Journal of Cancer Research and Clinical Oncology (2025) 151:171
https://doi.org/10.1007/s00432-025-06224-8

RESEARCH

Integration of transcriptome-wide association study and gene-based association analysis identifies candidate genes for Hodgkin lymphoma

Wen-Hui Jia[1] · Chang-Ling Huang[1] · Wen-Li Zhang[2] · Yong-Qiao He[2] · Wen-Qiong Xue[2] · Ying Liao[2] · Zhi-Yang Zhao[1] · Meng-Xuan Yang[1] · Lu Pei[2] · Wei-Hua Jia[1,2] · Tong-Min Wang[2]

Received: 28 January 2025 / Accepted: 4 May 2025
© The Author(s) 2025

Abstract

Background Genome-wide association studies (GWASs) have pinpointed many susceptibility loci for Hodgkin Lymphoma (HL), but their underlying biological mechanisms remain unclear.

Methods Utilizing GWAS data from the UK Biobank and FinnGen, along with expression quantitative trait loci (eQTL) statistics from the Genotype-Tissue Expression (GTEx) and the eQTL Catalogue, we carried out a large-scale gene-level association study using Omnibus Transcriptome Test with Expression Reference Summary data (OTTERS), and gene-based analysis with eQTL Multi-marker Analysis of Genomic Annotation (E-MAGMA).

Results We identified sixteen susceptibility genes for HL (FDR < 0.01), primarily immune-related, including *HLA-DQA1*, *HLA-DQA2*, *HLA-DQB1*, *HLA-DRB1*, *HLA-DRB5*, *HLA-DMA*, and *HLA-DPB1*, alongside genes involved in apoptosis, RNA processing, transcriptional regulation, and signal transduction. We identified five novel plausible genes, including *HLA-DMA*, *HLA-DPB1*, *LSM2*, *AAR2*, and *NOTCH4*.

Conclusion These findings highlight the role of the exogenous antigen presentation pathway in HL, shedding light on potential mechanisms.

Keywords Hodgkin lymphoma · Susceptibility genes · Human leukocyte antigen

Introduction

Hodgkin lymphoma (HL) is an aggressive B-cell tumor that features Hodgkin/Reed-Sternberg cells. It mainly consists of the classical HL (cHL), which makes up 90% of all cases, and nodular lymphocyte-predominant HL (NLPHL) (Connors et al. 2020). According to GLOBOCAN 2022, there were 82,409 new HL cases and 22,701 deaths worldwide (Bray et al. 2024). Key risk factors include genetic factors (Kharazmi et al. 2015), environmental influences (Ribeiro et al. 2021), and virus infections like Epstein-Barr virus (EBV) and human immunodeficiency virus (HIV) (Carbone et al. 2017). Extensive research has uncovered genetic susceptibility factors for HL. Linkage and candidate gene studies have found some associations of HLA (Human Leukocyte Antigen) class I and class II alleles with cHL (Kushekhar et al. 2014). Additionally, candidate gene studies have pinpointed susceptibility loci near non-HLA genes involved in immune regulation, carcinogen and folate metabolism, DNA repair, and other HL-related pathways (Sud et al. 2017b). GWASs have identified multiple susceptibility regions for HL, with the most significant associations mapped to the HLA class II region near *HLA-DRA*, *HLA-DRB1*, and *HLA-DRB9* (Cozen et al. 2012; Enciso-Mora et al. 2010; Sud et al. 2017a; Urayama et al. 2012). Other signals in the 6p21 region include rs2248462 (*MICB*), associated with overall cHL, as well as rs2734986 (*HLA-G/HLA-A*) and rs6904029 (*HCG9*), which are linked to EBV-positive classical HL (Urayama et al. 2012). Beyond the HLA region, susceptibility loci have been identified at 2p16.1, 3p24.1, 3q28, 5q31, 6q22.33, 6q23.3, 8q24.21, 10p14, 11q22.3, 11q23.1, 13q34,

✉ Wei-Hua Jia
jiawh@sysucc.org.cn

✉ Tong-Min Wang
wangtm@sysucc.org.cn

[1] School of Public Health, Sun Yat-sen University, Guangzhou 510080, China

[2] State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, China

Published online: 20 May 2025

16p11.2, 16p13.13, 19p13.3, and 20q13.12. These loci map to regions adjacent to genes involved in hematopoiesis and immune regulation (Chen et al. 2022; Cozen et al. 2014; Enciso-Mora et al. 2010; Frampton et al. 2013; Sud et al. 2017a; Urayama et al. 2012; Sud et al. 2018).

Despite these findings, many GWAS-identified variants reside in large haplotype blocks or non-coding regions, complicating the identification of their functional mechanisms and potential causal genes. In this context, eQTL in disease-relevant tissues provides complementary insights by linking disease-associated SNPs to gene expression, shedding light on underlying biological mechanisms. Advances in bioinformatics have led to the development of various approaches that combine GWAS and eQTL data. For instance, the OTTERS combines GWAS and eQTL summary statistics to find genes significantly associated with diseases and predict altered gene expression in cases (Dai et al. 2023). Another tool, E-MAGMA, refines SNP-to-gene mapping by incorporating tissue- or cell-type-specific eQTL data, followed by gene association analysis (Gerring et al. 2021). Combining these gene-level association analyses with different statistical methods reduces false discoveries and provides reliable, complementary results.

Here, we integrated GWAS summary data from the UK Biobank and FinnGen with seven eQTL datasets from the GTExV7 and the eQTL Catalogue to perform OTTERS and E-MAGMA analyses to pinpoint susceptibility genes for HL.

