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[14,15,16], 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], Richard S. Houlston[1]

1. Division of Genetics and Epidemiology, The Institute of Cancer Research, London, SW7 3RP, UK.
2. Division of Molecular Genetic Epidemiology, German Cancer Research Center, Heidelberg, 69120, Germany.
3. Center for Primary Health Care Research, Lund University, Malmö, 205 02, Sweden.
4. McGill University and Genome Quebec Innovation Centre, Department of Human Genetics, McGill University, Montreal, Quebec, Canada.
5. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK.
6. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, UK.
7. Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK.
8. Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, 69120, Germany.
9. Royal Marsden NHS Foundation Trust, London, SM2 5NG, UK.
10. Institute of Population Health, University of Manchester, M1 3BB, UK.
11. Division of Health Sciences, Warwick Medical School, Warwick University, CV4 7AL, UK.
12. University College London, Department of Applied Health Research, London, WC1E 7HB, UK.
13. University of Pisa, Department of Biology, Via Derna n 1, 56100 Pisa, Italy.
14. Human Genomic Research Group, Department of Biomedicine, University of Basel, 4031, Switzerland.
15. Institute of Human Genetics, University of Bonn, 53127, Germany.
16. Department of Genomics, Life & Brain Center, University of Bonn, 53127, Germany.
17. University of Duisburg–Essen, Essen, 47057, Germany.
18. Experimental Tumor Research, Center for Tumor Biology and Immunology, Clinic for Hematology, Oncology and Immunology, Philipps University, Hans-Meerwein-Str. 3, 35043 Marburg, Germany.
19. Division of Breast Cancer Research, The Institute of Cancer Research, London, SW7 3RP, UK.
20. University Hospital of Cologne, Department of Internal Medicine, Cologne, 50937, Germany.

[¥]Additional members from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium are provided in the end notes.

Key words: Hodgkin lymphoma, genetics, regulation, biology, lymphoma, risk, susceptibility, predisposition, B-cell, T-cell, cancer

Running title: Genome-wide association study of Hodgkin lymphoma

KEY POINTS:

- Variation at 6p21.31, 6q23.3, 11q23.1, 16p11.2 and 20q13.12 influences risk of HL
- Genetic predisposition implicates germinal centre dysfunction, disrupted T-cell function, and NF-κB activation in the pathogenesis of HL

ABSTRACT

To further our understanding of inherited susceptibility to Hodgkin lymphoma (HL), we performed a meta-analysis of seven genome-wide association studies totalling 5,325 HL cases and 22,423 controls. We identify five 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 five new and 13 known risk loci implicates dysfunction of the germinal centre 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.

INTRODUCTION

Hodgkin lymphoma (HL) comprises classical Hodgkin lymphoma (cHL) (~95% of cases) and nodular lymphocyte predominant HL (NLPHL, ~5% of cases)[1]. While cHL and NLPHL are defined by the Hodgkin and Reed-Sternberg (HRS) cell and the lymphocyte predominant (LP) cell respectively, both diseases are thought to arise from the malignant transformation germinal centre (GC) B-cell[2,3]. Furthermore, both cHL and NLPHL 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 supports 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 seven independent GWAS and report five new HL risk loci[11-13]. Integration of gene expression, histone modification and *in situ* promoter capture Hi-C data (PCHi-C) at the five new and the 13 known risk loci provides evidence for cell-type specificity in B- and T-cells and implicates dysfunction of the germinal centre 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 clinico-pathological 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 Organisation guidelines.

Genome-wide association studies

We used GWAS data generated on three non-overlapping case–control series of Northern European ancestry, which have been the subject of previous analyses (**Supplementary Tables 1** and **2**)[11]. The UK-GWAS was based on 622 cases ascertained through the Royal Marsden Hospital National Health Service Trust Family History study during 2004–2008[14], and 5,677 controls from the UK Wellcome Trust Case Control Consortium 2 (WTCCC2)[15]. The German-GWAS comprised 1,001 cases ascertained by the German Hodgkin Study Group during 1998–2007, and 2,092 controls from the Heinz Nixdorf Recall (HNR) study[16]. The UK-NSHLG-GWAS utilised 1,717 cases ascertained through the NSHLG (http://www.public.ukcrn.org.uk) from 2010 to 2013[11]. Controls comprised: (1) 2,976 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 SEARCH (Study of Epidemiology & Risk Factors in Cancer), recruited via GP practices in East Anglia (2003–2009), (2) 4,446 cancer-free women from across the UK via the Breast Cancer Association Consortium (BCAC). Details of the genotyping platform and quality control measures applied to each of the three GWAS have been described previously and are detailed in **Supplementary 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 (**Supplementary Figure 1**). Eigenvectors for the GWAS data sets were inferred using smartpca (part of EIGENSOFT) by merging cases and controls with Phase III 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 (**Supplementary Table 3**). SNPs with a call rate < 95% were excluded as were those with a MAF < 0.01 or displaying deviation from Hardy–Weinberg equilibrium (HWE) (*i.e.*, $P < 10^{-6}$, **Supplementary Table 4**). GWAS data were phased with SHAPEIT3[20], and imputed to >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 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 (Supplementary Fig. 2). The inflation factor λ_{1000} was based on the 90% least-significant SNP scaled to 1000 cases and 1000 controls.

