Regular Article

LYMPHOID NEOPLASIA

Genome-wide association study implicates immune dysfunction in the development of Hodgkin lymphoma

Amit Sud,[1] Hauke Thomsen,[2] Giulia Orlando,[1] Asta Försti,[2,3] Philip J. Law,[1] Peter Broderick,[1] Rosie Cooke,[1] Fadi Hariri,[4] Tomi Pastinen,[4] Douglas F. Easton,[5,6] Paul D. P. Pharoah,[5,6] Alison M. Dunning,[5] Julian Peto,[7] Federico Canzian,[8] Rosalind Eeles,[1,9] ZSofia Kote-Jarai,[1] Kenneth Muir,[10,11] Nora Pashayan,[6,12] Daniele Campa,[13] the PRACTICAL Consortium, Per Hoffmann,[14-16] Markus M. Nöthen,[15,16] Karl-Heinz Jöckel,[17] Elke Pogge von Strandmann,[18] Anthony J. Swerdlow,[1,19] Andreas Engert,[20] Nick Orr,[19] Kari Hemminki,[2,3] and Richard S. Houlston[1]

[1]Division of Genetics and Epidemiology, The Institute of Cancer Research, London, United Kingdom; [2]Division of Molecular Genetic Epidemiology, German Cancer Research Center, Heidelberg, Germany; [3]Center for Primary Health Care Research, Lund University, Malmö, Sweden; [4]Department of Human Genetics, McGill University and Genome Quebec Innovation Centre, McGill University, Montreal, QC, Canada; [5]Centre for Cancer Genetic Epidemiology, Department of Oncology, and [6]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; [7]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom; [8]Genomic Epidemiology Group, German Cancer Research Center, Heidelberg, Germany; [9]Royal Marsden NHS Foundation Trust, London, United Kingdom; [10]Institute of Population Health, University of Manchester, Manchester, United Kingdom; [11]Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, United Kingdom; [12]Department of Applied Health Research, University College London, London, United Kingdom; [13]Department of Biology, University of Pisa, Pisa, Italy; [14]Human Genomic Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland; [15]Institute of Human Genetics and [16]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; [17]University of Duisburg–Essen, Essen, Germany; [18]Experimental Tumor Research, Center for Tumor Biology and Immunology, Clinic for Hematology, Oncology and Immunology, Philipps University, Marburg, Germany; [19]Division of Breast Cancer Research, The Institute of Cancer Research, London, United Kingdom; and [20]Department of Internal Medicine, University Hospital of Cologne, Cologne, Germany

KEY POINTS

- **Variation at 6p21.31, 6q23.3, 11q23.1, 16p11.2, and 20q13.12 influences risk for HL.**

- **Genetic predisposition implicates germinal center dysfunction, disrupted T-cell function, and NF-κB activation in the pathogenesis of HL.**

To further our understanding of inherited susceptibility to Hodgkin lymphoma (HL), we performed a meta-analysis of 7 genome-wide association studies totaling 5325 HL cases and 22 423 control patients. We identify 5 new HL risk loci at 6p21.31 (rs649775; $P = 2.11 \times 10^{-10}$), 6q23.3 (rs1002658; $P = 2.97 \times 10^{-8}$), 11q23.1 (rs7111520; $P = 1.44 \times 10^{-11}$), 16p11.2 (rs6565176; $P = 4.00 \times 10^{-8}$), and 20q13.12 (rs2425752; $P = 2.01 \times 10^{-8}$). Integration of gene expression, histone modification, and in situ promoter capture Hi-C data at the 5 new and 13 known risk loci implicates dysfunction of the germinal center reaction, disrupted T-cell differentiation and function, and constitutive NF-κB activation as mechanisms of predisposition. These data provide further insights into the genetic susceptibility and biology of HL. (_Blood_. 2018;132(19):2040-2052)

Introduction

Hodgkin lymphoma (HL) comprises classical HL (cHL; ~95% of cases) and nodular lymphocyte predominant HL (~5% of cases).[1] Although cHL and nodular lymphocyte-predominant HL are defined by the Hodgkin and Reed-Sternberg (HRS) cell and the lymphocyte-predominant cell, respectively, both diseases are thought to arise from the malignant transformation of germinal center B cell.[2,3] Furthermore, both cHL and nodular lymphocyte-predominant HL demonstrate a paucity of these neoplastic B cells within a background of reactive inflammatory cells that includes large populations of CD4[+] T cells.[4,5]

A viral or infectious agent has long been considered a major etiological factor for HL, with Epstein-Barr virus (EBV) being the posited infectious agent.[6,7] However, the EBV genome is only identifiable in a variable number of HL cases, and epidemiological data support a causal role for the virus in EBV-positive HL only.[8] Evidence for genetic susceptibility to HL is provided by the elevated familial risk, as well as the high concordance between monozygotic twins.[9,10] More recently, genome-wide association studies (GWAS) have confirmed an HLA association for HL and have identified single nucleotide polymorphisms (SNPs) at 13 non-HLA loci influencing risk.[11,12]

To gain further insight into HL susceptibility, we have conducted a meta-analysis of data from 7 independent GWAS and report 5 new HL risk loci.[11-13] Integration of gene expression, histone modification, and in situ promoter capture Hi-C data (PCHi-C) at the 5 new and the 13 known risk loci provides evidence for cell-type specificity in B and T cells and implicates dysfunction of the germinal center reaction, disrupted T-cell differentiation and function, and constitutive NF-κB activation as mechanisms by which loci influence HL risk.

Materials and methods

Ethics

Collection of patient samples and associated clinicopathological information was undertaken with written informed consent. Relevant ethical review boards approved the individual studies in

accordance with the tenets of the Declaration of Helsinki (UK-GWAS MREC 03/1/096, German-GWAS University of Heidelberg 104/2004, and UK-GWAS-NSHLG MREC 09/MRE00/72). The diagnosis of HL in all cases was established in accordance with World Health Organization guidelines.

GWAS

We used GWAS data generated on 3 nonoverlapping case–control series of Northern European ancestry, which have been the subject of previous analyses (supplemental Tables 1 and 2, available on the *Blood* Web site).[11] The UK-GWAS was based on 622 cases ascertained through the Royal Marsden Hospital National Health Service Trust Family History study during 2004 to 2008,[14] and 5677 control patients from the UK Wellcome Trust Case Control Consortium 2.[15] The German-GWAS comprised 1001 cases ascertained by the German Hodgkin Study Group during 1998 to 2007, and 2092 control patients from the Heinz Nixdorf Recall study.[16] The UK-NSHLG-GWAS used 1717 cases ascertained through the National Study of Hodgkin Lymphoma Genetics (https://www.ukctg.nihr.ac.uk) from 2010 to 2013.[11] Control patients comprised 2976 cancer-free men recruited by the PRACTICAL Consortium, the UK Genetic Prostate Cancer Study (UKGPCS; age <65 years), a study conducted through the Royal Marsden NHS Foundation Trust and Study of Epidemiology & Risk Factors in Cancer, recruited via general practitioner practices in East Anglia (2003-2009); and 4446 cancer-free women from across the United Kingdom via the Breast Cancer Association Consortium. Details of the genotyping platform and quality control measures applied to each of the 3 GWAS have been described previously and are detailed in supplemental Tables 3 and 4.[11,14,17,18] Briefly, individuals with a low call rate (<95%), as well as all individuals evaluated to be of non-European ancestry, were excluded (supplemental Figure 1). Eigenvectors for the GWAS data sets were inferred using smartpca (part of EIGENSOFT) by merging cases and controls with phase 3 HapMap samples.[19] For apparent first-degree relative pairs, we excluded the control from a case–control pair or the individual with the lower call rate (supplemental Table 3). SNPs with a call rate below 95% were excluded, as were those with a minor allele frequency <0.01 or displaying deviation from Hardy–Weinberg equilibrium (ie, $P < 10^{-6}$; supplemental Table 4). GWAS data were phased with SHAPEIT3,[20] and imputed to more than 10 million SNP, using IMPUTE4 v1.0[21] and a merged reference panel consisting of data from 1000 Genomes Project (phase 3)[22] and UK10K (EGAD00001000776).[23] Imputation was conducted separately for each study from a set of SNPs common to cases and controls. Poorly imputed SNPs (defined by an information measure < 0.80) were excluded. Tests of association between SNPs and HL were performed using logistic regression under an additive genetic model in SNPTESTv2.5.2.[24] The adequacy of the case–control matching was evaluated using Q–Q plots of test statistics (supplemental Figure 2). The inflation factor λ_{1000} was based on the 90% least-significant SNP scaled to 1000 cases and 1000 controls.

