56]University College London Cancer Institute, London, England

[57]Institute of Oncology Studies and Prevention,

[58]Department of Epidemiology, Statens Serum Institut, København, Denmark

[59]School of Medicine, Stanford University, Stanford, California

[60]School of Biomedical Sciences, University of Calgary, Calgary, Canada

[61]Division of Internal Medicine, Mayo Clinic, Rochester, Minnesota

[62]Division of Cancer Biology, National Cancer Institute, Bethesda, Maryland
[63]Department of Clinical Pathology, Genetic Epidemiology Laboratory, University of Melbourne, Melbourne, Australia
[64]Precision Medicine, School of Clinical Sciences, Monash University, Melbourne, Australia
[65]Department of Internal Medicine and Biomedical Informatics, University of Utah, Salt Lake, Utah
[66]Clinical School, Concord Hospital, University of Sydney, Sydney, Australia
[67]Centre for Epidemiology and Intelligence, Cancer Council Australia, Melbourne, Australia
[68]Division of Cancer Epidemiology, Deutsches Krebsforschungszentrum, Heidelberg, Germany
[69]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California
[70]Icahn School of Medicine at Mount Sinai, Tisch Cancer Institute, New York, New York
[71]Environmental Epidemiology and Public Health, Imperial College London, London, England
[72]Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts
[73]Department of Family Medicine and Public Health Sciences, Wayne State University, Karmanos Cancer Institute, Detroit, Michigan
[74]Institute for Risk Assessment Science, Utrecht University, Utrecht, Netherlands
[75]Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany
[76]Cancer Epidemiology Unit, University of Oxford, Oxford, England
[77]Division of Cancer Epidemiology and Genetics, National Cancer Institute (NIH), Bethesda, Maryland
[78]Division of Cancer Epidemiology and Genetics, National Cancer Research Institute, London, England
[79]Division of Health Sciences Research, Mayo Clinic, Rochester, Minnesota
[80]School of Public Health, Brown University, Providence, Rhode Island

[81]Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut

[82]Unit of Infections and Cancer, Cancer Epidemiology Research Programme, IDIBELL, Institut Catala d' Oncologia, Barcelona, Spain

[83]Laboratories of Neurogenetics, University of California, San Francisco, California

[84]Centre de Recherche en Transplantation et Immunologie, Université de Nantes, Inserm, Nantes, France

[85]Centre Hospitalier Universitaire de Nantes, Institut de Transplantation Urologie Néphrologie (ITUN), Nantes, France

[86]Departments of Preventive Medicine and Pathology, Norris Comprehensive Cancer Center and Hospital, University of Southern California, Los Angeles, California

[87]Division of Epidemiology, Statens Serum Institut, Amager, Copenhagen

Correspondence

Pouya Khankhanian, Department of Neurology, Hospital of University of Pennsylvania, 3400 Spruce Street, 3W Gates Bldg (Neurology), Philadelphia, PA 19104.

Email: Pouya.Khankhanian@uphs.upenn.edu

Funding information

Partial support for all datasets within the UPDB is provided by the Huntsman Cancer Institute (HCI) and the HCI Comprehensive Cancer Center Support grant, Grant/Award Number: P30 CA42014; Pierre-Antoine Gourraud is supported by the ATIP-Avenir INSERM program and the Region Pays de Loire ConnecTalent, ARSEP Foundation (France) and the Nantes University Foundation., Grant/Award Number: na; Hans-Olov Adami is supported by the Karolinska Institutet Distinguished Professor Award, Grant/Award Number: Dnr: 2368/10-221; Karin E Smedby is supported by the Swedish Cancer Society, the Strategic Research Program in Epidemiology at Karolinska Institute and National Institutes of Health, Grant/Award Number: (2009/659, 02 6661) and (5R01 CA69669-02); Nicola Camp reports support from, Grant/Award Number: NIH R01CA134674; John J Spinelli is supported by the Canadian Institutes for

Health Research (CIHR); the Canadian Cancer Society; and the Michael Smith Foundation for Health Research, Grant/Award Number: na; The UCR is supported in part by NIH contract, Grant/Award Number: HHSN261201000026C; James R Cerhan is supported by NIH grants, Grant/Award Number: R01 CA92153, R01 CA200703, P50 CA97274, U01 CA1955; Lenka Foretova is supported by a grant MH CZ - DRO, Grant/Award Number: (MMCI, 00209805); David Conti is funded by the following grants:, Grant/Award Number: P01CA196569, R01CA201407, R01CA140561, ES024844, a; Karolinska Institutet Distinguished Professor Award ; NIH, Grant/Award Numbers: R01CA134674, R01 CA92153, R01 CA200703, P50 CA97274, U01 CA195568, NO1-CO-12400; Huntsman Cancer Institute (HCI) and the HCI Comprehensive Cancer Center Support grant, , Grant/Award Number: P30 CA42014; National Cancer Institute SEER Program, Grant/Award Number: HHSN261201000026C; MH CZ - DRO, Grant/Award Number: 00209805; Swedish Cancer Society, Grant/Award Number: 2009/659, 02 6661; Strategic Research Program in Epidemiology at Karolinska Institute and National Institutes of Health, Grant/Award Number: 5R01 CA69669-02; European Commission, Grant/Award Numbers: QLK4-CT-2000-00422, FOOD-CT-2006-; Spanish Ministry of Health, Grant/Award Numbers: CIBERESP, PI11/01810, RCESP C03/09, RTICESP C03/10, RD06/0020/0095; Marató de TV3 Foundation, Grant/Award Number: 051210; Agència de Gestiód'AjutsUniversitarisi de Recerca – Generalitat de Catalunya, Grant/Award Number: 2014SRG756 and 2009SGR1465; Federal Office for Radiation Protection grants , Grant/Award Number: StSch4261 and StSch4420; José Carreras Leukemia Foundation grant , Grant/Award Number: DJCLS-R12/23; German Federal Ministry for Education and Research, Grant/Award Number: BMBF-01-EO-; National Institutes of Health, Grant/Award Numbers: CA118444, CA148690, CA92153, K08CA134919, AI076544, NS032830, NS049477, NS19142, NS049510, NS26799, NS43559, NS067305, CA104021, RR020092, RR024992 and K23N/S048869, CA97274, R01 CA92153, CA186107, CA87969, CA49449, CA149445, CA098122, and CA134958; National Center for Advancing

Translational Science, Grant/Award Number: UL1 TR000135; Wellcome Trust, Grant/Award Number: 085475/B/08/Z, 085475/Z/08/Z, 075491/Z/04/Z, and 068545/Z/02; US National Multiple Sclerosis Society, Grant/Award Number: RG 4201-A-1; Neuropromise EU grant, Grant/Award Number: LSHM-CT-; Italian Foundation for Multiple Sclerosis, Grant/Award Number: 2002/R/40, 2005/R/10, 2008/R/11, and 2008; Association of British Neurologists, Spanish Ministry of Health, Grant/Award Number: FISPI060117; HYPERGENES, Grant/Award Number: HEALTH-F4-; Molecular Epidemiology of Non-Hodgkin Lymphoma Survival, Grant/Award Number: R01 CA129539; National Cancer Institute, Grant/Award Number: P30 CA086862, P30 CA15083; Italian Ministry for Education, University and Research, Grant/Award Number: PRIN 2007 prot.2007WEJLZB, PRIN 2009 prot. 20092ZELR2; Italian Association for Cancer Research, Grant/Award Number: 11855; Regional Administration of Sardinia, Grant/Award Number: LR7 CRP-59812/2012; Institut National du Cancer, Grant/Award Number: 2008-020; National Center for Advancing Translational Science, Grant/Award Number: UL1 TR000135; National Cancer Institute, Grant/Award Number: P30 CA015083; Australian National Health and Medical Research Council grants , Grant/Award Numbers: 209057, 251553, 504711, 396414, ID990920; Instituto de Salud Carlos III, Grant/Award Number: PI11/01810, PI14/01219, RCESP C03/09,; Agencia de Gestio d'Ajuts Universitaris i de Recerca (AGAUR) – Generalitat de Catalunya, Grant/Award Number: 2014SGR756; Lymphoma Foundation, Grant/Award Number: LF5541; Lymphoma Research Fund, Grant/Award Number: 74419; Robert and Kate Niehaus Clinical Cancer Genetics Research Initiative, Grant/Award Number: 57470; National Natural Science Foundation of China, Grant/Award Number: 61471078; National Cancer Institute, Grant/Award Number: R01 CA098661, P30 CA016087; National Institute of Environmental Health Sciences, Grant/Award Number: ES000260; NIH/NHLBI, Grant/Award Number: R37HL039693; GENEVA Coordinating Center, Grant/Award Number: U01 HG004446; Stockholm Country Council, Grant/Award Number: 20110229; Lundbeck Foundation Grant,