In addition to analysing data from these three GWAS, we made use of pre-processed association test statistics for HL risk from a meta-analysis of three additional GWAS (USC-IARC-UC-GWAS) comprising 1,816 HL cases and 7, 879 contols[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 due to heterogeneity were calculated; an I^2 value $\geq 75\%$ 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 DSMZ and were cultured at 37°C in RPMI 1640 supplemented with 10% heat inactivated FBS (Thermo Scientific). Cell line identity was confirmed by STR-profiling. Cells were regularly tested for mycoplasma contamination (PromoCell, PK-CA91).

ChIP-seq analysis

L-428: ChIP-seq was performed on H3K27Ac and H3K4me3, for using antibodies obtained from Diagenode. Briefly, after cell lysing, sonication of nuclei was performed 293 (UCD-300, BioRuptor) to obtain 150-500bp 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 3-6µg of antibody and 2-4 million of sonicated cell lysate. ChIP samples were de-crosslinked 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 PCR were performed, followed by size selection for 200-400bp fragments and final library purification (GeneRead Size Selection kit, 301 Qiagen). ChIP libraries were sequenced

using HiSeq 2000 (Illumina) with 100bp 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 >100 samples from >30 cell types from the Blueprint Epigenome Consortium were analysed[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 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 min. Cross-linked DNA was digested with HindIII (NEB, R0104) and chromatin ends were filled and marked with biotin-14-dATP (ThermoFisher, 19524-016). The resulting blunted ended fragments were ligated at 16°C in the nucleus with T4 DNA ligase (NEB, M0202) to minimise random ligation. DNA purified after crosslinking was reversed by proteinase K (Ambion, AM2546) treatment. DNA was sheared by sonication (Covaris, M220) and 200-650bp fragments selected. Biotin tag DNA was pulled down with streptavidin beads and ligated with Illumina paired end adapters (Illumina). Six cycles of PCR 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, non-coding, antisense, snRNA, miRNA and snoRNA transcripts (**Supplementary Data**). After library enrichment, a post-capture PCR 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 three independent biological replicates were combined to define 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 2Mb window of the

sentinel SNP, and with a score ≥ 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 three publicly available resources: (i) lymphobastoid cell line (LCL) expression from the MuTHER (n = 825) consortium[37]; (ii) LCL expression from the GTEx consortium (n = 114)[38]; (iii) 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 dataset.

Genetic correlation with infection

To estimate the genetic correlation between specific infections and all HL, and NSHL and MCHL subtypes[40], we used LD score regression. Summary statistics for self-reported infectious diseases from over 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 randomisation

We performed two-sample MR using SNPs associated with specific infection-related traits as IVs. SNPs associated with each of the infection-related traits at genome-wide significance (i.e. $P \leq 5.0 \times 10^{-8}$) were used as IVs[41]. We analysed infection-related traits for which >2 SNPs had been shown to be associated with the specific infection (tonsillectomy, mumps infection, childhood ear infection and yeast infections). To avoid co-linearity between SNPs for each trait, we excluded SNPs that were correlated (i.e. r^2 value of ≥ 0.01) within each trait, and only considered the SNPs with the strongest effect on the trait for use as IVs. Where data on an IV was not present in the outcome trait, a proxy was utilised (r^2>0.6). Details of the IVs used are detailed in **Supplementary Data.** For each SNP, we recovered the chromosome position, risk allele, association estimates (per-allele log-OR) and standard errors. The allele that was associated with increased risk of the exposure was considered the effect allele. The odds ratios (OR) of HL, NSHL and MCHL per unit of standard deviation increment for each infection-related trait, were estimated using the 'Mendelian randomisation' R package[42]. Given that traits analysed 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 analysed summary level GWAS data generated on HL cases and controls of European ancestry[11] from three sources (**Supplementary Tables 1-4**): (1) two GWAS of UK cases and controls and one GWAS of German cases and controls, totalling 3,077 cases and 14,546 controls (Discovery GWAS)[11]; (2) the Stanford Global Biobank Engine, an analysis of 432 HL cases from the UK Biobank[13] and (3) a meta-analysis of three published HL GWAS totalling 1,816 HL cases and 7,879 controls (USC-IARC-UC-GWAS)[12,25,26].