In addition to analyzing data from these 3 GWAS, we made use of preprocessed association test statistics for HL risk from a meta-analysis of 3 additional GWAS (USC-IARC-UC-GWAS) comprising 1816 HL cases and 7879 control individuals,[12,25,26] and an analysis of 432 HL cases and 337 208 unaffected individuals[13] from the UK Biobank, accessed through the Global Biobank Engine.

Meta-analysis

Meta-analyses were performed under a fixed-effects model, using META v1.7.[24] Cochran's Q-statistic to test for heterogeneity and the I^2 statistic to quantify the proportion of the total variation resulting from heterogeneity were calculated; an I^2 value at least 75% is considered to be characteristic of large heterogeneity.[27] Where the same controls were used in both the UK-GWAS and the USC-IARC-UC GWAS, these controls were excluded from the UK-GWAS association analysis.

Cell culture

L-428 HL cells were obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen and were cultured at 37°C in RPMI 1640 supplemented with 10% heat-inactivated fetal bovine serum (Thermo Fisher Scientific). Cell line identity was confirmed by STR-profiling. Cells were regularly tested for mycoplasma contamination (PromoCell, PK-CA91).

Chromatin immunoprecipitation-seq analysis

L-428 Chromatin immunoprecipitation (ChIP)-seq was performed on H3K27Ac and H3K4me3 for using antibodies obtained from Diagenode. Briefly, after cell lysing, sonication of nuclei was performed (UCD-300, BioRuptor) to obtain 150- to 500-bp fragments. ChIP reaction was performed on a Diagenode SX-8G IP-Star Compact using Diagenode automated Ideal Kit reagents (C01010011). Protein A beads were incubated for 10 hours with 3 to 6 μg antibody and 2 to 4 million sonicated cell lysate. ChIP samples were de-cross-linked at 65°C for 4 hours and subsequently treated for 30 minutes with RNAse Cocktail and proteinase K. DNA was then purified (MiniElute PCR purification kit, Qiagen), followed by library preparation according to manufacture (HTP Illumina library preparation kit, KAPA Biosystems). Fourteen cycles of polymerase chain reaction were performed, followed by size selection for 200- to 400-bp fragments and final library purification (GeneRead Size Selection kit, 301 Qiagen). ChIP libraries were sequenced using HiSeq 2000 (Illumina) with 100-bp single-ended reads. Generated raw reads were filtered for quality (Phred33 ≥ 30) and length (n ≥ 32), and adapter sequences were removed using Trimmomatic v0.2235. Reads passing filters were then aligned to the human reference (hg19), using BWA v0.6.1. Peak calls are obtained using MACS2 v 2.0.1.

Histone modification data from primary blood cells H3K27Ac and H3K4me3 data from more than 100 samples from more than 30 cell types from the Blueprint Epigenome Consortium were analyzed.[28]

Cell-specificity analysis

Overlap enrichment analysis of HL risk SNPs with H3K4me3 and H3K27Ac peaks was performed as described by Trynka et al.[29] Briefly, we evaluated whether the HL risk SNPs and SNPs in linkage disequilibrium (LD; $r^2 > 0.8$) with the sentinel SNP were enriched at H3K4me3 and H3K27Ac ChIP-seq peaks in blood cells and the HRS cell line L-428 by a permutation procedure with 10^5 iterations.

Promoter capture Hi-C

In situ Hi-C libraries for L-428 were prepared as previously described.[30,31] Briefly, 25 million cells were fixed in 1% formaldehyde for 10 minutes. Cross-linked DNA was digested with *Hind*III (NEB, R0104), and chromatin ends were filled and marked

with biotin-14-dATP (Thermo Fisher, 19524-016). The resulting blunted ended fragments were ligated at 16°C in the nucleus with T4 DNA ligase (NEB, M0202) to minimize random ligation. DNA purified after cross-linking was reversed by proteinase K (Ambion, AM2546) treatment. DNA was sheared by sonication (Covaris, M220), and 200- to 650-bp fragments were selected. Biotin tag DNA was pulled down with streptavidin beads and ligated with Illumina paired-end adapters (Illumina). Six cycles of polymerase chain reaction were performed to amplify libraries before capture. Promoter capture was based on 32 313 biotinylated 120-mer RNA baits (Agilent Technologies), targeting both ends of HindIII restriction fragments that overlap Ensembl promoters of protein-coding, noncoding, antisense, snRNA, miRNA, and snoRNA transcripts (supplemental Data). After library enrichment, a postcapture polymerase chain reaction amplification step was carried out using 6 amplification cycles. Hi-C and PCHi-C libraries were sequenced using HiSeq 2000 technology (Illumina). Reads were aligned to the GRCh37 build using Bowtie2 v2.2.6,[32] and identification of valid di-tags was performed using HiCUP v0.5.9.[33] To declare significant contacts, HiCUP output was processed using CHiCAGO v1.1.8.[34] Data from 3 independent biological replicates were combined to define a definitive set of contacts. Publicly accessible PCHi-C data generated in B- and T-cell populations were downloaded from the Open Science Framework.[35]

Chromatin interactions relevant to HL risk loci were defined as contacts overlapping with HL risk SNPs and SNPs in LD ($r^2 > 0.8$ with the sentinel SNP), with promoters within a 2-Mb window of the sentinel SNP, and with a score of at least 5.0.[34] Plotting of HL association data and chromatin contacts was performed using visPIG.[36]

Expression quantitative trait loci analysis

An analysis of associations between the SNPs ($r^2 > 0.8$) at each locus and tissue-specific changes in gene expression was performed using summary statistics from 3 publicly available resources: lymphoblastoid cell line (LCL) expression from the MuTHER (n = 825) consortium[37]; LCL expression from the GTEx consortium (n = 114)[38]; and CD4[+] and CD8[+] T cells from 313 individuals.[39] Statistical significance was assigned after correcting for the number of probes at each locus (microarray) or the number of transcripts at each locus (RNA-seq) for each expression data set.

Genetic correlation with infection

To estimate the genetic correlation between specific infections and all HL, and nodular sclerosis HL (NSHL) and mixed cellularity HL (MCHL) subtypes,[40] we used LD score regression. Summary statistics for self-reported infectious diseases from more than 200 000 participants in 23andMe included[41] chickenpox, shingles, cold sores, mononucleosis, mumps, hepatitis B, plantar warts, positive tuberculosis test results, Streptococcus throat infection, scarlet fever, pneumonia, bacterial meningitis, yeast infections, urinary tract infections, tonsillectomy, childhood ear infections, myringotomy, measles, hepatitis A, rheumatic fever, common colds, rubella, and chronic sinus infection.

Mendelian randomization

We performed 2-sample Mendelian randomization using SNPs associated with specific infection-related traits (ie, $P \leq 5.0 \times 10^{-8}$) as instrumental variables.[41] We analyzed infection-related traits for

which more than 2 SNPs had been shown to be associated with the specific infection (tonsillectomy, mumps infection, childhood ear infection, and yeast infections). To avoid colinearity between SNPs for each trait, we excluded SNPs that were correlated (ie, $r^2 \geq 0.01$) within each trait and only considered the SNPs with the strongest effect on the trait for use as instrumental variables. Where data on an instrumental variable was not present in the outcome trait, a proxy was used ($r^2 > 0.6$). Details of the instrumental variables used are detailed in supplemental Data. For each SNP, we recovered the chromosome position, risk allele, association estimates (per allele log-odds ratio [OR]), and standard errors. The allele that was associated with increased risk for the exposure was considered the effect allele. The ORs of HL, NSHL, and MCHL per unit of standard deviation increment for each infection-related trait were estimated using the Mendelian randomization R package.[42] Given that traits analyzed are binary outcomes, the maximum likelihood method was employed with the resulting causal effect estimate representing the odds for HL risk per unit increase in the log-OR for infection-related trait.