Grant/Award Number: R19-A2364; Danish Cancer Society Grant, Grant/Award Number: DP 08-155; National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts , Grant/Award Number: HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C; National Cancer Institute, Grant/Award Number: CA62006

1 | INTRODUCTION

It is well established that patients with autoimmune diseases (AD) such as rheumatoid arthritis (RA), Sjögren's syndrome, and systemic lupus erythematosus (SLE) are at increased risk of malignant lymphomas, that is, Hodgkin and non-Hodgkin lymphomas (NHL; Baecklund, Smedby, Sutton, Askling, & Rosenquist, 2014; Thun, Linet, Cerhan, Haiman, & Schottenfeld, 2017) (Hemminki, Försti, Sundquist, Sundquist, & Li, 2017). Different mechanisms may plausibly contribute to this association. For instance, an autoimmune reaction may involve chronic antigenic stimulation and inflammation, which may promote lymphoma development through heightened B- or T-cell activation (Baecklund et al., 2014). Increased risks of salivary gland marginal zone lymphomas (MZL) of B-cell origin in patients with Sjögren's syndrome and of small intestinal T-cell lymphomas in patients with celiac disease support such mechanisms (Baecklund et al., 2014). AD treatment might also contribute to the observed increased lymphoma risk, for example, through suppression of the immune system (Baecklund et al., 2014).

Although these mechanisms are intuitively an appealing explanation for the AD-NHL association, the association might also theoretically involve other risk factors shared by the two groups of diseases. In this regard, the current understanding of environmental risk factors possibly shared by ADs and NHLs, such as smoking, offers no convincing explanation for their mutual clustering (Thun et al., 2017; Deane et al., 2010; Park et al., 2009; Belbasis, Bellou, Evangelou, Ioannidis, & Tzoulaki, 2015; Smedby & Ponzoni, 2017; Ekström et al., 2003; Bernatsky et al., 2013). Further, meta-analyses of genome-wide association studies (GWAS) suggested genetic overlap between SLE and diffuse large B-cell lymphoma (DLBCL; Bernatsky et al., 2017), and between multiple sclerosis (MS) and Hodgkin lymphoma (Khankhanian et al., 2016) as a partial explanation of the accumulation of those two diseases among relatives.

Here, we use available GWAS data from three ADs, RA, SLE, and MS, and four NHL subtypes, DLBCL, chronic lymphocytic leukemia (CLL), follicular lymphoma (FL), and MZL, to explore genetic commonalities between the two disease groups.

2 | MATERIALS AND METHODS

2.1 | MS, RA, SLE, and NHL data set characteristics

The MS study consists of 9,772 cases and 17,376 controls from the Wellcome Trust Case Control Consortium 2 (WTCCC2) project (International Multiple Sclerosis Genetics Consortium, 2011; Table 1). Individuals in this data set were of European descent and originated from 15 geographic regions, including the USA, Australia, New Zealand, and numerous European countries. Included in this data set were summary-level association results for a total of 464,434 single nucleotide polymorphisms (SNPs). The genotyping platform was the Illumina Human 660-Quad platform; quality control was performed by the original authors and a log-additive genetic model was used.

The RA study consists of a combined 3,921 cases and 4,079 controls of European descent from a meta-analysis of two datasets: WTCCC1 (Wellcome Trust Case Control Consortium, 2007) and the epidemiological investigation of rheumatoid arthritis (EIRA) data set (Padyukov et al., 2011; Table 1). The combined data set (union) had summary-level association results for 650,312 SNPs. The genotyping platform was the Illumina 300K chip; imputation and quality control were performed by the original authors and a log-additive genetic model was used.

The SLE study consists of combined 7,219 cases and 15,991 controls of European descent from the Bentham et al. (2015) multicenter study (Table 1). The study had summary-level association results for 623,954 SNPs. The genotyping platform was the Illumina HumanOmni1-Quad BeadChip; quality control was performed by the

TABLE 1 The GWAS used for the meta-analysis

Disease	Study	Unique cases	Unique controls	Original genotyped SNPs
MS	WTCCC2 (12)	9,772	17,376	465,434
RA	WTCCC1 (13) and EIRA (14)	3,921	4,079	650,312
SLE	Bentham study (15)	7,219	15,991	623,954
DLBCL	Groupe d'Etude des Lymphomes de l'Adulte (16)	549	525	513,264
DLBCL	Mayo-DLBCL (16)	393	172	516,286
DLBCL	San Francisco (16)	254	749	290,454
DLBCL	Omni	2,421	5,991	607,957
CLL	San Francisco	213	See SF above	290,454
CLL	Omni	1,953	See Omni above	607,957
CLL	Utah	326	413	559,899
FL	San Francisco (SF)	210	See SF above	290,454
FL	San Francisco (SF2)	119	349	599,547
FL	Scandinavian lymphoma etiology (SCALE)	376	791	297,989
FL	Omni (22 sites)	1,981	See Omni above	607,957
MZL	Omni (22 sites)	741	See Omni above	607,957

Abbreviations: CLL, chronic lymphocytic leukemia; DLBCL, diffuse large B-cell lymphoma; EIRA, epidemiological investigation of rheumatoid arthritis; FL, follicular lymphoma; GWAS, genome-wide association studies; MS, multiple sclerosis; MZL, marginal zone lymphomas; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; SNPs, single nucleotide polymorphisms; WTCCC2, Wellcome Trust Case Control Consortium 2.

original authors and a log-additive genetic model was used.

The NHL study consists of cases and controls from multiple studies of four B-Cell NHL subtypes: DLBCL (Cerhan et al., 2014), FL (Skibola et al., 2014), CLL (Berndt et al., 2016), and MZL (Vijai et al., 2015) (Table 1). Individuals in this data set were also of European descent and originated from the USA and numerous European countries. Together, these datasets include summary-level association results (actual and imputed) for a total of 9,116,853 SNPs for DLBCL, 9,116,853 SNPs for CLL, 9,078,855 SNPs for FL, and 8,478,065 SNPs for MZL.

To generate a single working data set containing association results for each AD and each subtype of NHL, the datasets were merged according to SNP name, giving a final data set containing summary-level results for a total of approximately 460,000 overlapping SNPs for MS and each NHL subtype, 600,000 overlapping SNPs for RA and each NHL subtype, and 600,000 SNPs for SLE and each NHL subtype. R and Plink statistical software were used for all subsequent analyses (Purcell et al., 2007; R Core Team, 2013).