In a meta-analysis of data from the seven studies, we identified new genome-wide significant associations for HL (**Figure 1** and **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 nodular sclerosis Hodgkin lymphoma (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 five new risk SNPs. (**Supplementary 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 five new and 13 known HL risk loci we analysed H3K4me3 and H3K27Ac chromatin marks which annotate regulatory regions, in over 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 demonstrates the greatest enrichment of promoter interacting regions[35]. The H3K4me3 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 centre B-cell ($P = 6.3 \times 10^{-3}$) (**Supplementary 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 ($P = 7.9 \times 10^{-3}$) (**Supplementary Table 8**). Based on the co-localisation of variants with active chromatin marks, we calculated an enrichment scores for each genetic association (**Figure 2**)[29]. High SNP regulatory scores were also shown in T-cells 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 non-coding 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 lymphoblastoid cell lines (LCL) and in CD4+ and CD8+ T-cells. We identified eQTLs in LCL 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** and **Supplementary 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 analysed PCHi-C data. Firstly, 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 naïve and total B-cells, as well as CD4+ and CD8+ T-cells[35]. Secondly, we generated and analysed 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**, **Supplementary Figure 3** and **Supplementary 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 of infection[41].

To investigate co-heritability 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 over 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 Hodgkin lymphoma (**Supplementary Table 11**). Following on from this, for infections with greater than two genetically defined instrumental variables (IVs), we performed a Mendelian randomisation (MR) analysis to identify a potential causal relationship with HL. For tonsillectomy, yeast infections and childhood ear infections, no statistically significant associations were demonstrated (**Supplementary Tables 12**). A nominally significant positive association between self-reported mumps infection and HL was found (P = 0.04), however this was not significant after correction for multiple testing.

DISCUSSION

By utilising publicly available summary statistics we have increased the power of our study allowing us to identify five new HL risk loci, thus bringing the total number of HL risk loci to 18. Whilst 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 centre B-cells, CD4+ thymocytes, CD4+ T-cells and CD8+ T-cells. Furthermore, whilst 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 utilised PCHi-C and gene expression data in these cell types to identify targets subject to regulatory control by HL risk SNPs. While in part speculative, and requiring functional validation, integrating proximity, cell specificity of risk loci, gene expression and PCHi-C data, our analyses highlight three biological processes and their associated genes as a basis of HL susceptibility (**Table 2**): the germinal centre 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 finding 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 pre-eminent 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 impact on 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. Firstly, the resolution of the Hi-C library using *HindIII*, a 6-base pair cutter, is approximately 10kb. As such, we are unable to detect concordant chromatin contacts at risk loci which influence the expression of genes located <10kb. Secondly, it is recognised that the range at which gene expression is perturbed to influence disease risk, may be narrow and as such 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 is limited by broad B- and T-cell populations, it is possible that we have not captured the cell type to analyse expression. As such we would view both methods as complimentary in identifying target genes.

The established association between EBV and risk of HL, coupled with other epidemiological observations provides strong *a priori* evidence for infection being a major etiological risk factor for HL. While our MR analysis failed to implicate a causal relationship with any of the self-reported infection traits, we acknowledge that our study 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 since tonsillectomy has previously been linked to HL in some epidemiological observational studies[46].

In conclusion, 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 highlights dysfunction of the germinal centre reaction, perturbed T-cell function and constitutive NF-κB activation as mechanisms by which genetic risk loci influence HL pathogenesis.