Methods

GWAS population and data collection

Participants for this study were drawn from the UK Biobank and FinnGen. The UK Biobank enrolled over 500,000 participants from 2006 to 2010, collecting comprehensive phenotypic and genotypic data at baseline and conducting long-term follow-up to monitor health outcomes (Sudlow et al. 2015). We utilized GWAS summary statistics from the publicly accessible UKBB PheWeb (https://pheweb.org/UKB-TOPMed/) for the analysis. PheWeb provides ICD-based GWAS results derived from electronic health records, encompassing 1,403 PheWAS codes for binary traits, including HL (PheCode 201). ICD-10 codes C81.0 to C81.3 had been used to identify HL cases in the UK Biobank (259 cases, 402,715 controls). Association testing had been performed by SAIGE (Scalable and Accurate Implementation of Generalized Mixed Models), adjusting for sex, birth year, and the first four PCs (principal components) (Zhou et al. 2018). We included a total of 8,978,153 SNPs with MAF (Minor Allele Frequency) > 0.01 for further analysis. FinnGen is a large-scale project combining public and private efforts to analyze genomic and health data from 500,000 individuals enrolled in Finnish biobanks. The quality control steps were previously described. We used GWAS summary data from the December 2022 8th release (https://r8.finngen.fi/), encompassing 690 HL patients and 271,463 controls with 8,990,713 SNPs (MAF > 0.01). HL was determined using ICD-10 codes C81.0-3. Genome-wide associations had been performed by SAIGE and adjusted for age, sex, the first ten PCs, and genotyping batch (Kurki et al. 2023).

eQTL data source

Summary statistics for seven cis-eQTLs related to two HL-relevant tissues were sourced from the GTEx (version 7) (Battle et al. 2017) and the eQTL Catalogue, including GTExV7 EBV-transformed LCL ($n = 117$), GTExV7 Whole blood ($n = 369$), eQTL Catalogue GENCORD LCL ($n = 190$), eQTL Catalogue GEUVADIS LCL ($n = 445$), eQTL Catalogue TwinsUK LCL ($n = 418$), eQTL Catalogue TwinsUK blood ($n = 195$), and eQTL Catalogue Lepik 2017 blood ($n = 471$) (Table S1).

GWAS meta-analysis for HL

A GWAS meta-analysis of 949 HL cases and 674,178 controls from the UK Biobank and FinnGen was performed using METAL with a fixed-effects inverse variance weighted model (Willer et al. 2010). We included 7,689,304 variants common to the UK Biobank and FinnGen. Stepwise conditional analysis was conducted in GCTA-COJO (--cojo-slct) to identify independent SNPs (Yang et al. 2012). Linkage disequilibrium (LD) was estimated using 10,000 unrelated Europeans randomly selected from the UK Biobank. For non-HLA regions, loci with $P < 1 \times 10^{-6}$ and $r^2 < 0.1$ were included, while stricter criteria ($P < 5 \times 10^{-8}$ and $r^2 < 0.01$) were applied for the HLA region to identify independent susceptibility signals.

Fine mapping analysis

To identify the most likely causal variants in HL-associated genomic regions, we performed fine-mapping by SuSiE (Sum of Single Effects) (Wang et al. 2020) via the easyfinemap pipeline (version 0.4.4, https://jianhua-wang.github.io/easyfinemap/). SuSiE assigned posterior probabilities to variants, indicating their potential causality for HL. To define credible sets of potentially causal variants, we focused on a 500 kb window around the lead SNPs reaching the genome-wide significant threshold ($P < 5 \times 10^{-8}$)

and applied a stringent probability threshold of 0.95 for the credible set.

Gene-level association analysis

We executed TWAS utilizing OTTERS that amplifies statistical power by integrating five polygenic risk score (PRS) models alongside cis-eQTL training in two phases (Dai et al. 2023). In stage I, OTTERS employs four PRS methods with five models, including P-value thresholding (P+T) thresholding 0.05 and 0.001 with LD clumping (Privé et al. 2019), frequentist LASSO (Mak et al. 2017), Bayesian regression with continuous shrinkage priors (PRS-CS) (Ge et al. 2019), and a nonparametric Bayesian approach (SDPR) (Zhou et al. 2021) to estimate cis-eQTL weights combining eQTL summary data and reference LD from training samples. It then conducts gene-level association analysis with the GWAS summary statistics to generate Z scores and *P* values. In stage II, the aggregated Cauchy association test (ACAT) is applied to calculate the *P* value (Liu et al. 2019). Gene-phenotype associations were considered significant if the ACAT *P* value met the Benjamini-Hochberg (BH) correction threshold (FDR < 0.01 for each eQTL dataset), at least two PRS models showed *P* value < 0.05, and the Z score directions were consistent across all five models.

We also performed gene-based association analysis for HL through E-MAGMA (Gerring et al. 2021), which builds on the MAGMA framework and employs a multiple linear principal component regression model to improve statistical performance (de Leeuw et al. 2015). SNP annotation and gene association formed the two components of the analysis. For the annotation, we utilized eQTL summary statistics from seven eQTL datasets involved in the OTTERS, selecting SNP-gene expression associations with FDR-adjusted $P < 0.05$. SNPs were mapped to their associated genes to ensure high-confidence annotations. Genes that met the Benjamini-Hochberg correction threshold (FDR < 0.01) were identified as HL-related genes by E-MAGMA.

To minimize the false discovery rate, we defined the genes that reached the Benjamini-Hochberg correction threshold by both OTTERS and E-MAGMA in each eQTL database (FDR < 0.01) as HL susceptibility genes. By leveraging these two complementary methods, we aim to enhance the robustness of the results of gene-level association studies.