2.2 | SNP-level overlap between diseases

For each of the 12 cross-disease analyses, we followed a procedure used in other meta-analyses of complex genetic diseases (Khankhanian et al., 2016). For example, to assess the genetic overlap between MS and DLBCL, we first identified SNPs that associated independently with either disease. Then, in each disease, we grouped SNPs by increasing significance by establishing seven association thresholds ranging from $p < 5 \times 10^{-8}$ to $p < 5 \times 10^{-1}$. From the collection of SNPs that reached a given threshold, we selected only independent subsets ($r^2 < 0.1$ in CEU), preferentially keeping SNPs with lower p values. (The CEU are controls of Northern and Western European ancestry from CEPH (Center d'Etude du Polymorphism Humain) a collection based on 1000 Genomes and HapMap genotype data; r^2 was downloaded using SNAP software.) Each subset of SNPs for each AD was tested for association with each NHL subtype. Association test statistics were adjusted for multiple testing using Benjamini–Hochberg's false discovery rate (FDR) method based on the total number of SNPs in the subset. An FDR< 0.05 was considered statistically significant. The reverse process was performed to test each set of NHL SNPs and AD risk. SNP-level analyses were conducted for each combination of one autoimmune disease (MS, RA, and SLE) and one subtype of NHL (DLBCL, CLL, FL, and MZL).

2.3 | Polygenic risk scores

Polygenic risk scores (PRS) were calculated to test the cumulative effect of SNPs associated with each AD on NHL and vice versa. For example, for the comparison of SLE and DLBCL, sets of top independent SNPs were chosen as described above. The SLE-PRS and the DLBCL-PRS were calculated for each individual; the PRS is defined as the weighted sum of the number of risk alleles at each SNP in the set, weighted by the log odds ratio of association for each SNP (Khankhanian et al., 2016). We assessed the ability of the SLE-PRS to distinguish DLBCL cases from controls and the ability of the DLBCL-PRS to distinguish SLE cases from controls using the Nagelkerke R^2. This analysis was repeated for each combination of one NHL subtype (DLBCL, FL, CLL, and MZL) and one AD (SLE, RA, and MS).

2.4 | Meta-Analysis

To identify novel susceptibility loci in our merged data set, we combined summary results from each AD and each NHL subtype in a meta-analysis. For each pair of diseases, for all overlapping SNPs, discovery-level *p* values and odds ratios (OR) from the AD and NHL datasets (as provided by the authors of those original studies) were combined using a fixed-effects meta-analysis as implemented in the Plink software package. The p-value threshold for Cochrane's Q statistic was set to 0.05 to screen for heterogeneity in results across studies.

2.5 | Diseasome

To visualize the similarities between ADs and the NHL subtypes, we built a human disease network based on disease proximities, as previously described (Khankhanian et al., 2016; Himmelstein, Khankhanian, Baranzini, 2015). Briefly, proximity was calculated using a random walk with restart over a heterogeneous network wherein diseases are connected by shared genetic etiology, as determined by databases of previously published data. Two diseases with greater shared genetic etiology will have greater proximity due to a larger number of connections. The mean proximity between NHLs and ADs was compared to the mean proximity between NHLs and solid cancers with the Fisher test. Similarly, the mean proximity between NHL and solid cancers was compared to the mean proximity between NHL and all other diseases (Khankhanian et al., 2016; Himmelstein et al., 2015).

3 | RESULTS

3.1 | Overview

A total of 9,772 MS patients, 3,921 RA patients, 7,219 SLE patients, 3,617 DLBCL patients, 2,492 CLL patients, 2,686 FL patients, 741 MZL patients, and 46,436 total

Overlapping SNPs	**MS** 465,434 SNPs	**SLE** 623,954 SNPs	**RA** 603,325 SNPs
CLL 9,116,853 SNPs	460,908 SNPs	609,999 SNPs	594,537 SNPs
DLBCL 9,116,853 SNPs	460,981 SNPs	610,319 SNPs	594,517 SNPs
FLL 9,078,855 SNPs	461,003 SNPs	610,241 SNPs	594,551 SNPs
MZL 8,478,065 SNPs	459,792 SNPs	605,797 SNPs	593,751 SNPs

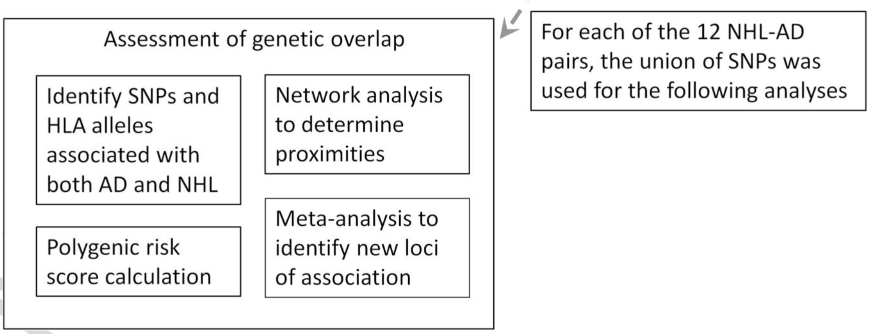

FIGURE 1 Study design and data analysis procedures. For each of the 12 pairs of diseases (three ADs and four NHLs), results from previous GWAS were used to assess genetic overlap between the two diseases. SNPs independently associated with both diseases were identified. Genetic risk scores were evaluated for genomewide overlap. Network analysis evaluated the proximity of these diseases in the context of other human diseases. After the evaluation of genetic overlap, we merged GWAS results for each AD-NHL in a meta-analysis to discover novel genes associated with both diseases. AD, autoimmune disease; GWAS, genome-wide association studies; NHL, non-Hodgkin lymphoma; SNPs, single nucleotide polymorphisms

controls were analyzed. Figure 1 gives an overview of study design and data analysis. For each of 12 pair-wise comparisons, comparing one of four NHL subtypes against one of three ADs, the following analyses are presented. First, we present SNPs that associated independently with both diseases. Next, we present PRS to assess the cumulative genome-wide effect of AD-associated SNPs on NHL and of NHL-associated SNPs on AD. To identify susceptibility genes common to each of 12 disease pairs, a series of 12 GWAS meta-analyses are presented. Finally, the three ADs and the four NHL subtypes from this study were mapped in a genetic diseasome, a network of diseases, with other ADs, NHLs, solid cancers, and other unrelated diseases, and relative proximity of diseases are presented.

3.2 | SNP and HLA-allele overlap between ADs and NHLs

Each of the three ADs was evaluated for SNP-level overlap with each of the four NHLs, resulting in 12 comparisons. The comparison of SLE versus DLBCL is detailed as an example (Table 2, Row 1). We identified 2,472 SNPs that associated with SLE at a significance threshold of $p < 5 \times 10^{-4}$. After discarding SNPs for which linkage disequilibrium (LD) information was not available, 1,718 SNPs remained to represent 389 independent regions (with $r^2 < 0.1$ as the threshold to define independence). Of the 389 SNPs (one SNP per independent region), two of these were significantly associated with DLBCL ($p < .05$ after Benjamini–Hochberg correction for 389 multiple tests). Similar results were found when DLBCL-associated SNPs were assessed for association with SLE (Table 2, Row 2). Details regarding the individual overlapping SNPs are given in Table S1.

The analysis was repeated for other ADs and other NHL subtypes. In each comparison, a relatively small number of overlapping regions was identified, at most 14. The greatest amount of overlap was observed in the comparisons between MS and CLL and between MS and FL; SLE had a smaller number of overlapping SNPs with the NHL subtypes, and RA had the smallest number of overlapping SNPs with the NHL subtypes. The differences in the amount of overlap between specific ADs were small, although it should be noted that this analysis was not equipped to make quantitative assertions about the significance of the difference in overlap (as these differences are highly dependent on other factors including the difference in power between studies).