ACKNOWLEDGMENTS

In the United Kingdom, Bloodwise (LLR; 10021) provided principal funding for the study. Support from Cancer Research UK (C1298/A8362) and the Lymphoma Research Trust is also acknowledged. A.S. is supported by a clinical fellowship from Cancer Research UK and charitable funds from the Royal Marsden Hospital. For the UK-GWAS, sample and data acquisition were supported by Breast Cancer Now, the European Union and the Lymphoma Research Trust. The UK-GWAS made use of control genotyping data generated by the WTCCC. We acknowledge use of genotype data from the British 1958 Birth Cohort DNA collection, which was funded by the Medical Research Council Grant G0000934 and the Wellcome Trust Grant 068545/Z/02. A full list of the investigators who contributed to the generation of the data is available from http://www.wtccc.org.uk. Funding for this project was provided by the Wellcome Trust under awards 076113 and 085475. Patients for the new GWAS were ascertained through the National Study of Hodgkin Lymphoma Genetics (http://www.public.ukcrn.org.uk) and we thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of Genotyping data. For their help with UK sample collection we thank Hayley Evans, James Griffin, Joanne Micic, Susan Blackmore, Beverley Smith, Deborah Hogben, Alison Butlin, Jill Wood, Margot Pelerin, Alison Hart, Katarzyna Tomczyk and Sarah Chilcott-Burns. The BCAC study would not have been possible without the contributions of the following: Manjeet K. Bolla, Qin Wang, Kyriaki Michailidou and Joe Dennis. BCAC is funded by Cancer Research UK (C1287/A10118, C1287/A16563). For the BBCS study, we thank Eileen Williams, Elaine Ryder-Mills, Kara Sargus. The BBCS is funded by Cancer Research UK and Breast Cancer Now and acknowledges NHS funding to the National Institute of Health Research (NIHR) Biomedical Research Centre (BRC) and the National Cancer Research Network (NCRN). We thank the participants and the investigators of EPIC (European Prospective Investigation into Cancer and Nutrition). The coordination of EPIC is financially supported by the European Commission (DG-SANCO) and the International Agency for Research on Cancer. The national cohorts are supported by: Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), Federal Ministry of Education and Research (BMBF) (Germany); the Hellenic Health Foundation, the Stavros Niarchos Foundation (Greece); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek

Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS), PI13/00061 to Granada, PI13/01162 to EPIC-Murcia, Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, ISCIII RETIC (RD06/0020) (Spain); Cancer Research UK (14136 to EPIC-Norfolk; C570/A16491 and C8221/A19170 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk, MR/M012190/1 to EPIC-Oxford) (United Kingdom). We thank the SEARCH and EPIC teams, which were funded by a programme grant from Cancer Research UK (C490/A10124) and supported by the UK NIHR BRC at the University of Cambridge. We thank Breast Cancer Now and the Institute of Cancer Research (ICR) for support and funding of the UKBGS, and the study participants, study staff, and the doctors, nurses and other health-care providers and health information sources who have contributed to the study. We acknowledge NHS funding to the Royal Marsden/ICR NIHR BRC. UKGPCS would like to thank The Institute of Cancer Research and The Everyman Campaign for funding support. The UKGPCS acknowledges The Prostate Cancer Research Foundation, Prostate Action, The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI), the NIHR funding to the NIHR Biomedical Research data managers and consultants for their work in the UKGPCS study and urologists and other persons involved in the planning, and data collection of the CAPS study. Genotyping of the OncoArray was funded by the US National Institutes of Health (NIH) (U19 CA 148537 for ELucidating Loci Involved in Prostate cancer SuscEptibility (ELLIPSE) project and X01HG007492 to the Center for Inherited Disease Research (CIDR) under contract number HHSN268201200008I). Additional analytic support was provided by NIH NCI U01 CA188392 (PI: Schumacher). The PRACTICAL consortium was supported by Cancer Research UK Grants C5047/A7357, C1287/A10118, C1287/A16563, C5047/A3354, C5047/A10692, C16913/A6135, European Commission's Seventh Framework Programme grant agreement no. 223175 (HEALTH-F2-2009-223175), and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON initiative). We would also like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation, Prostate Research Campaign UK (now Prostate Action), The Orchid Cancer Appeal, The National Cancer Research Network UK, The National Cancer Research Institute (NCRI) UK. We are grateful for support of NIHR funding to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. The APBC BioResource, which form part of the PRACTICAL consortium, consists of the following members: Wayne Tilley, Gail Risbridger, Renea Taylor, Judith A Clements, Lisa Horvath, Vanessa Hayes, Lisa Butler, Trina Yeadon, Allison Eckert, Pamela Saunders, Anne-Maree Haynes, Melissa Papargiris. CIHR EP1-120608 supported the ChIP-seq work (PI: Pastinen). German funding was provided by the German Cancer Aid, the Harald Huppert

Foundations, The German Federal Ministry of Education and Research (eMed, Cliommics 01ZX1309B), the Heinz Nixdorf Foundation (Germany), the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen and the Faculty of Medicine University Duisburg–Essen. This study makes use of data generated by the Blueprint Consortium. A full list of the investigators who contributed to the generation of the data is available from www.blueprint-epigenome.eu. Funding for the project was provided by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 282510 BLUEPRINT. We thank Jose Martin-Subero (Institut d'investigacions Biomèdiques August Pi i Sunyer) for his advice with respect to the analysis of ChIP-seq data from the Blueprint Consortium. We also thank 23andMe for providing us with summary-level statistics of genetic associations to infection-related traits. Finally, we are grateful to all the patients and individuals for their participation and the clinicians, investigators, other staff who contributed to sample and data collection.