Conditional analyses

To identify potential novel susceptibility genes, we performed E-MAGMA and OTTERS after conditioning on previously reported GWAS risk SNPs. We first extracted 79 SNPs from the GWAS Catalog that reached genome-wide significance ($P < 5 \times 10^{-8}$) for HL. We then performed LD clumping ($--r^2 < 0.1$) for the HLA region (30–34 Mb) and 1p13.2, identifying eight independent lead SNPs. Conditional regression analysis was then conducted on the HL GWAS data using GCTA-COJO (--cojo-cond). Finally, OTTERS and E-MAGMA analyses were performed using the conditional GWAS summary results, and genes that reached the Benjamini-Hochberg correction significant threshold ($FDR_{cojo} < 0.05$) in both methods were defined as novel susceptibility genes that were not captured by previously reported GWAS signals.

Gene set enrichment analysis

To investigate the biological significance of HL susceptibility genes, we applied the ClusterProfiler (Yu et al. 2012) to conduct enrichment analysis of the significant genes identified by both E-MAGMA and OTTERS. We assessed enrichment in Gene Ontology (GO) biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for humans. ClusterProfiler employs an Over-Representation Analysis (ORA) strategy, utilizing the hypergeometric distribution to calculate the *P*-value and determine whether gene sets related to known biological functions are enriched in the gene sets of interest. Gene sets with a Benjamini-Hochberg corrected FDR < 0.05 and an Enrichment Ratio ≥ 5% were considered significantly enriched.

Results

GWAS meta-analysis identifies five regions associated with HL risk

We performed a GWAS meta-analysis based on 949 HL patients and 674,178 controls from the UK Biobank and FinnGen cohorts. After quality control and meta-analysis, a total of 7,689,303 variants were retained. A total of twenty-two variants within the HLA region and one at 1p13.2 region were found to reach the genome-wide significance threshold ($P < 5 \times 10^{-8}$, Fig. 1). To identify independent genetic susceptibility signals in each region, we performed stepwise conditional analysis using GCTA-COJO on loci within each genetic susceptibility region (Yang et al. 2012). We identified two independent SNPs associated with HL surpassing $P < 5 \times 10^{-8}$, including rs9271406 at 6p21.32 (OR = 0.75, 95% CI = 0.68–0.82, $P_{meta} = 1.17 \times 10^{-10}$, $P_{COJO} = 1.17 \times 10^{-10}$) and rs1230666 at 1p13.2 (OR = 0.72, 95% CI = 0.64–0.81, $P_{meta} = 3.12 \times 10^{-8}$, $P_{COJO} = 3.12 \times 10^{-8}$; Table 1, Table S2). In addition, we identified three loci surpassing suggestive significance ($P < 1 \times 10^{-6}$), including 8q24.21, 17q25.3, and 20q13.33 (Table 1; Figs. 1 and

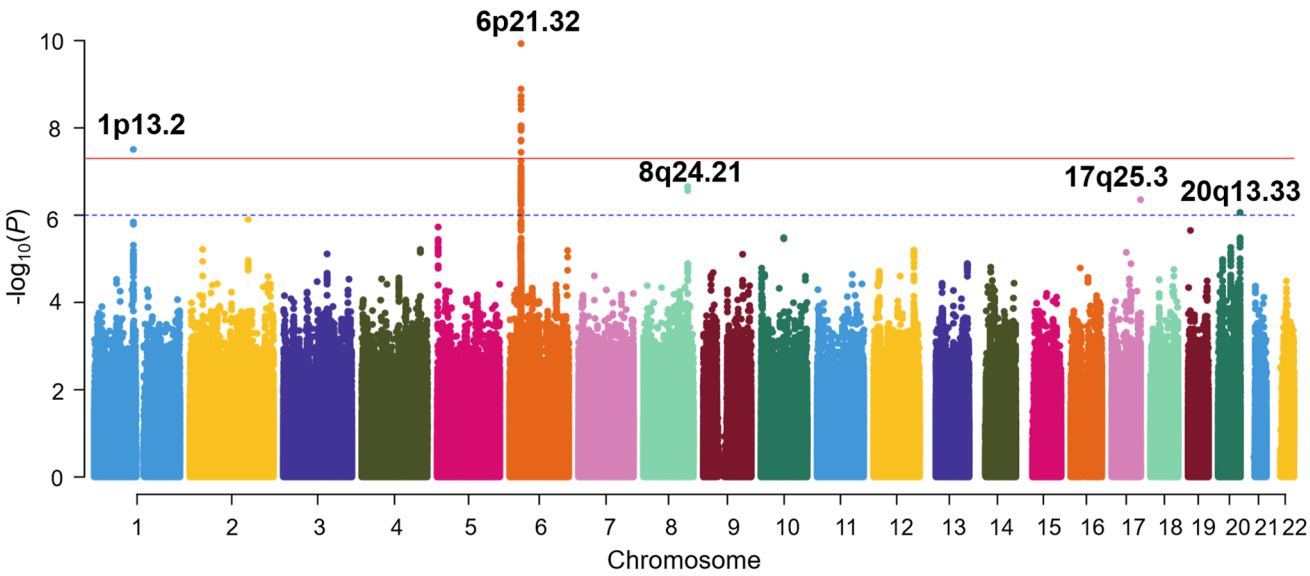

Fig. 1 Manhattan plot of GWAS meta-analysis for HL in 949 cases and 674,178 controls. The X-axis represents the chromosomal physical positions of SNPs, while the Y-axis represents the -log10 transformed *P*-values of the SNPs. The red line indicates the genome-wide significance threshold of $P = 5.0 \times 10^{-8}$, and the blue line indicates the suggestive threshold of $P = 1.0 \times 10^{-6}$

Table 1 GWAS meta-analysis identified five independent genetic susceptibility loci for HL