This analysis was repeated with the initial significance thresholds ranging from $p < 5 \times 10^{-8}$ to $p < 5 \times 10^{-1}$; while the results in Table 2 reflect a threshold of $p < 5 \times 10^{-4}$ as an example, a similar pattern of results held at other thresholds. Details of the SNPs comprising this overlap are given in Table S1.

3.3 | Polygenic risk-overlap between diseases

To assess the extent of genetic risk overlap between AD and NHL subtypes at the genome-wide level (including human leukocyte antigen (HLA) region), PRS, termed MS-PRS, RA-PRS, SLE-PRS, DLBCL-PRS, CLL-PRS, FL-PRS, and MZL-PRS, were calculated.

In each of the seven individual diseases, the mean PRS was higher in cases than in controls as expected. However, when the PRS of ADs were calculated in NHL subtypes, the score was not significantly different between cases and controls (Table S2). Similarly, when PRS of NHL subtypes were calculated in ADs, the scores were not significantly different between cases and controls.

3.4 | Meta-analysis

We combined each of the three ADs with each of the four NHL GWAS in a series of 12 meta-analyses to leverage increased statistical power for the discovery of novel SNPs associated with both diseases. In Table 3, we report a list of SNPs that had statistically significant association in a meta-analysis of an AD with an NHL subtype, but which did not meet the discovery threshold of significance in the AD alone nor in the NHL subtype alone (though they may not have met the strict definition of genome-wide significance threshold as defined in our study, some of these hits had been carried forward by the original authors to validation on additional samples and subsequently been reported as significant in the original discovery studies). SNPs that passed the analysis paper-wide significance threshold (the "paper-wide threshold" includes correction for total number of tests performed in the total of 12 meta-analyses reported in this paper) are reported in Table 3. SNPs that passed a study-wide significance (after correction only for the total number of SNPs in each meta-analysis) are shown in Table S3.

3.5 | Diseasome

We reviewed 87 diseases (Table 4) for which sufficient GWAS results were available in the public domain. Pair-wise proximities between these diseases were calculated based on the degree of genome-wide genetic overlap. A graph of the proximity space reveals a cluster of 19

TABLE 2 Overlap of SNPs between the three autoimmune diseases and the four non-Hodgkin lymphoma subtypes

Analysis	Number of SNPs that were significant at the threshold of 5e-04	Number of SNPs with LD info available in SNAP	Number of regions based on LD	Number of SNPs that were significant after correction (BH < 0.05)
SLE SNPs in DLBCL	2,472	1,718	389	2
DLBCL SNPs in SLE	524	334	190	3
SLE SNPs in CLL	2,471	1,718	389	5
CLL SNPs in SLE	895	625	240	4
SLE SNPs in FL	2,473	1,718	389	2
FL SNPs in SLE	558	393	206	5
SLE SNPs in MZL	2,462	1,718	389	4
MZL SNPs in SLE	390	259	168	6
RA SNPs in DLBCL	532	423	238	0
DLBCL SNPs in RA	504	425	190	1
RA SNPs in CLL	531	423	238	1
CLL SNPs in RA	844	724	232	1
RA SNPs in FL	532	423	238	3
FL SNPs in RA	471	395	199	4
RA SNPs in MZL	530	422	237	1
MZL SNPs in RA	351	283	148	2
MS SNPs in DLBCL	1,203	1,031	380	6
DLBCL SNPs in MS	366	329	195	6
MS SNPs in CLL	1,204	1,031	380	13
CLL SNPs in MS	659	586	244	14
MS SNPs in FL	1,203	1,031	380	11
FL SNPs in MS	402	369	203	12
MS SNPs in MZL	1,203	1,031	380	0
MZL SNPs in MS	253	219	152	1

Note: See supplementary table for details of each region.

Abbreviation: DLBCL, diffuse large B-cell lymphoma; FL, follicular lymphoma; GWAS, genome-wide association studies; LD, linkage disequilibrium; MS, multiple sclerosis; MZL, marginal zone lymphomas; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; SNPs, single nucleotide polymorphisms.

TABLE 3 SNPs that were significant in a meta-analysis of an autoimmune disease with a non-Hodgkin Lymphoma, but which did not meet the threshold of significance in the autoimmune disease alone nor in the non-Hodgkin Lymphoma alone

Study	SNP	*p* (AD)	OR (AD)	*p* (NHL)	OR (NHL)	*p* (Meta)	Corr. *p* (Meta)	Paper corr. *p* (Meta)	OR (Meta)	Chr	Gene(s) of interest	RA	RDS
CLL vs. MS	rs140522	3.85E-06	0.91	1.18E-05	0.86	6.49E-11	2.99E-05	4.32E-04	0.90	22	ODF3B	A	4
CLL vs. MS	rs6793295	1.48E-05	0.91	1.10E-04	0.87	1.86E-09	8.59E-04	1.24E-02	0.90	3	LRRC34	A	7
CLL vs. RA	rs3731714	1.33E-03	0.89	7.82E-07	0.84	7.05E-09	4.19E-03	4.69E-02	0.87	2	CASP10, PPIL3, CFLAR	G	1d
DLBCL vs. MS	rs2425752	1.70E-06	0.91	1.10E-02	0.92	5.10E-09	2.35E-03	3.39E-02	0.91	20	NCOA5	A	1d
MZL vs. RA	rs16947122	3.56E-02	1.57	4.99E-03	0.51	5.03E-09	2.99E-03	3.35E-02	1.86	12	FBXW8, HRK, TESC	C	5
MZL vs. RA	rs1364229	1.73E-04	1.30	1.66E-04	0.72	1.66E-10	9.86E-05	1.10E-03	1.35	16	CDH8	A	7
MZL vs. RA	rs7192064	9.63E-04	0.79	3.67E-04	0.74	6.55E-09	3.89E-03	4.36E-02	0.76	16	CDH8	G	
MZL vs. RA	rs2131402	2.50E-04	0.77	3.67E-04	0.74	1.51E-09	8.97E-04	1.01E-02	0.75	16	CDH8	G	6
CLL vs. SLE	rs1439112	1.80E-07	0.85	3.84E-03	1.10	7.09E-09	4.33E-03	4.72E-02	0.88	2	MGAT5	A	4
CLL vs. SLE	rs10936599	1.99E-05	0.87	5.01E-05	0.86	4.06E-09	2.48E-03	2.70E-02	0.87	3	MYNN, ACTRT3, TERC, LRRC34	C	5
CLL vs. SLE	rs1317082	1.50E-05	0.86	3.73E-05	0.86	2.25E-09	1.37E-03	1.50E-02	0.86	3	MYNN, ACTRT3, TERC, LRRC34	A	6
CLL vs. SLE	rs13069553	9.55E-06	0.86	4.16E-05	0.86	1.61E-09	9.83E-04	1.07E-02	0.86	3	MYNN, ACTRT3, TERC, LRRC34	A	5
CLL vs. SLE	rs7621631	1.36E-05	0.86	4.92E-05	0.86	2.69E-09	1.64E-03	1.79E-02	0.86	3	MYNN, ACTRT3, TERC, LRRC34	C	7
CLL vs. SLE	rs10069690	7.21E-04	1.12	5.56E-07	1.21	4.60E-09	2.81E-03	3.06E-02	1.16	5	TERT	T	5

Note: RA, risk allele. RDS, regulomeDB score. corr, Corrected for multiple hypothesis testing in a single meta-analysis. Paper corr, corrected for multiple hypothesis testing in 12 meta-analyses presented in this paper.
Abbreviations: AD, autoimmune diseases; CLL, chronic lymphocytic leukemia; DLBCL, diffuse large B-cell lymphoma; MS, multiple sclerosis; MZL, marginal zone lymphomas; NHL, non-Hodgkin lymphomas; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