*The PRACTICAL Consortium (http://practical.icr.ac.uk/):

OncoArray (in addition to those named in the author list):

Brian E. Henderson[21], Christopher A. Haiman[21], Sara Benlloch[1,6], Fredrick R. Schumacher[22,23], Ali Amin Al Olama[6,24], Sonja I. Berndt[25], David V. Conti[21], Fredrik Wiklund[26], Stephen Chanock[25], Victoria L. Stevens[27], Catherine M. Tangen[28], Jyotsna Batra[29,30], APCB BioResource[29], Judith Clements[29,30], Henrik Gronberg[25], Johanna Schleutker[30,31,32], Demetrius Albanes[24], Stephanie Weinstein[24], Alicja Wolk[34], Catharine West[35], Lorelei Mucci[36], Géraldine Cancel-Tassin[37,38], Stella Koutros[25], Karina Dalsgaard Sorensen[39,40], Eli Marie Grindedal[41], David E. Neal[42,43], Ruth C. Travis[44], Robert J. Hamilton[45], Sue Ann Ingles[21], Barry Rosenstein[46,47], Yong-Jie Lu[48], Graham G. Giles[49,50], Adam S. Kibel[51], Ana Vega[52], Manolis Kogevinas[53,54,55,56], Kathryn L. Penney[57], Jong Y. Park[58], Janet L. Stanford[59,60], Cezary Cybulski[61], Børge G. Nordestgaard[62,63], Hermann Brenner[64,65,66], Christiane Maier[67], Jeri Kim[68], Esther M. John[69,70], Manuel R. Teixeira[71,72], Susan L. Neuhausen[73], Kim De Ruyck[74], Azad Razack[75], Lisa F. Newcomb[59,76], Davor Lessel[78], Radka Kaneva[78], Nawaid Usmani[79,80], Frank Claessens[81], Paul A. Townsend[82], Manuela Gago Dominguez[83,84], Monique J. Roobol[85], Florence Menegaux[86]

21. Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA 90033, USA.
22. Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA.
23. Seidman Cancer Center, University Hospitals, Cleveland, OH 44106, USA.
24. University of Cambridge, Department of Clinical Neurosciences, Cambridge CB2 1TN, UK.
25. Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD 20892, USA.
26. Department of Medical Epidemiology and Biostatistics, Karolinska Institute, SE-171 77 Stockholm, Sweden.

27. Epidemiology Research Program, American Cancer Society, 250 Williams Street, Atlanta, GA 30303, USA.
28. SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.
29. Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Queensland 4059, Australia.
30. Translational Research Institute, Brisbane, Queensland 4102, Australia.
31. Department of Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, FI-20520 Turku, Finland.
32. Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, FI-20520, Turku, Finland.
33. BioMediTech, University of Tampere, Tampere 33100, Finland.
34. Division of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, SE-171 77 Stockholm, Sweden.
35. Institute of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Radiotherapy Related Research, The Christie Hospital NHS Foundation Trust, Manchester M13 9NT, UK.
36. Department of Epidemiology, Harvard School of Pubic Health, Boston, MA 02115, USA.
37. CeRePP, Tenon Hospital, 75020 Paris, France.
38. Sorbonne Université, GRC n°5 ONCOTYPE-URO, Tenon Hospital, 75020 Paris, France.
39. Department of Molecular Medicine, Aarhus University Hospital, 8000 Aarhus C , Denmark.
40. Department of Clinical Medicine, Aarhus University, 8000 Aarhus C, Denmark.
41. Department of Medical Genetics, Oslo University Hospital, N-0424 Oslo, Norway.
42. University of Cambridge, Department of Oncology, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.
43. Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK.
44. Cancer Epidemiology, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK.
45. Dept. of Surgical Oncology, Princess Margaret Cancer Centre, Toronto M5G 2M9, Canada.
46. Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York , NY 10029, USA.
47. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.
48. Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, John Vane Science Centre, London EC1M 6BQ, UK.
49. Cancer Epidemiology Centre, The Cancer Council Victoria, Melbourne, Victoria 3004, Australia.
50. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria 3053, Australia.
51. Division of Urologic Surgery, Brigham and Womens Hospital, Boston, MA 02115, USA.
52. Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica, CIBERER, IDIS, 15782 Santiago de Compostela, Spain.
53. Centre for Research in Environmental Epidemiology (CREAL), Barcelona Institute for Global Health (ISGlobal), 60803 Barcelona, Spain.
54. CIBER Epidemiología y Salud Pública (CIBERESP), 28029 Madrid, Spain.
55. IMIM (Hospital del Mar Research Institute), 08003, Barcelona, Spain.
56. Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain.
57. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA 02115, USA.
58. Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, Florida 33612, USA.

59. Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, USA.
60. Department of Epidemiology, School of Public Health, University of Washington, Seattle, Washington 98195, USA.
61. International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, 70-001 Szczecin, Poland.
62. Faculty of Health and Medical Sciences, University of Copenhagen, 1165 Copenhagen, Denmark.
63. Department of Clinical Biochemistry, Herlev and Gentofte Hospital, 2900 Copenhagen, University Hospital, Herlev, Denmark.
64. Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.
65. German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany.
66. Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), 69120 Heidelberg, Germany.
67. Institute for Human Genetics, University Hospital Ulm, 89081 Ulm, Germany.
68. The University of Texas M. D. Anderson Cancer Center, Department of Genitourinary Medical Oncology, Houston 77030, TX, USA.
69. Cancer Prevention Institute of California, Fremont, CA 94538, USA.
70. Department of Health Research & Policy (Epidemiology) and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA 94305 , USA.
71. Department of Genetics, Portuguese Oncology Institute of Porto, 4200-072 Porto, Portugal.
72. Biomedical Sciences Institute (ICBAS), University of Porto, 4200-072 Porto, Portugal.
73. Department of Population Sciences, Beckman Research Institute of the City of Hope, Duarte, CA 91016, USA.
74. Ghent University, Faculty of Medicine and Health Sciences, Basic Medical Sciences, 9000 Ghent, Belgium.
75. Department of Surgery, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.
76. Department of Urology, University of Washington, Seattle, WA 98105, USA.
77. Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany.
78. Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical University, 1431 Sofia, Bulgaria.
79. Department of Oncology, Cross Cancer Institute, University of Alberta, Edmonton, Alberta T6G 2R3, Canada.
80. Division of Radiation Oncology, Cross Cancer Institute, Edmonton, Alberta T6G 1Z2, Canada.
81. Molecular Endocrinology Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, 3000 Leuven, Belgium.
82. Institute of Cancer Sciences, Manchester Cancer Research Centre, University of Manchester, Manchester Academic Health Science Centre, St Mary's Hospital, Manchester M13 9WL, UK.
83. Genomic Medicine Group, Galician Foundation of Genomic Medicine, Instituto de Investigacion Sanitaria de Santiago de Compostela (IDIS), Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde, SERGAS, 15706 Santiago De Compostela, Spain.
84. University of California San Diego, Moores Cancer Center, La Jolla, CA 92093, USA.
85. Department of Urology, Erasmus University Medical Center, 3015 CE Rotterdam, the Netherlands.
86. Cancer & Environment Group, Center for Research in Epidemiology and Population Health (CESP), INSERM, University Paris-Sud, University Paris-Saclay, F-94805 Villejuif, France

Information of the consortium can be found at http://practical.icr.ac.uk

AUTHORSHIP

Contribution: A.S., R.S.H. and K.H. designed and provided overall project management. A.S. and R.S.H. drafted the manuscript. R.S.H., A.J.S. and N.O. provided samples for UK-GWAS and UK-NSHLG-GWAS, D.E., P.P., A.D., J.P., F.C., R.E., Z.K.-J, K.M., N.P. and D.C. provided control samples for the UK-NSHLG-GWAS. A.S. and P.J.L. performed bioinformatic and statistical analysis. In the UK P.B. and A.S. performed sample and laboratory coordination. A.S. and R.C. provided clinical data on the UK samples. A.S and G.O. performed capture Hi-C experiments. F.H. and T.P. performed ChIP-seq experiments. In Germany, H.T. and A.F. performed bioinformatic and statistical analyses; P.H. and M.M.N. were responsible for German-GWAS analysis; K.-H.J. provided the German control samples; E.P.v.S. and A.E. were responsible for German HL patients.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Amit Sud, Division of Genetics and Epidemiology, The Institute of Cancer Research, London. Tel: +44 (0) 208 722 4635, E-mail: amit.sud@icr.ac.uk

DATA AVAILABILITY

Sequencing data, which forms the reference panel for imputation, have been deposited in the European Genome-phenome Archive (EGA) under accession codes EGAD00001000776.

Transcriptional profiling data from the MuTHER consortium that support the findings of this work have been deposited in the European Bioinformatics Institute (Part of the European Molecular Biology Laboratory, EMBL-EBI) under accession code E-TABM-1140.

ChIP-seq data from the Blueprint Epigenome Consortium are available from http://dcc.blueprint-epigenome.eu/#/home.