## Results

Association analysis

We analyzed summary level GWAS data generated on HL cases and controls of European ancestry[11] from 3 sources (supplemental Tables 1-4): 2 GWAS of UK cases and controls and 1 GWAS of German cases and controls, totaling 3077 cases and 14 546 control patients (Discovery GWAS)[11]; the Stanford Global Biobank Engine, an analysis of 432 HL cases from the UK Biobank[13]; and a meta-analysis of 3 published HL GWAS totaling 1816 HL cases and 7879 control patients (USC-IARC-UC-GWAS).[12,25,26]

In a meta-analysis of data from the 7 studies, we identified new genome-wide significant associations for HL (Figure 1; Table 1) at 6p21.31 (rs649775; $P = 2.11 \times 10^{-10}$, marking ITPR3-UQCC2-IP6K3), 6q23.3 (rs1002658; $P = 2.97 \times 10^{-8}$, marking OLIG3-TNFAIP3), 11q23.1 (rs7111520; $P = 1.44 \times 10^{-11}$, marking POU2AF1), 16p11.2 (rs6565176; $P = 4.00 \times 10^{-8}$, marking MAPK3-CORO1A), and 20q13.13 (rs2425752; $P = 2.01 \times 10^{-8}$, marking NCOA5-CD40). In addition, we identified a promising association at 1p13.2 (rs2476601; $P = 4.20 \times 10^{-7}$, marking PTPN22).

The bimodal incidence of HL and the higher rate of NSHL and EBV-negative HL in young adults suggest differences in the etiology of HL subtypes.[8] Case-only analysis, however, provided no evidence for an age or histological subtype association for the 5 new-risk SNPs. (supplemental Tables 5 and 6).

Cell specificity of associations

Trynka et al have recently shown that chromatin marks highlighting regulatory regions overlap with phenotype-associated variants in a cell-type-specific manner.[29] To examine for cell-type specificity of the 5 new and 13 known HL risk loci, we analyzed H3K4me3 and H3K27Ac chromatin marks, which annotate regulatory regions, in more than 125 samples from 38 hematopoietic cell types from BLUEPRINT[28,29] and the HRS cell line L-428. The H3K27Ac histone mark is predominantly associated with enhancers, and of all the histone marks, it demonstrates the greatest enrichment of promoter-interacting regions.[35] The H3K4me3

SUD et al
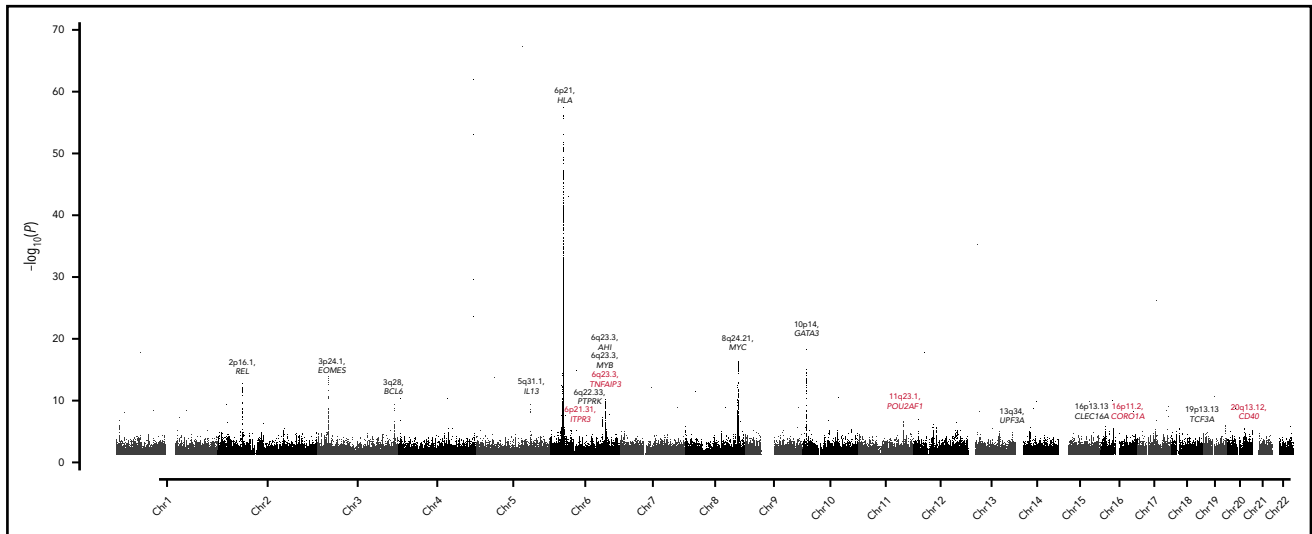
Figure 1. Genome-wide meta-analysis P values of Hodgkin's lymphoma risk (−log₁₀P, y-axis) plotted against their chromosomal positions (x-axis). Novel HL risk loci and candidate gene are in orange.

histone mark is predominantly associated with promoters and transcribed regions and has previously been shown to be the most phenotypically cell-type-specific.[29,43] Cell types showing the strongest enrichment of risk SNPs at H3K4me3 marks were CD4+ T cells from venous blood ($P = 2.9 \times 10^{-3}$), CD3−CD4+CD8+-positive thymocytes ($P = 5.7 \times 10^{-3}$), and tonsillar-derived germinal center B cell ($P = 6.3 \times 10^{-3}$; supplemental Table 7). Cell types with the strongest enrichment of risk SNPs at H3K27Ac marks were CD8+ T cells from venous blood ($P = 3.0 \times 10^{-4}$), CD3+CD4+CD8+ thymocytes ($P = 5.6 \times 10^{-4}$), CD4+ thymocytes ($P = 2.7 \times 10^{-3}$), and L-428 ($P = 7.9 \times 10^{-3}$; supplemental Table 8). Based on the colocalization of variants with active chromatin marks, we calculated an enrichment score for each genetic association (Figure 2).[29] High SNP regulatory scores were also shown in T cells at 3p24.1, 6q22.33, 6q23.3, and 10p14 risk loci; in B cells at 2p16.1, 3q28, 8q24.21, 11q23.1, and 20q13.12 risk loci; and in HRS cells at 3p24.1, 5q31.1, 6q22.33, 6q23.3, 10p14, 13q34 16p13.13, and 20q13.12.

Identification of candidate target genes at HL risk loci

Most GWAS loci map to noncoding regions of the genome and influence gene regulation.[44] Hence, to gain insight into the biological mechanisms for the associations at the 5 new and 13 known HL risk loci, we first performed expression quantitative trait locus (eQTL) analysis on expression data in B-cell LCLs and in CD4+ and CD8+ T cells. We identified eQTLs in LCLs at 6p21.31 (*ITPR3*), 6q23.3 (*AHI1*, *ALDH8A1*), 10p14 (*GATA3*), 11q23.1 (*COLCA1*, *COLCA2*), 13q34 (*UPF3A*, *CDC16*), 16p13.13 (*SOCS1*), 16p11.2 (*MAPK3*, *BOLA2*), and 20q13.12 (*WFDC10B*); in CD4+ T cells at 6q23.3 (*AHI1*) and 13q34 (*CDC16*); and in CD8+ T cell at 3p24.1 (*EOMES*), 6q23.3 (*AHI1*), and 13q34 (*CDC16*; Table 2; supplemental Tables 9 and 10).

Chromatin looping interactions between enhancer elements and promotors are central to regulation of gene expression.[45] To link risk loci to candidate target genes, we analyzed PCHi-C data. First, we examined physical interactions at genomic regions annotated by HL risk loci (including variants with an $r^2 > 0.8$),

using publicly accessible PCHi-C in naive and total B cells, as well as CD4+ and CD8+ T cells.[35] Second, we generated and analyzed PCHi-C data for the HRS cell line L-428. We observed concordance between H3K27Ac peaks and chromatin contacts in B, T, and HRS cells for specific HL risk loci. Notable chromatin contacts were found in the B lineage at 2p16.1 (*REL*), 6p21.31 (*BAK1*), 8q24.21 (*MYC*, *PVT1*), 13q34 (*RASA3*), 16p.13.13 (*RMI2*), and 20q13.12 (*CD40*); in the T-lineage at 3p24.1 (*EOMES*, *AZI2*), 6p21.31 (*BAK1*), 6q22.33 (*THEMIS*, *PTPRK*), 6q23.3 (*MYB*), 13q34 (*RASA3*), and 16p13.13 (*SOCS1*, *RMI2*); and in L-428 at 3p24.1 (*AZI2*, *CMC1*), 6q23.3 (*MYB*), 6q23.3 (*TNFAIP3*), and 16p13.13 (*SOCS1*, *RMI2*; Table 2; supplemental Figure 3; supplemental Data).