TABLE 4 Classification of immune and neoplastic diseases from the diseasome

Autoimmune diseases	Hematologic cancers
Alopecia areata (AR)	Chronic lymphocytic leukemia (CLL)
Ankylosing spondylitis (AS)	Hodgkin lymphoma (HL)
Behcet's disease (Beh)	Mutiple myeloma (MM)
Celiac disease (Cel)	Diffuse large B-cell lymphoma (DLBCL)
Crohn's Disease (CD)	Follicular lymphoma (FL)
Graves' Disease (GD)	Marginal zone lymphoma (MZL)
IgA glomerulonephritis (IgA)	
Kawasaki disease (Kaw)	**Solid cancers**
Multiple Sclerosis (MS)	Basal cell carcinoma (BCC)
Primary biliary cirrhosis (PBC)	Bladder carcinoma (BlC)
Psoriasis (Ps)	Breast carcinoma (BrC)
Psoriatic arthritis (PsA)	Central nervous system cancer (CNS)
Rheumatoid arthritis (RA)	Esophageal carcinoma (EsC)
Sclerosing cholangitis (PSC)	Lung Carcinoma (LuC)
Systemic lupus erythematosus (SLE)	Lung adenocarcinoma (LuA)
Systemic scleroderma (SS)	Melanoma (Mel)
Type 1 diabetes mellitus (T1D)	Neuroblastoma (NB)
Ulcerative colitis (UC)	Ovarian carcinoma (OvC)
Vitiligo (Vit)	Pancreatic carcinoma (PaC)
	Prostate carcinoma (PrC)
	Renal cell carcinoma (RCC)
	Squamous cell carcinoma (SCC)
	Stomach carcinoma (StC)
	Thyoid carcinoma (ThC)

autoimmune diseases, a cluster of many of the 16 available solid cancers, and a cluster of the four NHLs, which has closer common genetic risk overlap with autoimmune diseases than with solid cancers in this two-dimensional projection (Figure 2, Panel 1). The mean pair-wise proximity metric between NHL subtypes and autoimmune diseases was higher than the mean proximity between NHLs and solid cancers (0.0049 vs. 0.0023, $p = .0041$, Figure 2, Panel 2). The mean pair-wise proximity between NHL variants and solid cancers was higher than the mean proximity between NHL and all other diseases (0.0023 vs. 0.0012, $p = .00066$, Figure 2, Panel 2).

4 | DISCUSSION

In an effort to understand the association between AD and NHL, we performed a series of analyses exploring the genetic overlap between four NHL subtypes and three ADs. We found that only a small number of risk loci associated with NHL were also associated with AD risk, and, conversely, that only a small number of AD risk loci were associated with risk of the NHL subtypes studied. Polygenic risk score analysis, which considers a large number of genes and places less relative weight on the top few genes, did not demonstrate the significant genome-wide polygenic overlap between any of the NHL subtypes and any of the AD examined in this study. Diseasome analysis, in contrast to polygenic risk score analysis, places larger relative weight on a fewer number of confirmed top genes. Diseasome analysis revealed that the NHL subtypes tend to occupy a common genetic risk neighborhood and that this common neighborhood is closer to the group of ADs than to the group of solid cancers. Thus, we conclude that while few risk loci overlap between any pair of the studied diseases, there is not enough genetic overlap found in this study to explain an important proportion of increased risk (less than one percent of disease risk explained based on PRS analysis, Table S2).

Altogether, within the limitations inherent in the available data our findings provide little evidence that shared genetic risk factors are a major explanation for the increased risk of malignant B-cell lymphomas in patients with autoimmune diseases, such as RA and SLE (Baecklund et al., 2014). As this is also the case for known environmental risk factors (Thun et al., 2017; Deane et al., 2010) (Park et al., 2009; Belbasis et al., 2015; Smedby & Ponzoni, 2017; Ekström et al., 2003; Bernatsky et al., 2013), other mechanisms, such as inflammation and chronic antigenic stimulation which increase B- and T-cell receptor rearrangement and B-cell somatic hypermutation, and/or AD treatment with immunosuppressive or biologic therapy, seem likely to be more significant contributors to the long-standing association between the two disease groups. The collective findings further suggest that monitoring and managing inflammation or other factors associated with the disease course as the way to reducing the risk of malignant B-cell lymphoma in patients with AD (Baecklund et al., 2006).

A series of 12 meta-analyses of the three individual ADs with the four individual NHL subtypes demonstrated seven regions which passed a genome-wide threshold of significance in the 12 meta-analyses, which would not have been discovered in the analysis of the individual diseases due to limited power (Table 3). The corresponding effect sizes were modest and total risk

COLOR FIG

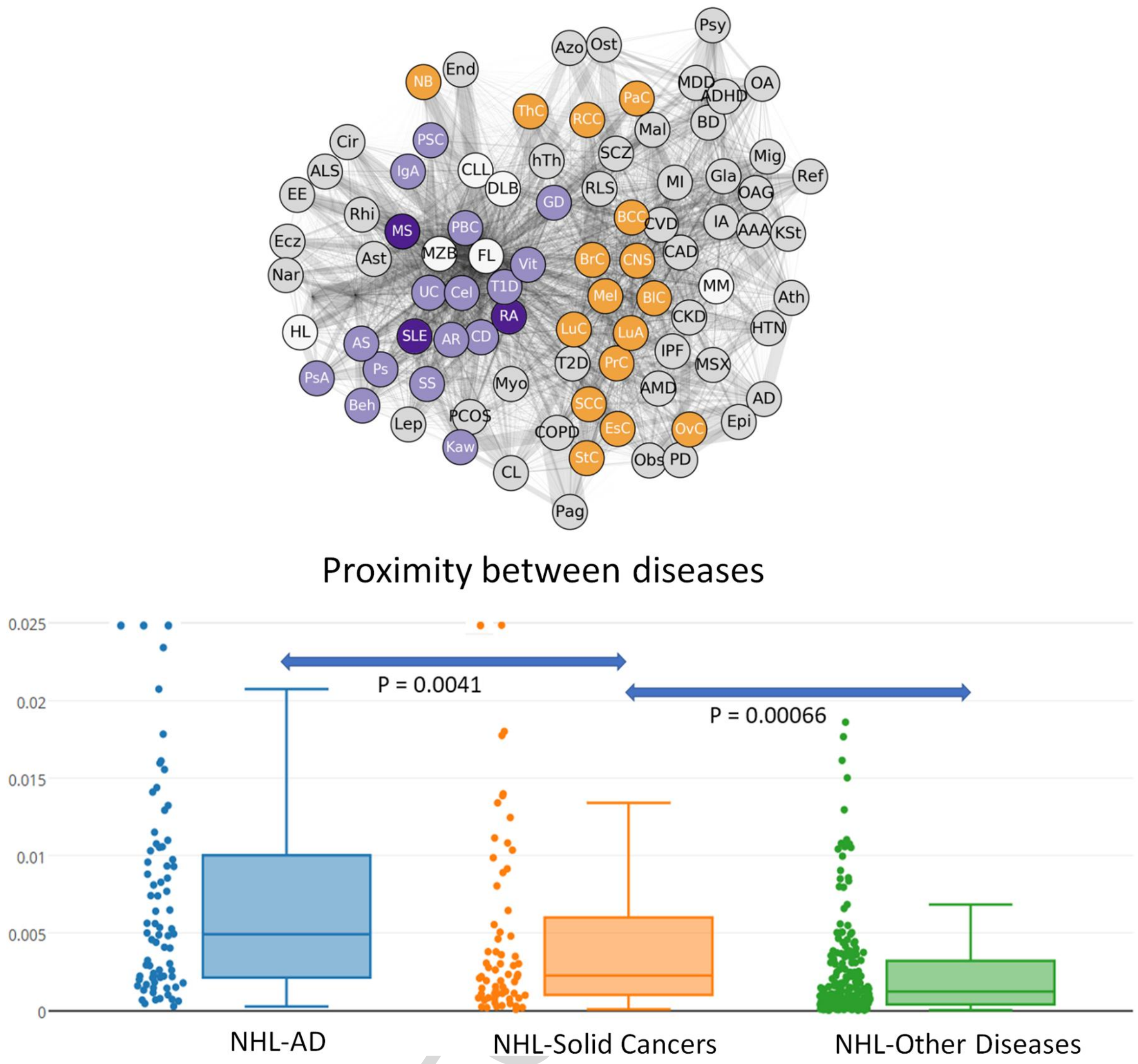

FIGURE 2 Panel 1. A graph of autoimmune diseases (purple), solid cancers (orange), hematologic cancers (white), and other diseases (gray). The thickness of lines indicates greater levels of genetic overlap (proximity between diseases). Panel 2. The proximity between NHLs and ADs (blue) is greater than the proximity between NHLs and solid cancers (orange), which is greater than the proximity between NHLs and other diseases (green). ADs, autoimmune diseases; NHLs, non-Hodgkin lymphomas