SNP	Position [a]	Cytoband	Allele [b]	Nearby Gene	Study	EAF		OR (95% CI) [c]	P^c
						Case	Control		
rs9271406	6:32,587,588	6p21.32	G/A	*HLA-DQA1*	UK Biobank	0.44	0.50	0.79(0.67,0.94)	8.50×10^{-3}
					FinnGen	0.46	0.54	0.73(0.66,0.81)	2.79×10^{-9}
					Meta	0.45	0.52	0.75(0.68,0.82)	1.17×10^{-10}
rs1230666	1:114,173,410	1p13.2	G/A	*MAGI3*	UK Biobank	0.79	0.85	0.61(0.48,0.77)	6.90×10^{-5}
					FinnGen	0.79	0.83	0.76(0.67,0.87)	6.08×10^{-5}
					Meta	0.79	0.84	0.72(0.64,0.81)	3.12×10^{-8}
rs71520688	8:129,056,888	8q24.21	T/C	*PVT1*	UK Biobank	0.12	0.10	1.38(1.03, 1.85)	0.03
					FinnGen	0.16	0.12	1.42(1.23,1.65)	2.30×10^{-6}
					Meta	0.14	0.11	1.41(1.24,1.61)	2.21×10^{-7}
rs78604106	17:80,920,842	17q25.3	T/C	*B3GNTL1*	UK Biobank	0.05	0.04	1.55(0.99,2.44)	0.06
					FinnGen	0.09	0.06	1.58(1.30,1.91)	2.95×10^{-6}
					Meta	0.07	0.05	1.57(1.32,1.88)	4.42×10^{-7}
rs2427513	20:61,755,392	20q13.33	A/C	*YTHDF1*	UK Biobank	0.48	0.44	1.17(0.99,1.40)	0.07
					FinnGen	0.55	0.49	1.28(1.15,1.42)	3.23×10^{-6}
					Meta	0.52	0.46	1.25(1.14,1.37)	8.68×10^{-7}

[a]: The chromosome, genomic positions of susceptibility loci according to the hg19 reference

[b]: Effective allele/reference allele

[c]: Odds ratios (ORs), 95% confidence intervals (CI) and *P* values of GWAS analysis in each study and meta-analysis

Meta-analysis of three study samples was performed using a fixed-effect model

SNP: Single nucleotide polymorphism

EAF: Effective allele frequency

2). The genetic inflation factor (λ) was 1.022, indicating no apparent population stratification (Figure S1).

Fine mapping reveals credibly causal variants

We performed fine-mapping for the 2 regions (1p13.2, 6p21.32) reaching the genome-wide significant threshold by including all variants up- and downstream of the two lead SNPs at a 250 kb window from the HL meta GWAS using SuSiE based on linkage disequilibrium (LD) (Wang et al. 2020). Two variants with causality (posterior probability > 0.5) were found, including rs9271406 at 6p21.32 (PP_SUSIE = 0.53) and rs1230666 at 1p13.2 (PP_SUSIE = 0.61) (Table S3).