Shared susceptibility with infection

The association between EBV with HL, coupled with epidemiological reports of HL also being associated with non-EBV infections,[46-49] suggests shared susceptibility a priori. Support for such an assertion is provided by a recent report implicating a number of the HL loci, including 6q23.3, 16p11.2, and 20q13.12, as well as the HLA region, as determinants of risk for infection.[41]

To investigate coheritability between HL and susceptibility to infection, we implemented cross-trait LD score regression.[40] Using summary-level GWAS data, we estimated genetic correlations between HL and more than 20 self-reported infections in 200 000 23andMe participants.[41] Overall, no statistically significant correlation was shown between any specific infection and HL, NSHL, or mixed cellularity HL (supplemental Table 11). Following on from this, for infections with greater than 2 genetically defined instrumental variables, we performed a Mendelian randomization analysis to identify a potential causal relationship with HL. For tonsillectomy, yeast infections, and childhood ear infections, no statistically significant associations were demonstrated (supplemental Table 12). A nominally significant positive association between self-reported mumps infection and HL was found ($P = .04$); however, this was not significant after correction for multiple testing.

Table 1. Summary results for newly identified Hodgkin lymphoma risk loci

Locus	Risk Nearest genes*	Risk Allele (freq)	Risk Position (hg19, bp)	Discovery GWAS meta-analysis P	Discovery GWAS meta-analysis OR (95% CI)	UK Biobank P	UK Biobank OR (95% CI)	USC-IARC-UC-GWAS P	USC-IARC-UC-GWAS OR (95% CI)	Meta-analysis P	Meta-analysis OR (95% CI)	Meta-analysis I², %	Meta-analysis P_{het}
1p13.2, rs2476601	PTPN22	A (0.12)	114377568	3.92×10^{-3}	1.15 (1.04-1.26)	3.21×10^{-4}	1.42 (1.17-1.72)	3.70×10^{-3}	1.24 (1.07-1.44)†	4.20×10^{-7}	1.21 (1.12-1.30)	20	.29
6p21.31, rs649775	ITPR3-UQCC2-IP6K3	A (0.11)	33684313	4.00×10^{-6}	1.25 (1.14-1.38)	—	—	8.22×10^{-6}	1.36 (1.19-1.55)	2.11×10^{-10}	1.29 (1.19-1.40)	0	.54
6q23.3, rs1002658	OLIG3-TNFAIP3	T (0.18)	137981584	3.86×10^{-6}	1.19 (1.11-1.28)	—	—	2.15×10^{-3}	1.18 (1.06-1.31)	2.97×10^{-8}	1.19 (1.12-1.26)	0	.53
11q23.1, rs7111520	POU2AF1	A (0.70)	111249611	4.33×10^{-7}	1.17 (1.10-1.24)	—	—	4.39×10^{-6}	1.24 (1.13-1.35)	1.44×10^{-11}	1.19 (1.13-1.25)	0	.68
16p11.2, rs565176	MAPK3-CORO1A	T (0.48)	30174926	8.64×10^{-6}	1.14 (1.08-1.21)	3.44×10^{-4}	1.28 (1.10-1.23)	—	—	4.00×10^{-8}	1.16 (1.10-1.22)	0	.46
20q13.12, rs2425752	NCOA5-CD40	T (0.23)	44702120	2.23×10^{-4}	1.13 (1.06-1.20)	2.94×10^{-4}	1.30 (1.12-1.50)	3.77×10^{-3}	1.14 (1.09-1.20)	2.01×10^{-8}	1.15 (1.10-1.21)	56	.06

freq, frequency; I², proportion of the total variation resulting from heterogeneity.

*Genes at each risk locus are given for identification purposes only and does not necessarily indicate biological functionality.

†Summary statistics from 1200 patients with cHL and 6417 control patients.[26]

SUD et al

Figure 2. Heat map of SNP scores for H3K27Ac and H3K4me3 at each Hodgkin lymphoma risk locus. SNP score calculated as per Trynka et al.[29] For each SNP at a given locus, the score represents the height of the closest ChIP-seq peak divided by the distance to the summit in each cell line, normalized across all immune cell types. Thus, a SNP within a chromatin mark that is active in only 1 cell type will have a high score of 1 (red) in that cell type and 0 (white) in others. In contrast, a SNP close to chromatin marks that are not cell-type-specific will have similarly modest scores across cell types. Genes at each risk locus are given for identification purposes only, and do not necessarily indicate biological functionality.

Discussion

By using publicly available summary statistics, we have increased the power of our study, allowing us to identify 5 new HL risk loci, thus bringing the total number of HL risk loci to 18. Although our reliance on such data has restrained our ability to examine subtype-specific effects, it is likely that the newly described risk loci have generic effects on HL susceptibility, as with the known risk loci at 5q31.1 and 19p13.3.[11]

At the new and known HL risk loci, we observed an enrichment of active regulatory regions in germinal center B cells, CD4[+] thymocytes, CD4[+] T cells, and CD8[+] T cells. Furthermore, although some HL risk loci locate to H3K27Ac peaks in both B and T cells, a number display lineage specificity. Motivated by this finding, we have used PCHi-C and gene expression data in these cell types to identify targets subject to regulatory control by HL risk SNPs. Although in part speculative, and requiring functional validation, integrating proximity, cell specificity of risk loci, gene expression, and PCHi-C data, our analyses highlight 3 biological processes and their associated genes as a basis of HL susceptibility (Table 2): the germinal center reaction (2p16.1, REL[50]; 3q28, BCL6 and mir-28[51,52]; 6p21, HLA[53]; 6q23.3, MYB[54]; 8q24.21, MYC[55]; 11q23.1, POU2AF1[56]; 16p11.2, MAPK3[57]; 19p13.3, TCF3[58]; 20q13.12, CD40),[59,60] T-cell differentiation and function (3p24.1, EOMES[61]; 5q31,1, IL13[62]; 6q22.33, PTPRK and THEMIS[63,64]; 6q23.3, MYB[65]; 6q23.3, AHI1[66]; 10p14, GATA3[67]; 16p13.1, SOCS1 and CLEC16A[68,69]; 16p11.2, MAPK3 and CORO1A[70,71]), and constitutive NF-κB activation (2p16.1, REL[72]; 3p24.1, AZI2[73]; 6q23.3, TNFAIP3[74]; 20q13.12, CD40[75,76]).

Our findings extend the relationship between germline genetics and tumor biology,[44] as evidenced by enrichment of active chromatin marks for HL risk loci in L-428, and the findings of many of the target genes for HL GWAS associations are subject to somatic alterations in HRS cells; namely, REL,[77] TNFAIP3, and SOCS1.[78-80] The composite cellular basis of the HL tumor represents a preeminent example of the importance of the cellular microenvironment for the development of cancer. Hence, it is entirely conceivable that some of the HL risk loci may affect the development of the B-cell tumor indirectly. Support for such an assertion is the observation of T-cell specificity, as well as the finding of an eQTL at 3p24.1 (EOMES) in CD8[+] T cells. Notably, Eomes[Hi] T-bet[Lo] PD-1[Hi]CD8[+] T cells are considered to delineate a key subset of exhausted CD8[+] T cells,[81,82] which may contribute to an immunosuppressive tumor microenvironment and is a feature of peripheral blood T cells in HL.[83]

There are a number of reasons for the observed lack of concordance between the PCHi-C and eQTL analysis at risk loci. First, the resolution of the Hi-C library using HindIII, a 6-base pair cutter, is approximately 10 kb. As such, we are unable to detect concordant chromatin contacts at risk loci, which influence the expression of genes located less than 10 kb. Second, it is recognized that the range at which gene expression is perturbed to influence disease risk may be narrow, and as such, it may not be detected by an eQTL analysis. Finally, given the risk loci are likely to act in specific cell populations, and our expression data are limited by broad B- and T-cell populations, it is possible that we have not captured the cell type to analyze expression. As such, we would view both methods as complimentary in identifying target genes.