explained was low, however, the genes in these regions are discussed in a Supporting Text. In brief, the list comprises genes involved in other cell proliferation and specifically hematopoiesis, telomerase activity, and antigen presentation (via, e.g., *MGAT5*). Many of these genes have since been implicated in the ADs and NHLs examined in this manuscript (as larger meta-analyses of the individual ADs and NHLs have been published), which lends credibility to the present findings and supports the potential advantage of the cross-disease meta-analysis approach. Given the availability of studies of the individual ADs and NHLs with larger sample sizes, a repeat meta-analysis would be possible.

There are noteworthy limitations to this study. First, this is a post hoc secondary endpoint analysis; validation in an independent data set would be required to confirm the specific meta-analysis findings, and a series of in vitro and in vivo studies would be required to elucidate mechanisms and imply causation. Some of the individual NHL subtype GWAS were of relatively small sample size and therefore, the statistical power in these analyses was limited. A lack of whole-exome coverage in a genome-wide study is another limitation; GWAS offers incomplete coverage and an imperfect view of the human genome compared with newer methods. An expansion cross-disease analysis to larger datasets with greater coverage would be of significant value. We completed 12 parallel meta-analyses, which further imposed a limitation on power; the multiple-hypothesis correction for this additional layer of hypothesis testing raised the threshold of genome-wide significance by one order of magnitude and thus limited the power of new discovery. There were many meta-analyses hits that reached genome-wide significance but not paper-wide significance; the vast majority of these were the hits that have been confirmed in recent published literature, suggesting that perhaps future meta-analyses should focus on individual disease pairs, thus avoiding the additional limitation of parallel meta-analysis. The diseasome analysis was limited by an inability to control for overlap in the control datasets of the individual GWAS used to construct the diseasome. In particular, for the diseases that were not classified as NHL or AD, caution against overinterpretation of clusters of diseases with shared GWAS controls is warranted.

The three ADs and the four NHL subtypes presented here were selected because data were available and we were able to create a relationship with the respective consortia. It would be of value for future endeavors to study other autoimmune diseases such as Sjögren's syndrome and other lymphomas such as Hodgkin's lymphoma via a similar analysis pipeline, especially given the observed epidemiologic links between those other syndromes and the ones presented in this study.

5 | CONCLUSION

Within the limits of this GWAS-based cross-disease analysis, we estimated that the shared genetic risk between the three autoimmune diseases and four non-Hodgkin lymphoma subtypes is limited to a handful of genes. This finding suggests that genetic etiology is not the primary driver in the observed epidemiologic link between AD and NHL, but rather the link may be driven by nongenetic factors, such as chronic antigenic stimulation and inflammation or immune-modulating treatment. A meta-analysis of ADs with NHLs suggested new candidate genes to explain the limited shared genetic risk, with roles in the cell cycle, apoptosis, and telomere length. Further meta-analyses of genetic variants in autoimmune diseases and lymphomas with larger datasets and deeper sequencing may provide further insight into mechanisms common to the two groups of diseases.

ACKNOWLEDGMENTS

Support for authors

All 92 authors agree that there are no conflicts of interest to report.

Hans-Olov Adami is supported by the Karolinska Institutet Distinguished Professor Award (Dnr: 2368/10-221).

Nicola Camp reports support from NIH R01CA134674. 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 Comprehensive 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.

James R Cerhan is supported by NIH grants R01 CA92153, R01 CA200703, P50 CA97274, U01 CA195568, and the Henry J Predolin Foundation.

David Conti is funded by the following grants: P01CA196569, R01CA201407, R01CA140561, ES024844, and R01ES016813.

Lenka Foretova is supported by a grant MH CZ - DRO (MMCI, 00209805).

Pierre-Antoine Gourraud is supported by the ATIP-Avenir INSERM program and the Region Pays de Loire ConnecTalent, ARSEP Foundation (France) and the Nantes University Foundation.

Karin E Smedby is supported by the Swedish Cancer Society (2009/659, 02 6661), the Strategic Research Program in Epidemiology at Karolinska Institute and National Institutes of Health (5R01 CA69669-02).

John J Spinelli is supported by the Canadian Institutes for Health Research (CIHR); the Canadian Cancer Society; and the Michael Smith Foundation for Health Research

Support for 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, HHSN261201000006C).

BC (J.S., A.B.W.): Canadian Institutes for Health Research (CIHR). Canadian Cancer Society. Michael Smith Foundation for Health Research.

CLL Meta-Analysis: We thank I. Brock, K. Butterbach, A. Chabrier, D. Chan-Lam, D. Connley, H. Cramp, R. Cutting, C. Dalley, H. Dykes, A. Gabbas, P. Gaddam, P. Hui, L. Irish, L. Jacobus, S. Kaul, L. Klareskog, A. Lai, J. Lunde, M. McAdams, L. Padyukov, D. Parisi, V. Rajamanickam, T. Rattle, L. Rigacci, R. Sargent, G. Specchia, M. Stagner, P. Taylor, C. Tornow, J. WiIliams and G. Wood. The overall GWAS project was supported by the intramural program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, US National Institutes of Health.

CPSII (L.T.): The American Cancer Society funds the creation, maintenance, and updating of the CPS-II cohort. The authors thank the CPS-II participants and Study Management Group for their invaluable contributions to this research. 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.

EIRA: The authors acknowledge the help of Peter K. Gregersen and Jane Worthington in access to NARAC and WTCCC data. They also thank Robert M. Plenge for careful reading of the manuscript and critical discussions; Kian Mun Chan, Boon Yeong Goh, Wee Yang Meah, Jameelah B. S. Mohamed, Jason Ong, Eileen Ping, and Sigeeta Rajaram for their invaluable laboratory assistance; the participating patients with RA and controls and all rheumatologists for recruiting patients in the EIRA study; and Marie-Louise Serra, Camilla Bengtsson, Eva Jemseby, and Lena Nise for their invaluable contributions to the collection of data and maintenance of the database.

ELCCS (E.R.): The ECSG Lymphoma Case-Control Study (ELCCS) were funded by Bloodwise and Leukemia & 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 references 2014SRG756 and 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.

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), the National Center for Advancing Translational Science (UL1 TR000135).

GELA (G.S.): The French National Cancer Institute (INCa).