Hi-C data from the Blueprint Epigenome Consortium are available from https://osf.io/u8tzp/.

ChIP-seq data for the HRS cell line L-428 are deposited under the accession number EGAS00001003033

Hi-C data for the HRS cell line L-428 are deposited in EGA under accession number EGAS00001003032.

Summary statistics for genetic susceptibility to infection-related traits are available upon request from 23andMe. Please visit research/23andme.com/collaborate to request access to these datasets.

REFERENCES

1. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization (WHO) classification of lymphoid neoplasms. *Blood*. 2016.
2. Kanzler H, Kuppers 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. *The Journal of Experimental Medicine*. 2008;205(10):2251-2268.
4. Küppers R. The biology of Hodgkin's lymphoma. *Nature Reviews Cancer*. 2008;9:15.
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 on 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. *New England Journal of Medicine*. 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. Genome-wide association study of classical Hodgkin lymphoma identifies key regulators of disease susceptibility. *Nature Communications*. 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. *Nature Communications*. 2014;5:3856.
13. DeBoever C, Tanigawa Y, McInnes G, et al. Medical relevance of protein-truncating variants across 337,208 individuals in the UK Biobank study. *bioRxiv*. 2017.
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). *Nature Genetics*. 2010;42:1126.
15. The Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661.
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. *American Heart Journal*. 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. *Nature protocols*. 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. *Nature Communications*. 2013;4:2549.
19. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. *PLOS Genetics*. 2006;2(12):e190.
20. O'Connell J, Sharp K, Shrine N, et al. Haplotype estimation for biobank-scale data sets. *Nature Genetics*. 2016;48:817.
21. Bycroft C, Freeman C, Petkova D, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*. 2017.

22. The Genomes Project C. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061.
23. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun*. 2015;6:8111.
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. *JNCI: Journal of the National Cancer Institute*. 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, Abrignani S, Adams D, et al. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *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. *Nature Genetics*. 2012;45:124.
30. Rao SS, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159(7):1665-1680.
31. Orlando G, Kinnersley B, Houlston RS. Capture Hi-C Library Generation and Analysis to Detect Chromatin Interactions. *Current Protocols in Human Genetics*. 2018;0(0):e63.
32. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature 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. *F1000Research*. 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. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 2016;167(5):1369-1384.e1319.
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. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature genetics*. 2012;44(10):1084-1089.
38. Consortium GT. Genetic effects on gene expression across human tissues. *Nature*. 2017;550:204.
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 Genetics*. 2017;13(3):e1006643.
40. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nature Genetics*. 2015;47:1236.
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. *Nature Communications*. 2017;8(1):599.
42. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization Analysis With Multiple Genetic Variants Using Summarized Data. *Genetic Epidemiology*. 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;473:43.
44. Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. *Nature Reviews Cancer*. 2017;17:692.

45. Mifsud B, Tavares-Cadete F, Young AN, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*. 2015;47:598.
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. *British Journal of 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. *British Journal of 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. *International Journal of 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-kappaB transcription factor subunits. *J Exp Med*. 2014;211(10):2103-2118.
51. Bartolome-Izquierdo N, de Yebenes VG, Alvarez-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 Photoactivatable 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. *Nature Immunology*. 2012;13:1092.
56. Schubart DB, Rolink A, Kosco-Vilbois MH, Botteri F, Matthias P. B-cell-specif ic 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. *Molecular Immunology*. 2015;67(2, Part 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. *Nature Reviews Immunology*. 2015;15:137.
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. *Annual Review of Immunology*. 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, 3rd, 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 IFNgamma(+) CD4(+) T cells. *Neurol Neuroimmunol Neuroinflamm*. 2018;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 Communication and Signaling*. 2015;13(1):22.
71. Fischer AM, Katayama CD, Pages G, Pouyssegur 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 IκB Kinase-Related Kinases That Potentiates NF-κB Signaling. *Molecular and Cellular Biology*. 2003;23(21):7780-7793.
74. Wertz IE, O'Rourke KM, Zhou H, et al. De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-[kappa]B 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. *Journal of Biological Chemistry*. 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.
77. Martin-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. *The Journal of Experimental Medicine*. 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 (New York, NY)*. 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. *Nature Immunology*. 2005;6:1236.
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, Dorken 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. *The Journal of Immunology*. 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. *Hematological Oncology*. 1990;8(4):191-204.
92. Stanelle J, Doring C, Hansmann ML, Kuppers 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 Research*. 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, Jorgensen SF, Maggadottir SM, et al. Association of CLEC16A with human common variable immunodeficiency disorder and role in murine B cells. *Nat Commun*. 2015;6:6804.
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. *Journal of Clinical Immunology*. 2014;34(7):871-890.
101. Yasuda T, Sanjo H, Pages 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. *The Journal of Experimental Medicine*. 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.