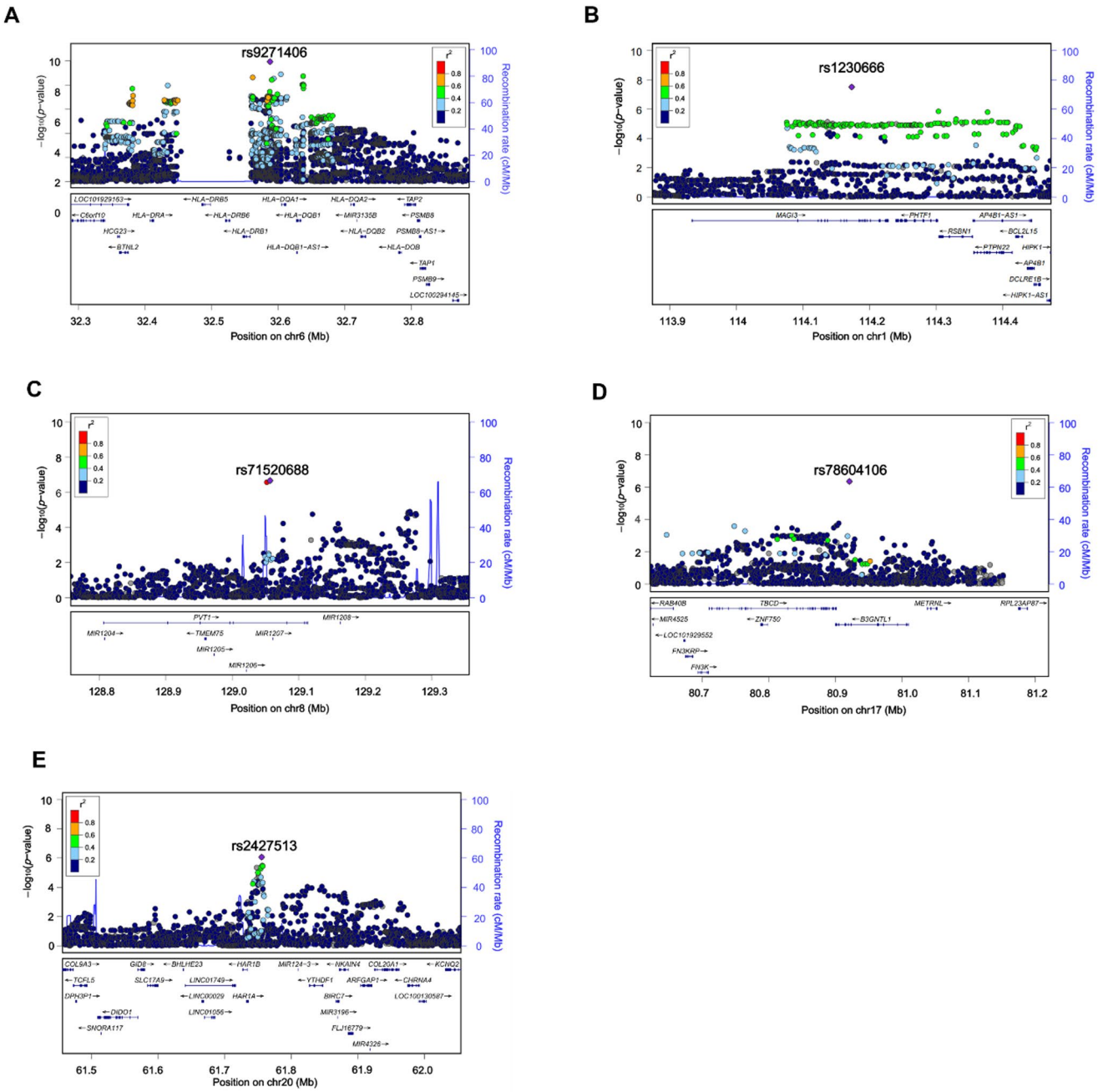

Fig. 2 Regional plots of five genomic susceptibility loci for HL. (**A**) 6p21.32 region (rs9271406), (**B**) 1p13.2 region (rs1230666), (**C**) 8q24.21 region (rs71520688), (**D**) 17q25.3 region (rs78604106), and (**E**) 20q13.33 region (rs2427513). SNPs located within 250 kb upstream or downstream of the lead SNPs (purple diamonds) were displayed with their -log10 transformed *P*-values (y-axis) relative to genomic position based on the hg19 reference (x-axis). Recombination rates (blue lines) and linkage disequilibrium (LD) patterns (color-coded from blue to red, corresponding to r^2 ranging from 0 to 1) were estimated using data from the 1000 Genomes Project of European ancestry

OTTERS and E-MAGMA discover sixteen susceptibility genes for HL

We performed gene-level association analysis using both OTTERS (Dai et al. 2023) and E-MAGMA (Gerring et al. 2021). A total of 19,127 genes were used for OTTERS analysis, and 14,500 genes were used for E-MAGMA. Sixteen genes showed statistical significance (FDR < 0.01 in both OTTERS and E-MAGMA) within the same eQTL dataset (Table 2, Table S4, Figure S2). These included HLA class II genes (*HLA-DQA1*, *HLA-DQA2*, *HLA-DQB1*, *HLA-DRB1*, *HLA-DRB5*, *HLA-DMA*, and *HLA-DPB1*, Figure S3,

Table 2 OTTERS and E-MAGMA jointly identified sixteen susceptibility genes for HL

Gene	Position [a]	eQTL dataset	E-MAGMA				OTTERS			
			P [b]	FDR [b]	P_{COJO} [c]	FDR_{COJO} [c]	P [d]	FDR [b]	P_{COJO} [e]	FDR_{COJO} [e]
BCL2L15	1:114,420,790–114,425,480	TwinsUK LCL	5.71×10^{-6}	5.00×10^{-3}	0.07	0.12	6.63×10^{-6}	1.35×10^{-3}	0.26	0.41
AIF1	6:31,582,961–31,583,880	Lepik 2017 blood	1.07×10^{-6}	1.30×10^{-3}	0.07	0.12	5.49×10^{-5}	3.18×10^{-3}	0.01	0.03
LSM2	6:31,765,173–31,769,967	GEUVADIS LCL	3.54×10^{-6}	4.18×10^{-3}	3.86×10^{-3}	0.02	9.00×10^{-7}	1.85×10^{-3}	4.21×10^{-3}	0.02
NOTCH4	6:32,162,620–32,177,232	GENCORD LCL	7.62×10^{-7}	8.37×10^{-4}	1.83×10^{-3}	0.01	2.56×10^{-4}	1.13×10^{-3}	0.01	0.03
TSBP1-AS1	6:32,223,488–32,228,552	TwinsUK LCL	2.34×10^{-7}	3.51×10^{-4}	0.02	0.08	1.04×10^{-4}	9.50×10^{-3}	0.09	0.19
HLA-DRB5	6:32,485,120–32,491,592	Lepik 2017 blood	1.47×10^{-7}	2.01×10^{-4}	0.05	0.12	7.68×10^{-5}	4.12×10^{-3}	0.10	0.19
HLA-DRB1	6:32,546,546–32,552,086	GTExV7 Whole blood	8.21×10^{-9}	8.69×10^{-5}	0.26	0.26	2.20×10^{-7}	2.45×10^{-4}	0.18	0.31
		TwinsUK blood	1.20×10^{-9}	4.03×10^{-6}	0.09	0.12	3.68×10^{-5}	4.15×10^{-3}	0.72	0.87
		Lepik 2017 blood	2.27×10^{-9}	8.29×10^{-6}	0.09	0.12	2.82×10^{-8}	1.07×10^{-5}	1.39×10^{-4}	1.08×10^{-3}
		GENCORD LCL	4.08×10^{-9}	9.82×10^{-6}	0.07	0.12	6.11×10^{-10}	2.83×10^{-6}	1.76×10^{-3}	7.72×10^{-3}
HLA-DQA1	6:32,595,956–32,605,398	TwinsUK blood	1.09×10^{-8}	1.83×10^{-5}	0.18	0.19	1.18×10^{-4}	8.82×10^{-3}	0.40	0.55
		GENCORD LCL	8.33×10^{-8}	1.10×10^{-4}	0.12	0.13	2.89×10^{-6}	1.22×10^{-3}	0.88	0.92
HLA-DQB1	6:32,627,244–32,631,702	TwinsUK blood	7.28×10^{-10}	4.03×10^{-6}	0.05	0.11	5.91×10^{-5}	5.88×10^{-3}	0.94	0.94
		GENCORD LCL	3.49×10^{-9}	9.82×10^{-6}	0.05	0.11	9.87×10^{-6}	2.86×10^{-3}	0.83	0.91
		TwinsUK LCL	7.32×10^{-11}	7.69×10^{-7}	0.03	0.10	1.93×10^{-5}	2.83×10^{-3}	0.82	0.91
HLA-DQA2	6:32,709,119–32,712,056	GEUVADIS LCL	3.57×10^{-9}	1.69×10^{-5}	0.09	0.12	3.39×10^{-10}	1.16×10^{-6}	0.07	0.16
		GENCORD LCL	4.47×10^{-9}	9.82×10^{-6}	0.08	0.12	1.76×10^{-13}	1.63×10^{-9}	1.89×10^{-4}	1.08×10^{-3}
TAP2	6:32,789,610–32,798,084	Lepik 2017 blood	6.06×10^{-10}	6.64×10^{-6}	0.09	0.12	3.87×10^{-6}	4.31×10^{-4}	0.37	0.54
PSMB9	6:32,811,913–32,819,638	Lepik 2017 blood	1.42×10^{-8}	3.11×10^{-5}	5.33×10^{-3}	0.02	2.22×10^{-4}	8.61×10^{-3}	0.03	0.07