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

IMSGC-WTCCC2: The principal funding for this study was provided by the Wellcome Trust (085475/B/08/Z, 085475/Z/08/Z, 075491/Z/04/Z, and 068545/Z/02). The work was also supported by National Institutes of Health (AI076544, NS032830, NS049477, NS19142, NS049510, NS26799, NS43559, NS067305, CA104021, RR020092, RR024992 and K23N/S048869), US National Multiple Sclerosis Society (RG 4201-A-1), Nancy Davis Foundation, Cambridge NIHR Biomedical Research Center, UK Medical Research Council (G0700061, G0000934), Multiple Sclerosis Society of Great Britain and Northern Ireland (898/08), Wolfson Royal Society Merit Award, Peter Doherty fellowship, Lagrange Fellowship, Harry Weaver Neuroscience Scholarships, Australian National Health and Medical Research Council (NHMRC), Australian Research Council Linkage Program Grant, JHH Charitable Trust Fund, Multiple Sclerosis Research Australia, Health Research Council New Zealand, National MS Society of New Zealand, Wetenschappelijk Onderzoek Multiple Sclerose, Bayer Chair on Fundamental Genetic Research regarding the Neuroimmunological Aspects of Multiple Sclerosis, Biogen Idec Chair Translational Research in Multiple Sclerosis, FWO-Vlaanderen, Belgian Neurological Society, Danish Multiple Sclerosis Society, Neuropromise EU grant (LSHM-CT-2005-018637), Center of Excellence for Disease Genetics of the Academy of Finland, Sigrid Juselius Foundation, Helsinki University Central Hospital Research Foundation, Bundesministerium für Bildung und Technologie (KKNMS consortium Control MS), Deutsche Forschungsgemeinschaft, Institut National de la Santé et de la Recherche Médicale (INSERM), Association pour la Recherche sur la Sclérose En Plaques (ARSEP), Association Française contre les Myopathies (AFM), Italian

Foundation for Multiple Sclerosis (FISM grants 2002/R/40, 2005/R/10, 2008/R/11, and 2008/R/15), Italian Ministry of Health (grant Giovani Ricercatori 2007 - D.lgs 502/92), Regione Piemonte (grants 2003, 2004, 2008, 2009), CRT Foundation, Turin, Moorfields/UCL Institute of Ophthalmology NIHR Biomedical Research Center, Norwegian MS Register and Biobank, Research Council of Norway, South-Eastern and Western Norway regional Health Authories, Ullevål University Hospital Scientific Advisory Council, Haukeland University Hospital, Amici Centro Sclerosi Multipla del San Raffaele (ACESM), Association of British Neurologists, Spanish Ministry of Health (FISPI060117), Bibbi and Niels Jensens Foundation, Montel Williams foundation, Hjärnfonden and Swedish medical research council (8691), Stockholm County Council (562183), Swedish Council for Working life and Social Research, Gemeinnützige Hertie Stiftung, Northern California Kaiser Permanente members and Polpharma Foundation, and Washington University Institute of Clinical and Translational Sciences—Brain, Behavioral and Performance Unit. We acknowledge use of data from the British 1958 Birth Cohort, the UK National Blood Service, the popgen biobank, the KORA and MONICA Augsburg studies, the Accelerated Cure Project, the Brigham & Women's Hospital PhenoGenetic Project, the Swedish CAD project, the Norwegian Bone Marrow Donor Registry, the Children's Hospital of Philadelphia (CHOP), the Swedish Breast Cancer study, BRC-REFGENSEP (Pitié-Salpêtrière Center d'Investigation Clinique (CIC) and Généthon) and HYPERGENES (HEALTH-F4-2007-201550). Projects received support from the German Ministry of Education and Research, the Helmholtz Zentrum München—National Research Center, the German National Genome Research Network (NGFN), the LMUinnovativ, the Knut and Alice Wallenberg Foundation, the Center for Applied Genomics from the Children's Hospital of Philadelphia Development Award, the Agency for Science & Technology and Research of Singapore, and the Susan G. Komen Breast Cancer Foundation.

Iowa-Mayo SPORE (G.W., J.R.C., T.E.W.): National Institutes of Health (CA97274). NCI Specialized Programs of Research Excellence (SPORE) in Human Cancer (P50 CA97274). Molecular Epidemiology of Non-Hodgkin Lymphoma Survival (R01 CA129539), National Cancer Institute (P30 CA086862, P30 CA15083), 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 (LR7 CRP-59812/2012), Fondazione Banco di Sardegna 2010–2012.

LYSA (G.S, H.G): Institut National du Cancer (INCa, Paris) grant 2008-020

Mayo Clinic Case-Control (J.R.C.): National Institutes of Health (R01 CA92153). National Center for Advancing Translational Science (UL1 TR000135). National Cancer Institute (P30 CA015083).

MCCS (G.G.G., G.S.): The Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian National Health and Medical Research Council grants 209057, 251553, 504711, 396414 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the National Death Index and the Australian Cancer Database. The MCCS was made possible by the contribution of many people, including the original investigators, the teams that recruited the participants and continue working on follow-up, and the many thousands of Melbourne residents who continue to participate in the study.

MCC-Spain: The MCC-Spain study is funded by the Instituto de Salud Carlos III (ISCIII – Spanish Government) (PI11/01810, PI14/01219, RCESP C03/09, and CIBERESP); the Agencia de Gestio d'Ajuts Universitaris i de Recerca (AGAUR) – Generalitat de Catalunya (Catalonian Government) (2014SGR756). Nadia García and Marleny Vergara (ICO-IDIBELL) provided technical support for this study.

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. ENCODE, U01 HG007033 (J.V.).

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 NHL study (C.M.Vajdic): It 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.

NSFC: the National Natural Science Foundation of China (No. 61471078).

NYUWHS: 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.

SLE: The Health and Retirement Study genetic data were obtained from dbGaP under accession phs000187.v1; the study is sponsored by the National Institute on Aging (grant numbers U01AG009740, RC2AG036495, and RC4AG039029) and was conducted by the University of Michigan. The melanoma study data were obtained from dbGaP under accession number phs000187.v1.p1. Research support to collect data and develop an application to support this project was provided by 3P50CA093459, 5P50CA097007, 5R01ES011740, and 5R01CA133996. Funding support for the Genes and Blood Clotting study was provided through the NIH/NHLBI (R37HL039693). The Genes and Blood Clotting Study is one of the Phase 3 studies as part of the Gene Environment Association Studies (GENEVA) under GEI. Assistance with genotype cleaning was provided by the GENEVA Coordinating Center (U01 HG004446). Funding support for DNA extraction and genotyping, which was performed at the Broad Institute, was provided by NIH/NHLBI (R37HL039693). Additional support was provided by the Howard Hughes Medical Institute. The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession number phs000304.v1.p1. The CGEMS prostate cancer study data were obtained from dbGaP under accession phs000207v1

SCALE (K.E.S., H.O.A., H.H.): Swedish Cancer Society (2009/659).Stockholm Country Council (20110229) 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.

UCSF/UCSF2 (E.A.H., P.M.B, C.F.S.): The UCSF studies were supported by the NCI, National Institutes of Health, CA87014, CA1046282, CA122663 and CA154643, and R01CA87014, R01CA45614, R03CA14397, and R03CA150037 (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.

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 Center. We thank the NCRI Haemato-oncology Clinical Studies Group, colleagues in the North Trent Cancer Network the North Trent Haemato-oncology Database.

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, HHSN-268201100002C, HHSN268201100003C, HHSN268201-100004C, and HHSN271201100004C.

WTCCC1: The principal funder of this project was the Wellcome Trust. Case collections were funded by: Arthritis Research Campaign, BDA Research, British Heart Foundation, British Hypertension Society, Diabetes UK, Glaxo-Smith Kline Research and Development, Juvenile Diabetes Research Foundation, National Association for Colitis and Crohn's disease, SHERT (The Scottish Hospitals Endowment Research Trust), St Bartholomew's and The Royal London Charitable Foundation, UK Medical Research Council, UK NHS R&D and the Wellcome Trust. Statistical analyses were funded by a Commonwealth Scholarship, EU, EPSRC, Fundação para a Ciência e a Tecnologia (Portugal), National Institutes of Health, National Science Foundation and the Wellcome Trust. We acknowledge the many physicians, research fellows and research nurses who contributed to the various case collections, and the collection teams and senior management of the UK Blood Services responsible for the UK Blood Services Collection. For the 1958 Birth Cohort, venous blood collection was funded by the UK Medical Research Council and cell-line production, DNA extraction, and processing by the Juvenile Diabetes Research Foundation and the Wellcome Trust.