The established association between EBV and risk for HL, coupled with other epidemiological observations, provides strong a priori evidence for infection being a major etiological risk factor for HL. Although our Mendelian randomization analysis did not implicate a causal relationship with any of the self-reported infection traits, we acknowledge that our study

Table 2. Integration of expression quantitative trait loci, histone modification, promoter capture Hi-C data at non-HLA Hodgkin lymphoma risk loci to identify candidate causal genes at Hodgkin lymphoma risk loci

Locus	Sentinel SNP	Position, (bp, hg19)	Gene(s) in LD block	Coding variant(s)	Promotor/ UTR variant(s)	Expression quantitative trait loci in LCLs	Expression quantitative trait loci in T cell	H3K27Ac histone peak*	Hi-C contact(s) in naive or total B cells	Hi-C contact(s) in T cells	Hi-C contact(s) in HRS cell	Evidence of perturbation in HL	Candidate biological mechanism
2p16.1	rs2420518	61054980						Naive B cell	REL	REL		REL[84]	Constitutive NF-κB activation (REL)[72], Altered B-cell differentiation and germinal center reaction (REL)[50]
3p24.1	rs3806624	27764623	EOMES		EOMES, (3'-UTR)		EOMES (↑) (CD8⁺)	Effector memory CD8⁺ T cell, plasma cell, L-428	AZI2, CMC1, NEK10, OXSM, NGLY1, ZCWPW2	EOMES, AZI2, CMC1, NEK10, OXSM, NGLY1, ZCWPW2	AZI2, CMC1,	EOMES[85],	Exhausted CD8 T-cell phenotype (EOMES)[61,81,82], Constitutive NF-κB activation (AZI2),[73]
3q28	rs4459895	187954414	LPP					CD38⁻ naive B cell, naive B cell, germinal center B cell, L-428				BCL6[86]	Dysfunction of B-cell germinal center reaction (BCL6, mir-28),[51,52]
5q31.1	rs848	131996500	IL-13	(p.Gln144Arg)	IL-13, (3'-UTR)			L-428				IL-13[87]	Altered CD4⁺ T-cell function (IL-13)[62]
6p21.31	rs649775	33684313	ITPR3, UQCC2, IP6K3		IP6K3, (3'-UTR)	ITPR3 (↓)		CD4⁺ T cell, CD8⁺ T cell, effector memory CD8⁺ T cell, naive B cell, class-switched memory B cell	BAK1, SYNGAP1, GGNBP1, LINC00336	BAK1, GRM4, SYNGAP1, KIFC1, CUTA, PHF1, GGNBP1, LINC00336			Altered B-cell differentiation (ITPR3)[98]
6q22.33	rs9482849	128288536	PTPRK					CD4⁺ T cell, central memory CD4⁺ T cell, CD8⁺ T cell, effector memory CD8⁺ T cell, L-428	PTPRK, THEMIS			PTPRK[89]	Altered T-cell differentiation (PTPRK, THEMIS),[63,64]
6q23.3	rs9402684	135419305	HBS1L					CD3⁻CD4⁻CD8⁺ thymocyte, CD3⁺CD4⁺CD8⁺ thymocyte, CD4⁺ T cell, central memory CD4⁺ T cell, CD8⁺ T cell, effector memory CD8⁺ T cell, germinal center B cell, plasma cell, L-428	MYB,	MYB,	MYB,		Altered T-cell differentiation (MYB)[65,90], Altered B-cell differentiation and germinal center reaction (MYB),[54]
6q23.3	rs6928977	135626348	AHI1			AHI1 (↑), ALDH8A1 (↑)	AHI1 (CD4⁺ and CD8⁺) (↑)	CD3⁺CD4⁺CD8⁺ thymocyte					Altered T-cell differentiation (AHI1)[66]
6q23.3	rs1002658	137981584						L-428	RP11-204P2.3		TNFAIP3,	TNFAIP3[3,78]	Constitutive NF-κB activation (TNFAIP3)[74]
8q24.21	rs34748721	129195943						naive B cell, class switched memory B cell	CASC11, MYC, PVT1, RNU1-106P, MIR1207			MYC[91]	Dysfunction of B-cell germinal center reaction (MYC)[55,76]

*SNPs (r² < 2.5 kilobases from ChIP-seq peak).

SUD et al

Table 2. (continued)

Locus	Sentinel SNP	Position, (bp, hg19)	Gene(s) in LD block	Coding variant(s)	Promotor/ UTR variant(s)	Expression quantitative trait loci in LCLs	Expression quantitative trait loci in T cell	H3K27Ac histone peak*	Hi-C contact(s) in naïve or total B cells	Hi-C contact(s) in T cells	Hi-C contact(s) in HRS cell	Evidence of perturbation in HL	Candidate biological mechanism
10p14	rs2388486	8099021	GATA3			GATA3(↓)		CD3⁻CD4⁺CD8⁺ thymocyte, CD3⁺CD4⁺CD8⁺ thymocyte, CD4⁺ T cell, CD8⁺ T cell, effector memory CD8⁺ T cell, germinal center B cell				GATA3[92]	Altered T-cell differentiation (GATA3)[6,7]
10p14	rs3781093	8101927	GATA3			GATA3 (↓)		CD4⁺ T cell, CD8⁺ T cell				GATA3	Altered CD4⁺ T-cell differentiation (GATA3)[6,7], Altered B-cell differentiation (GATA3)[93]
11q23.1	rs7111520	111249611	POU2AF1			COLCA1 (↑), COLCA2 (↑)		CD4⁺ T cell, central memory CD4⁺ T cell, CD8⁺ T cell, CD38⁻ B cell, CD38⁻ naïve B cell, naïve B cell, germinal center B cell, unswitched memory B cell, class switched memory B cell, plasma cell	FDX1	FDX1, PPP3R1B, ALG9, FDXACB1, DIXDC1		POU2AF1[94]	Dysfunctional germinal center reaction (POU2AF1)[5,6]
13q34	rs112998813	115059729	UPF3A			CDC16 (↑), UPF3A (↓)	CDC16 (CD4⁺ and CD8⁺) †	CD4⁺ T cell, central memory CD4⁺ T cell, CD8⁺ T cell, effector memory CD8⁺ T cell, CD38⁻ naïve B cell, naïve B cell, germinal center B cell, unswitched memory B cell, class-switched memory B cell, plasma cell		RASA3, TMEM255B,, GASA6		CDC16,[85]	Disrupted cell cycle regulation (CDC16)[95], Dysfunction of mRNA surveillance (UPF3A)[96]
16p13.13	rs34972832	11198938	CLEC16A			SOCS1 (↑)		CD4⁺ T cell, central memory CD4⁺ T cell, CD8⁺ T cell, effector memory CD8⁺ T cell, naïve B cell, germinal center B cell, class switched memory B cell, plasma cell, L-424	RMI2	SOCS1, RMI2, PRM2,, PRM3,, TNP2, HNRNPCP4	SOCS1, RMI2	SOCS1[80]	T-cell dysfunction (SOCS1)[98], Altered T-cell differentiation (CLEC16A)[69], B-cell dysfunction (CLEC16A)[97], Genomic instability (RMI2)[98]
16p11.2	rs6565176	30174926	CORO1A		CORO1A (5′-UTR)	MAPK3 (↓), BOLA2 (↓),		CD4⁺ T cell, CD8⁺ T cell, effector memory CD8⁺ T cell, naïve B cell, class-switched memory B cell				MAPK3,[99] CORO1A[100]	T-cell dysfunction (CORO1A and MAPK3)[33,70,71], Dysfunction of B-cell germinal center reaction (MAPK3)[7,101,102]
19p13.3	rs2012125	1630341	TCF3A					CD38⁻ B cell, naïve B cell, class-switched memory B cell					Dysfunction of B-cell germinal center reaction (TCF3A)[103]
20q13.12	rs2425752	44702120	NCOA5, CD40			WFDC10B (↑)		Central memory CD4⁺ T cell, CD8⁺ T cell, effector memory CD8⁺ T cell, naïve B cell, germinal center B cell, L-428	CD40	TP53RK		CD40[104]	Dysfunctional germinal center reaction (CD40)[93,60], Constitutive NF-κB activation (CD40)[75,76]

*SNPs (r^2 < 2.5 kilobases from ChIP-seq peak).

had limited power. It is, however, possible that pleiotropism between the 6p21.1, 6q23.3, 16p11.2, and 20q13.12 risk loci for HL and tonsillectomy is consistent with some form of a shared biological basis. This is intriguing, as tonsillectomy has previously been linked to HL in some epidemiological observational studies.[46]

In summary, our study provides further evidence for inherited susceptibility to HL and support for cell-type specificity at HL risk loci. Furthermore, through the integration of gene expression, histone modification and in situ PCHi-C data, our data highlight dysfunction of the germinal center reaction, perturbed T-cell function, and constitutive NF-κB activation as mechanisms by which genetic risk loci influence HL pathogenesis.