Table 2 (continued)

Gene	Position [a]	eQTL dataset	E-MAGMA				OTTERS			
			P [b]	FDR [b]	P_{COJO} [c]	FDR_{COJO} [c]	P [d]	FDR [b]	P_{COJO} [e]	FDR_{COJO} [e]
HLA-DMA	6:32,916,390–32,926,631	TwinsUK LCL	4.13×10^{-7}	5.43×10^{-4}	9.22×10^{-4}	0.01	3.60×10^{-5}	4.12×10^{-3}	1.97×10^{-4}	1.08×10^{-3}
BRD2	6:32,936,437–32,942,860	Lepik 2017 blood	1.83×10^{-6}	2.00×10^{-3}	0.24	0.25	1.91×10^{-4}	7.83×10^{-3}	0.70	0.87
HLA-DPB1	6:33,043,703–33,049,341	TwinsUK LCL	4.84×10^{-6}	4.62×10^{-3}	2.61×10^{-5}	5.75×10^{-4}	2.95×10^{-5}	3.72×10^{-3}	9.08×10^{-6}	2.00×10^{-4}
AAR2	20:34,824,381–34,841,611	TwinsUK blood	6.32×10^{-6}	6.07×10^{-3}	-	-	1.50×10^{-5}	2.09×10^{-3}	-	-
		TwinsUK LCL	6.96×10^{-6}	5.63×10^{-3}	-	-	1.62×10^{-5}	2.56×10^{-3}	-	-

[a]: The chromosome, start, and end position of the susceptibility genes according to the hg19 reference

[b]: *P* values and Benjamini-Hochberg correction *P* values for the E-MAGMA analysis

[c]: *P* values and Benjamini-Hochberg correction *P* values for the E-MAGMA analysis after conditioning on the previously reported GWAS lead variants of HL

[d]: *P* values and Benjamini-Hochberg correction *P* values for the OTTERS analysis

[e]: *P* values and Benjamini-Hochberg correction *P* values for the OTTERS analysis after conditioning on the previously reported GWAS lead variants of HL

Figure S4) and non-HLA genes involved in immune-related pathways (*AIF1*, *PSMB9* and *TAP2*, Figure S5). Additionally, genes associated with RNA processing (*AAR2* and *LSM2*, Figure S6), apoptosis (*BCL2L15*, Figure S7), signal transduction (*NOTCH4*, FigureS8), transcriptional regulation (*BRD2*, Figure S9), and other biological functions (*TSBP1-AS1*, Figure S10) were identified. The association between lead SNPs and target gene expression is detailed in Table S5. At the HLA locus, the observed effects were driven by the lead SNP or its LD proxy SNP with $P_{GWAS} < 1.0\times10^{-6}$ and showed statistical significance with the expression of target genes ($P_{eQTL} < 0.05$). The lead SNP rs1230666 at 1p13.2 was significantly associated with *BCL2L15* expression in TwinsUK LCL ($P_{eQTL} = 7.03\times10^{-9}$).

Conditional analyses ascertain five new susceptibility genes for HL

After conditional analysis of previously reported independent HL lead SNPs, most of the sixteen HL susceptibility genes did not show statistical significance in gene-level association analysis ($FDR_{COJO} < 0.05$ in both OTTERS and E-MAGMA), except for *HLA-DMA* ($FDR_{COJO} = 0.01$ for E-MAGMA and $FDR_{COJO} = 1.08\times10^{-3}$ for OTTERS in the eQTL Catalogue TwinsUK LCL), *HLA-DPB1* ($FDR_{COJO} = 5.75\times10^{-4}$ for E-MAGMA and $FDR_{COJO} = 2.00\times10^{-4}$ for OTTERS in the eQTL Catalogue TwinsUK LCL), *NOTCH4* ($FDR_{COJO} = 0.01$ for E-MAGMA and $FDR_{COJO} = 0.03$ for OTTERS in the eQTL Catalogue GENCORD LCL) and *LSM2* ($FDR_{COJO} = 0.02$ for E-MAGMA and $FDR_{COJO} = 0.02$ for OTTERS in the eQTL Catalogue GEUVADIS LCL), suggesting that the published GWAS lead SNPs may explain most gene associations, while *HLA-DMA*, *HLA-DPB1*, *NOTCH4*, and *LSM2* are conditionally independent of the reported SNPs (Table 2). Additionally, *AAR2* was identified as a novel susceptibility gene in our study since no susceptibility variants have been previously reported at 20q11.23.