Table 1: Summary results for newly identified Hodgkin lymphoma risk loci. Freq, frequency; bp, base pair; OR, odds ratio; CI, confidence interval; I^2 proportion of the total variation due to heterogeneity.[¥] Summary statistics from 1,200 cHL patients and 6,417 controls[26]. †Genes at each risk locus are given for identification purposes only and does not necessarily indicate biological functionality.

		Risk		Discovery GWAS meta-analysis		UK Biobank		USC-IARC-UC-GWAS		Meta-analysis			
Locus	**Nearest genes†**	**allele (freq)**	**Position (hg19, bp)**	***P* value**	**OR (95% CI)**	***P* value**	**OR (95% CI)**	***P* value**	**OR (95% CI)**	***P* value**	**OR (95% CI)**	**I^2 (%)**	**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	0.29
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	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	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	0.68
16p11.2, rs6565176	*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	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	0.06

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. bp, base pair; cHL, classical Hodgkin lymphoma; NSHL; nodular sclerosis Hodgkin lymphoma; SNP, single nucleotide polymorphism; LCL, lymphoblastoid cell lines; LD, linkage disequilibrium. ¥SNPs (r^2 < 2.5 kilobases from ChIP-seq peak).

Locus	Sentinel SNP	Position (bp, hg19)	Gene(s) in LD block	Coding variant(s)	Promotor /UTR variant(s)	Expression quantitative trait loci in LCL	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
2p16.1	rs2420518	61054980						Naïve B-cell	*REL*	*REL*		*REL*[84]	Constitutive NF-κB activation (*REL*)[72] Altered B-cell differentiation and germinal centre 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- naïve B-cell Naïve B-cell Germinal centre B-cell L-428				*BCL6*[86]	Dysfunction of B-cell germinal centre 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 Naïve B-cell Class switched memory B-cell	*BAK1*, *SYNGAP1*, *GGNBP1*, *LINC00336*	*BAK1*, *GRM4*, *SYNGAP1*, *KIFC1*, *CUTA*, *PHF1*, *GGNBP1*, *LINC003336*			Altered B-cell differentiation (*ITPR3*)[88]
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 centre B-cell Plasma cell L-428	*MYB*		*MYB*		Altered T-cell differentiation (*MYB*)[65,90] Altered B-cell differentiation and germinal centre 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[78]	Constitutive NF-κB activation (*TNFAIP3*)[74]
8q24.21	rs34748721	129195943						Naïve B-cell Class switched memory B-cell	*CASC11, MYC, PVT1, RNU1-106P, MIR1207*			*MYC*[91]	Dysfunction of B-cell germinal centre reaction (*MYC*)[55,76]
10p14	rs2388486	8099021	*GATA3*			*GATA3*(↓)		CD3- CD4+ CD8+ thymocyte CD3+ CD4+ CD8+ thymocyte CD4+ T-cell CD8+ T-cell				*GATA3*[92]	Altered T-cell differentiation (*GATA3*)[67]

								Effector memory CD8+ T-cell Germinal centre B-cell					
10p14	rs3781093	8101927	*GATA3*			*GATA3* (↓)		CD4+ T-cell CD8+ T-cell				*GATA3*	Altered CD4+ T-cell differentiation (*GATA3*)[67] 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 centre B-cell Unswitched memory B-cell Class switched memory B-cell Plasma cell	FDX1	FDX1, PPP3R1B, ALG9, FDXACB1, DIXDC1		*POU2AF1*[94]	Dysfunctional germinal centre reaction (*POU2AF1*)[56]
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 centre 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 centre 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*)[68]. 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*)[63,70,71] Dysfunction of B-cell germinal centre reaction (*MAPK3*)[57,101,102]
19p13.3	rs2012125	1630341	*TCF3A*					CD38- B-cell Naïve B-cell Class switched memory B-cell					Dysfunction of B-cell germinal centre 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 centre B-cell L-428	CD40	TP53RK		*CD40*[104]	Dysfunctional germinal centre reaction (*CD40*)[53,60] Constitutive NF-κB activation (*CD40*)[75,76]

FIGURE LEGENDS

Figure 1: Genome-wide meta-analysis *P* values of Hodgkin's lymphoma risk ($-\log_{10}P$, *y* axis) plotted against their chromosomal positions (*x* axis). Novel HL risk loci and candidate gene are in orange.

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 each cell line, normalised across all immune cell types. Thus, a SNP within a chromatin mark that is active in only one 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.