Authorship

Contribution: A.S., K.H., and R.S.H. designed and provided overall project management; A.S. and R.S.H. drafted the manuscript; A.J.S., N.O., and R.S.H. provided samples for UK-GWAS and UK-NSHLG-GWAS;

Footnotes

Appendix: study group members

The members of the PRACTICAL Consortium (http://practical.icr.ac.uk/) are: OncoArray (in addition to those named in the author list): Brian E. Henderson (University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA), Christopher A. Haiman (University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA), Sara Benlloch (The Institute of Cancer Research, London, United Kingdom/University of Cambridge, Cambridge, United Kingdom), Fredrick R. Schumacher (Case Western Reserve University, Cleveland, OH/Seidman Cancer Center, University Hospitals, Cleveland, OH), Ali Amin Al Olama (University of Cambridge, Cambridge, United Kingdom), Sonja I. Berndt (National Institutes of Health, Bethesda, MD), David V. Conti (University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA), Fredrik Wiklund (Karolinska Institute, Stockholm, Sweden), Stephen Chanock (National Institutes of Health, Bethesda, MD), Victoria L. Stevens (American Cancer Society, Atlanta, GA), Catherine M. Tangen (Fred Hutchinson Cancer Research Center, Seattle, WA), Jyotsna Batra (Queensland University of Technology/Translational Research Institute, Brisbane, Queensland, Australia), APCB BioResource (Queensland University of Technology, Brisbane, Queensland, Australia), Judith Clements (Queensland University of Technology/Translational Research Institute, Brisbane, Queensland, Australia), Henrik Gronberg (National Institutes of Health, Bethesda, MD), Johanna Schleutker (Translational Research Institute, Brisbane, Queensland, Australia/University of Turku/Turku University Hospital, Turku, Finland), Demetrius Albanes (University of Cambridge, Cambridge, United Kingdom), Stephanie Weinstein (University of Cambridge, Cambridge, United Kingdom), Alicja Wolk (Karolinska Institutet, Stockholm, Sweden), Catharine West (University of Manchester, Manchester, United Kingdom), Lorelei Mucci (Harvard School of Public Health, Boston, MA), Géraldine Cancel-Tassin (Tenon Hospital/Sorbonne Université, Paris, France), Stella Koutros (National Institutes of Health, Bethesda, MD), Karina Dalsgaard Sorensen (Aarhus University Hospital/Aarhus University, Aarhus C, Denmark), Eli Marie Grindedal (Oslo University Hospital, Oslo, Norway), David E. Neal (University of Cambridge/Addenbrooke's Hospital/Cancer Research UK, Cambridge, United Kingdom), Ruth C. Travis (University of Oxford, Oxford, United Kingdom), Robert J. Hamilton (Princess Margaret Cancer Centre, Toronto, Canada), Sue Ann Ingles (University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA), Barry Rosenstein (Icahn School of Medicine at Mount Sinai, New York, NY), Yong-Jie Lu (Queen Mary University of London, London, United Kingdom), Graham G. Giles (The Cancer Council Victoria/The University of Melbourne, Melbourne, Victoria, Australia), Adam S. Kibel (Brigham and Womens Hospital, Boston, MA), Ana Vega (Fundación Pública Galega de Medicina Xenómica, Santiago de Compostela, Spain), Manolis Kogevinas (Barcelona Institute for Global Health, Barcelona, Spain/CIBER Epidemiología y Salud Pública, Madrid, Spain/Hospital del Mar Research Institute, Barcelona, Spain/Universitat Pompeu Fabra, Barcelona, Spain), Kathryn L. Penney (Brigham and Women's Hospital/Harvard Medical School, Boston, MA), Jong Y. Park (Moffitt Cancer Center, Tampa, FL), Janet L. Stanford (Fred Hutchinson Cancer Research Center/University of Washington, Seattle, WA), Cezary Cybulski (Pomeranian Medical University, Szczecin, Poland), Børge G. Nordestgaard (University of Copenhagen, Copenhagen, Denmark/Copenhagen University Hospital, Herlev, Denmark), Hermann Brenner (German Cancer Research Center/National Center for Tumor Diseases, Heidelberg, Germany), Christiane Maier (University Hospital Ulm, Ulm, Germany), Jeri Kim (The University of Texas M. D. Anderson Cancer Center, Houston, TX), Esther M. John (Cancer Prevention Institute of California, Fremont, CA/Stanford University School of Medicine, Stanford, CA), Manuel R. Teixeira (Portuguese Oncology Institute of Porto/University of Porto, Porto, Portugal), Susan L. Neuhausen (Beckman Research Institute of the City of Hope, Duarte, CA), Kim De Ruyck (Ghent University, Ghent, Belgium), Azad Razack (University of Malaya, Kuala Lumpur, Malaysia), Lisa F. Newcomb (Fred Hutchinson Cancer Research Center/University of Washington, Seattle, WA), Davor Lessel (Medical University, Sofia, Bulgaria), Radka Kaneva (Medical University, Sofia, Bulgaria), Nawaid Usmani (University of Alberta/Cross Cancer Institute, Edmonton, Alberta, Canada), Frank Claessens (KU Leuven, Leuven, Belgium), Paul A. Townsend (University of Manchester/St Mary's Hospital, Manchester, United Kingdom), Manuela Gago Dominguez (Servicio Galego de Saúde, Santiago De Compostela, Spain/University of California San Diego, La Jolla, CA), Monique J. Roobol (CE Rotterdam, the Netherlands), and Florence Menegaux (University Paris-Sud/University Paris-Saclay, Villejuif, France).

REFERENCES

1. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* 2016;127(20):2375-2390.

2. Kanzler H, Küppers R, Hansmann ML, Rajewsky K. Hodgkin and Reed-Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells. *J Exp Med.* 1996;184(4):1495-1505.

3. Brune V, Tiacci E, Pfeil I, et al. Origin and pathogenesis of nodular lymphocyte-predominant Hodgkin lymphoma as revealed by global gene expression analysis. *J Exp Med.* 2008;205(10):2251-2268.

4. Küppers R. The biology of Hodgkin's lymphoma. *Nat Rev Cancer.* 2009;9(1):15-27.

5. Greaves P, Clear A, Owen A, et al. Defining characteristics of classical Hodgkin lymphoma microenvironment T-helper cells. *Blood.* 2013;122(16):2856-2863.

6. MacMahon B. Epidemiological evidence of the nature of Hodgkin's disease. *Cancer.* 1957;10(5):1045-1054.

7. Weiss LM, Movahed LA, Warnke RA, Sklar J. Detection of Epstein-Barr viral genomes in Reed-Sternberg cells of Hodgkin's disease. *N Engl J Med.* 1989;320(8):502-506.

8. Hjalgrim H. On the aetiology of Hodgkin lymphoma. *Dan Med J.* 2012;59(7):B4485.

9. Kharazmi E, Fallah M, Pukkala E, et al. Risk of familial classical Hodgkin lymphoma by relationship, histology, age, and sex: a joint study from five Nordic countries. *Blood.* 2015;126(17):1990-1995.

10. Mack TM, Cozen W, Shibata DK, et al. Concordance for Hodgkin's disease in identical twins suggesting genetic susceptibility to the young-adult form of the disease. *N Engl J Med.* 1995;332(7):413-418.

11. Sud A, Thomsen H, Law PJ, et al; PRACTICAL consortium. Genome-wide association study of classical Hodgkin lymphoma identifies key regulators of disease susceptibility. *Nat Commun.* 2017;8(1):1892.

12. Cozen W, Timofeeva MN, Li D, et al. A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. *Nat Commun.* 2014;5(1):3856.

13. DeBoever C, anigawa Y, McInnes G, et al. Medical relevance of protein-truncating variants across 337,208 individuals in the UK Biobank study. *Nature Communications.* 2018;9;1612.