HL susceptibility genes are primarily enriched in immune-related functions

We employed ClusterProfiler to conduct gene set enrichment analysis on sixteen HL genetic susceptibility genes and found that they were significantly enriched in 113 GO gene sets, 24 KEGG pathways (FDR < 0.05, Enrichment Ratio ≥ 5%), most of which were immune response-related pathways. In the GO gene sets, significant terms included peptide antigen binding (GO: 0042605, $FDR = 2.00\times10^{-16}$, Enrichment Ratio = 53.33%), MHC class II protein complex (GO: 0042613, $FDR = 3.22\times10^{-17}$, Enrichment Ratio = 46.67%), and antigen processing and presentation of exogenous peptide antigen (GO: 0002478, $FDR = 1.54\times10^{-16}$, Enrichment Ratio = 53.33%) (Fig. 3A, Table S6). The KEGG significantly enriched pathways included antigen processing and presentation (hsa04612, $FDR = 9.38\times10^{-14}$, Enrichment Ratio = 72.73%), Asthma (hsa05310, $FDR = 6.17\times10^{-14}$, Enrichment Ratio = 63.64%) and Epstein-Barr virus

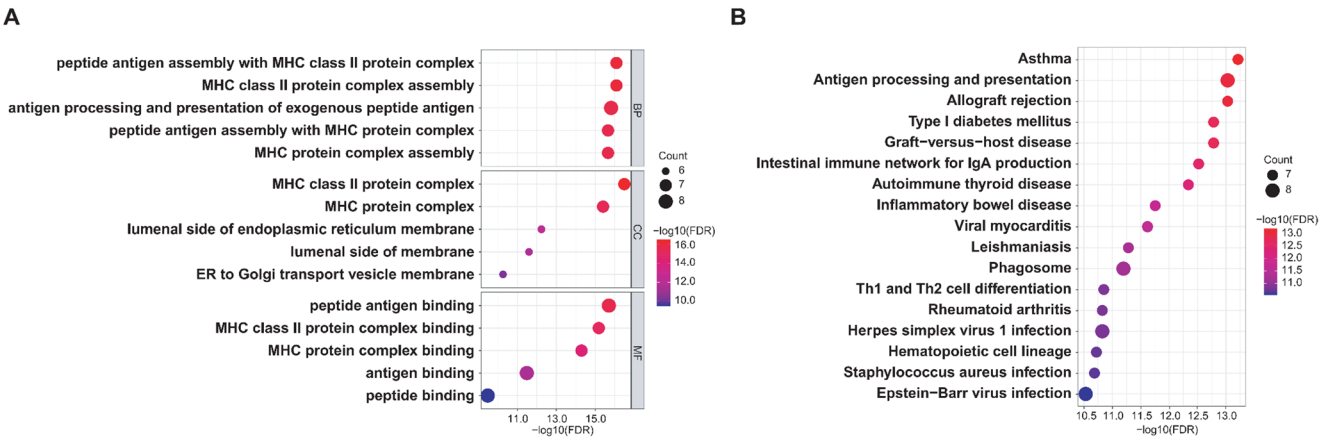

Fig. 3 GO and KEGG enrichment analysis of susceptibility genes for HL. (**A**) Bubble chart for GO enrichment analysis. (**B**) Bubble chart for KEGG enrichment analysis. Sixteen significant genes (FDR < 0.01) identified by both E-MAGMA and OTTERS analysis were used in the enrichment analysis. GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes

infection (hsa05169, FDR = 2.94×10^{-11}, Enrichment Ratio = 72.73%), along with other immune-related disease pathways (Fig. 3B, Table S7).

Discussion

GWASs have identified numerous susceptibility loci for HL, but applying the discoveries to functional or therapeutic contexts remains challenging. In this research, we performed a meta-GWAS for HL followed by an extensive gene-level association analysis, revealing sixteen susceptibility genes primarily involved in immune-related biological pathways. Among these genes, *HLA-DMA*, *HLA-DPB1*, *NOTCH4*, *AAR2* and *LSM2* were identified as novel susceptibility genes independent of previously reported GWAS signals.

We discovered two novel HLA class II genes, *HLA-DMA* and *HLA-DPB1*, as new susceptibility genes for HL. After adjusting for previously identified GWAS lead SNPs, the expression of *HLA-DMA* and *HLA-DPB1* remained significantly associated with HL risk. We also replicated previous findings of some HLA class II genes, such as *HLA-DQA1*, *HLA-DQB1*, *HLA-DRB1*, and *HLA-DRB5* (An et al. 2023). Classical HLA class II molecules (HLA-DR, -DP, and -DQ), expressed on antigen-presenting cells, trigger adaptive immunity by presenting exogenous antigens to $CD4^+$ T cells (Horton et al. 2004). Non-classical HLA class II molecules, like HLA-DM, are primarily involved in facilitating the binding of exogenous antigenic peptides to HLA class II molecules (Morris et al. 1994). The resulting antigen-HLA II complex is transported to the cell surface, where it interacts with $CD4^+$ T-cell receptors (TCRs), thereby initiating the immune response (Neefjes et al. 2011). HL is a lymphoid neoplasm originating from germinal center B cells, distinguished by Hodgkin/Reed-Sternberg cells surrounded by numerous reactive immune cells, including $CD4^+$ T follicular helper cells. Hodgkin/Reed-Sternberg cells rely heavily on their microenvironment for survival (Greaves et al. 2013; Küppers 2009). The identified HL-associated SNP rs9271406 at the HLA locus may influence HL risk by regulation of HLA class II gene expression, which might affect exogenous antigen presentation and alter the interaction between $CD4^+$ T follicular helper cells and germinal center B cells and thereby contribute to the development HL (Sud et al. 2017a).