YALE (T.Z.): National Cancer Institute (CA62006).

Grant numbers for authors

Hans-Olov Adami is supported by the Karolinska Institutet Distinguished Professor Award (Dnr: 2368/10-221).

Nicola Camp reports support from NIH R01CA134674. 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 Comprehensive 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.

James R Cerhan is supported by NIH grants R01 CA92153, R01 CA200703, P50 CA97274, U01 CA195568, and the Henry J Predolin Foundation.

David Conti is funded by the following grants: P01CA196569, R01CA201407, R01CA140561, ES024844, and R01ES016813.

Lenka Foretova is supported by a grant MH CZ - DRO (MMCI, 00209805).

Pierre-Antoine Gourraud is supported by the ATIP-Avenir INSERM program and the Region Pays de Loire ConnecTalent, ARSEP Foundation (France) and the Nantes University Foundation.

Karin E Smedby is supported by the Swedish Cancer Society (2009/659, 02 6661), the Strategic Research Program in Epidemiology at Karolinska Institute and National Institutes of Health (5R01 CA69669-02).

John J Spinelli is supported by the Canadian Institutes for Health Research (CIHR); the Canadian Cancer Society; and the Michael Smith Foundation for Health Research

DATA AVAILABILITY

Individual cohorts contributing to the meta-analysis should be contacted directly as each cohort has different data access policies. We have included citations for data sources in the reference section.

KEY MESSAGES

Within the limits of this GWAS-based cross-disease analysis, the shared genetic risk between SLE, RA, MS, and four common B-cell NHL types was limited to few weakly-associated loci and explained little total disease risk. Candidate genes with roles in the cell cycle, apoptosis, and telomere length should be considered in future analyses of shared genetic susceptibility to these conditions. Further meta-analyses of genetic variants in autoimmune diseases and lymphomas with larger datasets and deeper sequencing may provide further insight into mechanisms common to the two groups of diseases.

ORCID

Pouya Khankhanian http://orcid.org/0000-0001-8075-4176

REFERENCES

Baecklund, E., Iliadou, A., Askling, J., Ekbom, A., Backlin, C., Granath, F., ... Klareskog, L. (2006). Association of chronic inflammation, not its treatment, with increased lymphoma risk in rheumatoid arthritis. *Arthritis and Rheumatism*, *54*, 692–701.

Baecklund, E., Smedby, K. E., Sutton, L. A., Askling, J., & Rosenquist, R. (2014). Lymphoma development in patients with autoimmune and inflammatory disorders - What are the driving forces? *Seminars in Cancer Biology*, *24*, 61–70.

Belbasis, L., Bellou, V., Evangelou, E., Ioannidis, J., & Tzoulaki, I. (2015). Environmental risk factors and multiple sclerosis: An umbrella review of systematic reviews and meta-analyses. *Lancet Neurol*, *14*, 263–273.

Bentham, J., Morris, D., Cunninghame Graham, D., Pinder, C., Tombleson, P., Behrens, T., ... Vyse, T. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nature Genetics*, *47*, 1457–1464.

Bernatsky, S., Ramsey-Goldman, R., Labrecque, J., Joseph, L., Boivin, J. F., Petri, M., ... Clarcke, A. E. (2013). Cancer risk in systemic lupus: An updated international multi-centre cohort study. *Journal of Autoimmunity*, *42*, 130–135.

Bernatsky, S., Velásquez García, H. A., Spinelli, J. J., Gaffney, P., Smedby, K. E., Ramsey-Goldman, R., ... Angelucci, E. (2017). *Lupus-related single nucleotide polymorphisms and risk of diffuse large B-cell lymphoma. Lupus Sci Med*, *4*(1), 187.

Berndt, S., Camp, N., Skibola, C., Vijai, J., Wang, Z., Gu, J., ... Slager, S. (2016). Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat Commun*, *7*, 1–9.

Cerhan, J., Berndt, S., Vijai, J., Ghesquières, H., McKay, J., Wang, S., ... Chanock, S. (2014). Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nature Genetics*, *46*, 1233–1238.

Himmelstein, D., Khankhanian, P., & Baranzini, S. (2015). A human disease network from GWAS loci (Version v1.0). *Zenodo*. http://doi.org/10.5281/zenodo.23025

Deane, K., O'Donnell, C., Hueber, W., Majka, D., Lazar, A., Derber, L., ... Holers, V. (2010). The number of elevated cytokines and chemokines in preclinical seropositive rheumatoid arthritis predicts time to diagnosis in an age-dependent manner. *Arthritis and Rheumatism*, *62*, 3161–3172.

Ekström, K., Hjalgrim, H., Brandt, L., Baecklund, E., Klareskog, L., Ekbom, A., & Askling, J. (2003). Risk of malignant lymphomas in patients with rheumatoid arthritis and in their first-degree relatives. *Arthritis and Rheumatism*, *48*, 963–970.

~~Hemminki, K., Försti, A., Sundquist, K., Sundquist, J., & Li, X. (2017). Familial associations of lymphoma and myeloma with autoimmune diseases. *Blood Cancer Journal*, *7*, e515–e515.~~

International Multiple Sclerosis Genetics Consortium, T. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, *476*, 214–219.

Khankhanian, P., Cozen, W., Himmelstein, D., Madireddy, L., Din, L., VanDen Berg, A., ... Hjalgrim, H. (2016). Meta-analysis of genome-wide association studies reveals genetic overlap between Hodgkin lymphoma and multiple sclerosis. *International Journal of Epidemiology*, *45*, 728–740.

Padyukov, L., Seielstad, M., Ong, R. T., Ding, B., Rönnelid, J., Seddighzadeh, M., ... group, E. S. (2011). A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Annals of the Rheumatic Diseases*, *70*(2), 259–265. 2.

Park, Y. -W., Kee, S. -J., Cho, Y. -N., Lee, E. -H., Lee, H. -Y., Kim, E. -M., ... Kang, H. -S. (2009). Impaired differentiation and cytotoxicity of natural killer cells in systemic lupus erythematosus. *Arthritis and rheumatism*, *60*, 1753–1763.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., ... Sham, P. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575.

R Core Team, R. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Skibola, C., Berndt, S., Vijai, J., Conde, L., Wang, Z., Yeager, M., ... Rothman, N. (2014). Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the HLA region. *American Journal of Human Genetics*, *95*, 462–471.

~~Thun, M., Linet, M., Cerhan, J., Haiman, C., & Schottenfeld, D. (2017). *Schottenfeld and Fraumeni Cancer Epidemiology and Prevention* (Fourth Edition ed.). Oxford University Press~~ Q6

Smedby, K., & Ponzoni, M. (2017). The aetiology of B-cell lymphoid malignancies with a focus on chronic inflammation and infections. *Journal of Internal Medicine*, *282*, 360–370

Vijai, J., Wang, Z., Berndt, S., Skibola, C., Slager, S., Bracci, P., ... Nieters, A. (2015). A genome-wide association study of marginal zone lymphoma shows association to the HLA region. *Nature Communications*, *6*, 5751.

Wellcome Trust Case Control Consortium, T. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*(7145), 661–678.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Din L, Sheikh M, Kosaraju N, et al. Genetic overlap between autoimmune diseases and non-Hodgkin lymphoma subtypes. *Genet. Epidemiol.* 2019;1–21. https://doi.org/10.1002/gepi.22242