14. Enciso-Mora V, Broderick P, Ma Y, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). *Nat Genet.* 2010;42(12):1126-1130.

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

16. Schmermund A, Möhlenkamp S, Stang A, et al. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am Heart J.* 2002;144(2):212-218.

17. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc.* 2010;5(9):1564-1573.

18. Frampton M, da Silva Filho MI, Broderick P, et al. Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's lymphoma. *Nat Commun.* 2013;4(1):2549.

19. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2(12):e190.

20. O'Connell J, Sharp K, Shrine N, et al. Haplotype estimation for biobank-scale data sets. *Nat Genet.* 2016;48(7):817-820.

21. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv.* 2017;166298.

22. Abecasis GR, Altshuler D, Auton A, et al; 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing [published correction appears in *Nature.* 2011;473(7348):544]. *Nature.* 2010;467(7319):1061-1073.

23. Huang J, Howie B, McCarthy S, et al; UK10K Consortium. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun.* 2015;6(1):8111.

24. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39(7):906-913.

25. Khankhanian P, Cozen W, Himmelstein DS, et al. Meta-analysis of genome-wide association studies reveals genetic overlap between Hodgkin lymphoma and multiple sclerosis. *Int J Epidemiol.* 2016;45(3):728-740.

26. Urayama KY, Jarrett RF, Hjalgrim H, et al. Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. *J Natl Cancer Inst.* 2012;104(3):240-253.

27. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med.* 2002;21(11):1539-1558.

28. Stunnenberg HG, Hirst M, et al; International Human Epigenome Consortium. The International Human Epigenome Consortium: a blueprint for scientific collaboration and discovery [published correction appears in *Cell.* 2016;167(7):1897]. *Cell.* 2016;167(5):1145-1149.

29. Trynka G, Sandor C, Han B, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet.* 2013;45(2):124-130.

30. Rao SS, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping [published correction appears in *Cell.* 2015;162(3):687-688]. *Cell.* 2014;159(7):1665-1680.

31. Orlando G, Kinnersley B, Houlston RS. Capture Hi-C library generation and analysis to detect chromatin interactions. *Curr Protoc Hum Genet.* 2018;98(1):e63.

32. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):357-359.

33. Wingett S, Ewels P, Furlan-Magaril M, et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000 Res.* 2015;4:1310.

34. Cairns J, Freire-Pritchett P, Wingett SW, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* 2016;17(1):127.

35. Javierre BM, Burren OS, Wilder SP, et al; BLUEPRINT Consortium. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell.* 2016;167(5):1369-1384.

36. Scales M, Jäger R, Migliorini G, Houlston RS, Henrion MYR. visPIG--a web tool for producing multi-region, multi-track, multi-scale plots of genetic data. *PLoS One.* 2014;9(9):e107497.

37. Grundberg E, Small KS, Hedman ÅK, et al; Multiple Tissue Human Expression Resource (MuTHER) Consortium. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet.* 2012;44(10):1084-1089.

38. Battle A, Brown CD, Engelhardt BE, et al; eQTL manuscript working group. Genetic effects on gene expression across human tissues. *Nature.* 2017;550(7675):204-213.

39. Kasela S, Kisand K, Tserel L, et al. Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLoS Genet.* 2017;13(3):e1006643.

40. Bulik-Sullivan B, Finucane HK, Anttila V, et al; Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11):1236-1241.

41. Tian C, Hromatka BS, Kiefer AK, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun.* 2017;8(1):599.

42. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol.* 2013;37(7):658-665.

43. Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature.* 2011;473(7345):43-49.

44. Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nat Rev Cancer.* 2017;17(11):692-704.

45. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet.* 2015;47(6):598-606.

46. Vestergaard H, Westergaard T, Wohlfahrt J, Hjalgrim H, Melbye M. Tonsillitis,

tonsillectomy and Hodgkin's lymphoma. *Int J Cancer.* 2010;127(3):633-637.

47. Benharroch D, Shemer-Avni Y, Myint YY, et al. Measles virus: evidence of an association with Hodgkin's disease. *Br J Cancer.* 2004;91(3):572-579.

48. Anderson LA, Atman AA, McShane CM, Titmarsh GJ, Engels EA, Koshiol J. Common infection-related conditions and risk of lymphoid malignancies in older individuals. *Br J Cancer.* 2014;110(11):2796-2803.

49. Linabery AM, Erhardt EB, Fonstad RK, et al. Infectious, autoimmune and allergic diseases and risk of Hodgkin lymphoma in children and adolescents: a Children's Oncology Group study. *Int J Cancer.* 2014;135(6): 1454-1469.

50. Heise N, De Silva NS, Silva K, et al. Germinal center B cell maintenance and differentiation are controlled by distinct NF-κB transcription factor subunits. *J Exp Med.* 2014;211(10): 2103-2118.

51. Bartolomé-Izquierdo N, de Yébenes VG, Álvarez-Prado AF, et al. miR-28 regulates the germinal center reaction and blocks tumor growth in preclinical models of non-Hodgkin lymphoma. *Blood.* 2017;129(17):2408-2419.

52. Dent AL, Shaffer AL, Yu X, Allman D, Staudt LM. Control of inflammation, cytokine expression, and germinal center formation by BCL-6. *Science.* 1997;276(5312):589-592.

53. Victora GD, Schwickert TA, Fooksman DR, et al. Germinal center dynamics revealed by multiphoton microscopy with a photo-activatable fluorescent reporter. *Cell.* 2010; 143(4):592-605.

54. Lefebvre C, Rajbhandari P, Alvarez MJ, et al. A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol Syst Biol.* 2010;6:377.

55. Calado DP, Sasaki Y, Godinho SA, et al. The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. *Nat Immunol.* 2012;13(11): 1092-1100.

56. Schubart DB, Rolink A, Kosco-Vilbois MH, Botteri F, Matthias P. B-cell-specific coactivator OBF-1/OCA-B/Bob1 required for immune response and germinal centre formation. *Nature.* 1996;383(6600):538-542.

57. Adem J, Hämäläinen A, Ropponen A, et al. ERK1/2 has an essential role in B cell receptor- and CD40-induced signaling in an in vitro model of germinal center B cell selection. *Mol Immunol.* 2015;67(2 Pt B): 240-247.

58. Kwon K, Hutter C, Sun Q, et al. Instructive role of the transcription factor E2A in early B lymphopoiesis and germinal center B cell development. *Immunity.* 2008;28(6): 751-762.

59. De Silva NS, Klein U. Dynamics of B cells in germinal centres. *Nat Rev Immunol.* 2015; 15(3):137-148.

60. Kawabe T, Naka T, Yoshida K, et al. The immune responses in CD40-deficient mice: impaired immunoglobulin class switching and germinal center formation. *Immunity.* 1994;1(3):167-178.

61. Pearce EL, Mullen AC, Martins GA, et al. Control of effector CD8+ T cell function by the transcription factor Eomesodermin. *Science.* 2003;302(5647):1041-1043.

62. Wynn TA. IL-13 effector functions. *Annu Rev Immunol.* 2003;21(1):425-456.

63. Erdenebayar N, Maekawa Y, Nishida J, Kitamura A, Yasutomo K. Protein-tyrosine phosphatase-kappa regulates CD4+ T cell development through ERK1/2-mediated signaling. *Biochem Biophys Res Commun.* 2009;390(3):489-493.

64. Lesourne R, Uehara S, Lee J, et al. Themis, a T cell-specific protein important for late thymocyte development. *Nat Immunol.* 2009;10(8):840-847.

65. Allen RD III, Bender TP, Siu G. c-Myb is essential for early T cell development. *Genes Dev.* 1999;13(9):1073-1078.

66. Kaskow BJ, Buttrick TS, Klein HU, et al. MS AHI1 genetic risk promotes IFNγ+ CD4+ T cells. *Neurol Neuroimmunol Neuroinflamm.* 2017;5(1):e414.

67. Ting CN, Olson MC, Barton KP, Leiden JM. Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature.* 1996;384(6608):474-478.

68. Takahashi R, Nishimoto S, Muto G, et al. SOCS1 is essential for regulatory T cell functions by preventing loss of Foxp3 expression as well as IFN-gamma and IL-17A production. *J Exp Med.* 2011;208(10): 2055-2067.

69. Schuster C, Gerold KD, Schober K, et al. The autoimmunity-associated gene CLEC16A modulates thymic epithelial cell autophagy and alters T cell selection. *Immunity.* 2015; 42(5):942-952.

70. Siegmund K, Thuille N, Posch N, Fresser F, Baier G. Novel protein kinase C θ: coronin 1A complex in T lymphocytes. *Cell Commun Signal.* 2015;13(1):22.

71. Fischer AM, Katayama CD, Pagès G, Pouysségur J, Hedrick SM. The role of erk1 and erk2 in multiple stages of T cell development. *Immunity.* 2005;23(4):431-443.

72. Gilmore TD. The Rel/NF-kappaB signal transduction pathway: introduction. *Oncogene.* 1999;18(49):6842-6844.

73. Fujita F, Taniguchi Y, Kato T, et al. Identification of NAP1, a regulatory subunit of IkappaB kinase-related kinases that potentiates NF-kappaB signaling. *Mol Cell Biol.* 2003;23(21):7780-7793.

74. Wertz IE, O'Rourke KM, Zhou H, et al. Deubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling. *Nature.* 2004;430(7000):694-699.

75. Sarma V, Lin Z, Clark L, et al. Activation of the B-cell surface receptor CD40 induces A20, a novel zinc finger protein that inhibits apoptosis. *J Biol Chem.* 1995;270(21): 12343-12346.

76. Luo W, Weisel F, Shlomchik MJ. B cell receptor and CD40 signaling are rewired for synergistic induction of the c-Myc transcription factor in germinal center B cells. *Immunity.* 2018;48(2):313-326.

77. Martín-Subero JI, Gesk S, Harder L, et al. Recurrent involvement of the REL and BCL11A loci in classical Hodgkin lymphoma. *Blood.* 2002;99(4):1474-1477.

78. Schmitz R, Hansmann M-L, Bohle V, et al. TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma. *J Exp Med.* 2009;206(5): 981-989.

79. Reichel J, Chadburn A, Rubinstein PG, et al. Flow sorting and exome sequencing reveal the oncogenome of primary Hodgkin and Reed-Sternberg cells. *Blood.* 2015;125(7): 1061-1072.

80. Baus D, Pfitzner E. Specific function of STAT3, SOCS1, and SOCS3 in the regulation of proliferation and survival of classical Hodgkin lymphoma cells. *Int J Cancer.* 2006; 118(6):1404-1413.

81. Paley MA, Kroy DC, Odorizzi PM, et al. Progenitor and terminal subsets of CD8+ T cells cooperate to contain chronic viral infection. *Science.* 2012;338(6111): 1220-1225.

82. Intlekofer AM, Takemoto N, Wherry EJ, et al. Effector and memory CD8+ T cell fate coupled by T-bet and eomesodermin [published correction appears in *Nat Immunol.* 2006;7(1):113]. *Nat Immunol.* 2005;6(12): 1236-1244.

83. Yamamoto R, Nishikori M, Kitawaki T, et al. PD-1-PD-1 ligand interaction contributes to immunosuppressive microenvironment of Hodgkin lymphoma. *Blood.* 2008;111(6): 3220-3224.

84. Krappmann D, Emmerich F, Kordes U, Scharschmidt E, Dörken B, Scheidereit C. Molecular mechanisms of constitutive NF-kappaB/Rel activation in Hodgkin/Reed-Sternberg cells. *Oncogene.* 1999;18(4): 943-953.

85. Steidl C, Diepstra A, Lee T, et al. Gene expression profiling of microdissected Hodgkin Reed-Sternberg cells correlates with treatment outcome in classical Hodgkin lymphoma. *Blood.* 2012;120(17):3530-3540.

86. Wlodarska I, Nooyen P, Maes B, et al. Frequent occurrence of BCL6 rearrangements in nodular lymphocyte predominance Hodgkin lymphoma but not in classical Hodgkin lymphoma. *Blood.* 2003;101(2): 706-710.

87. Skinnider BF, Elia AJ, Gascoyne RD, et al. Interleukin 13 and interleukin 13 receptor are frequently expressed by Hodgkin and Reed-Sternberg cells of Hodgkin lymphoma. *Blood.* 2001;97(1):250-255.

88. Tang H, Wang H, Lin Q, et al. Loss of IP₃ receptor-mediated Ca²⁺ release in mouse B cells results in abnormal B cell development and function. *J Immunol.* 2017;199(2): 570-580.

89. Flavell JR, Baumforth KR, Wood VH, et al. Down-regulation of the TGF-beta target gene, PTPRK, by the Epstein-Barr virus encoded EBNA1 contributes to the growth and survival of Hodgkin lymphoma cells. *Blood.* 2008;111(1):292-301.

90. Chen Z, Stelekati E, Kurachi M, et al. miR-150 regulates memory CD8 T cell differentiation

via c-Myb. *Cell Reports.* 2017;20(11): 2584-2597.

91. Jücker M, Schaadt M, Diehl V, Poppema S, Jones D, Tesch H. Heterogeneous expression of proto-oncogenes in Hodgkin's disease derived cell lines. *Hematol Oncol.* 1990; 8(4):191-204.

92. Stanelle J, Döring C, Hansmann ML, Küppers R. Mechanisms of aberrant GATA3 expression in classical Hodgkin lymphoma and its consequences for the cytokine profile of Hodgkin and Reed/Sternberg cells. *Blood.* 2010;116(20):4202-4211.

93. Banerjee A, Northrup D, Boukarabila H, Jacobsen SE, Allman D. Transcriptional repression of Gata3 is essential for early B cell commitment. *Immunity.* 2013;38(5):930-942.

94. Re D, Müschen M, Ahmadi T, et al. Oct-2 and Bob-1 deficiency in Hodgkin and Reed Sternberg cells. *Cancer Res.* 2001;61(5): 2080-2084.

95. Schreiber A, Stengel F, Zhang Z, et al. Structural basis for the subunit assembly of the anaphase-promoting complex. *Nature.* 2011;470(7333):227-232.

96. Chan WK, Bhalla AD, Le Hir H, et al. A UPF3-mediated regulatory switch that maintains RNA surveillance. *Nat Struct Mol Biol.* 2009; 16(7):747-753.

97. Li J, Jørgensen SF, Maggadottir SM, et al. Association of CLEC16A with human common variable immunodeficiency disorder and role in murine B cells. *Nat Commun.* 2015;6(1):6804.

98. Singh TR, Ali AM, Busygina V, et al. BLAP18/RMI2, a novel OB-fold-containing protein, is an essential component of the Bloom helicase-double Holliday junction dissolvasome. *Genes Dev.* 2008;22(20): 2856-2868.

99. Watanabe M, Sasaki M, Itoh K, et al. JunB induced by constitutive CD30-extracellular signal-regulated kinase 1/2 mitogen-activated protein kinase signaling activates the CD30 promoter in anaplastic large cell lymphoma and Reed-Sternberg cells of Hodgkin lymphoma. *Cancer Res.* 2005; 65(17):7628-7634.

100. Stray-Pedersen A, Jouanguy E, Crequer A, et al. Compound heterozygous CORO1A mutations in siblings with a mucocutaneous-immunodeficiency syndrome of epidermodysplasia verruciformis-HPV, molluscum contagiosum and granulomatous tuberculoid leprosy. *J Clin Immunol.* 2014; 34(7):871-890.

101. Yasuda T, Sanjo H, Pagès G, et al. Erk kinases link pre-B cell receptor signaling to transcriptional events required for early B cell expansion. *Immunity.* 2008;28(4):499-508.

102. Yasuda T, Hayakawa F, Kurahashi S, et al. B cell receptor-ERK1/2 signal cancels PAX5-dependent repression of BLIMP1 through PAX5 phosphorylation: a mechanism of antigen-triggering plasma cell differentiation. *J Immunol.* 2012;188(12):6127-6134.

103. Wöhner M, Tagoh H, Bilic I, et al. Molecular functions of the transcription factors E2A and E2-2 in controlling germinal center B cell and plasma cell development. *J Exp Med.* 2016;213(7):1201-1221.

104. Carbone A, Gloghini A, Gattei V, et al. Expression of functional CD40 antigen on Reed-Sternberg cells and Hodgkin's disease cell lines. *Blood.* 1995;85(3):780-789.