Moreover, we identified *BCL2L15* as a susceptibility gene for HL, which was regulated by lead SNP rs1230666 at 1p13.2, a region previously linked to HL only at the SNP level (Sud et al. 2018). After adjusting for the reported SNP rs2476601 at this region, the association of *BCL2L15* expression with HL was no longer significant, indicating that the relationship between *BCL2L15* and HL risk might be captured by rs2476601. Furthermore, rs1230666 identified in this study was in linkage disequilibrium with rs2476601 ($r^2 = 0.55$), supporting the hypothesis that these variants may influence HL risk through a shared regulatory network that modulates the expression of *BCL2L15*. *BCL2L15* is a pro-apoptotic gene within the Bcl-2 protein family (Coultas et al. 2003; Pavlou et al. 2012). Pro-apoptotic and anti-apoptotic elements in the Bcl-2 family interact intricately to determine whether B cells survive or undergo apoptosis (Adams et al. 2018; Perini et al. 2018). Lower expression of certain pro-apoptotic regulators is a common feature among various B cell lymphoma subtypes (Ashkenazi et al. 2017). Therefore, we hypothesize that *BCL2L15* may be associated with HL risk by participating in the regulation of B-cell apoptosis.

There are several limitations in our study. Firstly, we focused solely on European descent, which may constrain the generalization of our findings to other ethnic populations.

Additional research involving various demographic groups is essential to confirm wider applicability and relevance. Although the GWAS summary data we obtained were corrected using SAIGE to mitigate potential biases due to case-control imbalance, future investigations with a larger sample size of cases are essential to confirm and discover susceptibility regions. Moreover, while our computational findings are solid, experimental confirmation is necessary to establish the biological relevance of the genes and pathways linked to HL pathogenesis. Functional studies will be critical for advancing targeted therapies and precision medicine.

Conclusions

Overall, we conducted a meta-GWAS followed by gene-level association analysis to identify HL susceptibility genes. Sixteen susceptibility genes were identified, including five novel genes (*HLA-DMA*, *HLA-DPB1*, *NOTCH4*, *AAR2*, and *LSM2*) independent of known GWAS signals. These results underscore the importance of the exogenous antigen presentation pathway in HL development, offering insights into potential mechanisms.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00432-025-06224-8.

Acknowledgements This research has been conducted using the UK Biobank Resource under Application Number 58450 and FinnGen. We are grateful for the efforts of the participants and researchers in making this work feasible. We also appreciate the eQTL Catalogue and the GTEx for sharing the eQTL summary data.

Author contributions Wen-Hui Jia: Formal analysis, visualization, validation, methodology, writing-original draft. Chang-Ling Huang: Data curation, methodology, formal analysis. Wen-Li Zhang: Data curation, methodology, formal analysis. Yong-Qiao He: Data curation. Wen-Qiong Xue: Data curation. Ying Liao: Data curation. Zhi-Yang Zhao: Data curation. Meng-Xuan Yang: Data curation. Lu Pei: Data curation. Wei-Hua Jia: Conceptualization, resources, supervision, project administration, writing-review and editing. Tong-Min Wang: Conceptualization, supervision, validation, methodology, writing-review and editing. All authors have read and agreed to the published version of the manuscript.

Funding This study was funded by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0501000); the National Natural Science Foundation of China (82373656, 82273705, 82473703, 82404339); the Science and Technology Planning Project of Guangzhou, China (2024A04J4560, 2024A04J00693); the Young Talent Support Project of Guangzhou Association for Science and Technology (QT2024-030); the Fundamental Research Funds for the Central Universities, Sun Yat-sen University (24ykqb002, 24qnpy292); the Cancer Innovative Research Program of Sun Yat-sen University Cancer Center (CIRP-SYSUCC-0017); the Young Talents Program of Sun Yat-sen University Cancer Center (YTP-SYSUCC-0076, YTP-SYSUCC-0081); the Young Science and Technology Talent Support Program of Guangdong Precision Medicine Application Association (YSTTGDPMAA202502); the Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, China.

Data availability No datasets were generated or analysed during the current study.

Declarations

Ethical approval The North West Multi-Centre Research Ethics Committee (MREC) has approved UK Biobank as a Research Tissue Bank (RTB) (21/NW/0157), allowing researchers to proceed without additional ethical approval. The FinnGen data used in this study were obtained from publicly available GWAS summary datasets, approved by the relevant ethics committees.

Inform consent All the participants involved in this study have provided written informed permission.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Adams CM, Clark-Garvey S, Porcu P, Eischen CM (2018) Targeting the Bcl-2 family in B cell lymphoma. Front Oncol 8:636. https://doi.org/10.3389/fonc.2018.00636

An Y, Lee C (2023) Identification and interpretation of eQTL and eGenes for hodgkin lymphoma susceptibility. Genes (Basel) 14. https://doi.org/10.3390/genes14061142

Ashkenazi A, Fairbrother WJ, Leverson JD, Souers AJ (2017) From basic apoptosis discoveries to advanced selective BCL-2 family inhibitors. Nat Rev Drug Discovery 16:273–284. https://doi.org/10.1038/nrd.2016.253

Battle A, Brown CD, Engelhardt BE, Montgomery SB (2017) Genetic effects on gene expression across human tissues. Nature 550:204–213. https://doi.org/10.1038/nature24277

Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74:229–263. https://doi.org/10.3322/caac.21834

Carbone A, Gloghini A, Caruso A, De Paoli P, Dolcetti R (2017) The impact of EBV and HIV infection on the microenvironmental niche underlying hodgkin lymphoma pathogenesis. Int J Cancer 140:1233–1245. https://doi.org/10.1002/ijc.30473

Chen C, Song N, Dong Q et al (2022) Association of Single-Nucleotide variants in the human leukocyte antigen and other loci with childhood hodgkin lymphoma. JAMA Netw Open 5:e2225647. https://doi.org/10.1001/jamanetworkopen.2022.25647

Connors JM, Cozen W, Steidl C, Carbone A, Hoppe RT, Flechtner HH, Bartlett NL (2020) Hodgkin lymphoma. Nat Rev Dis Primers 6:61. https://doi.org/10.1038/s41572-020-0189-6

Coultas L, Pellegrini M, Visvader JE et al (2003) Bfk: a novel weakly proapoptotic member of the Bcl-2 protein family with a BH3 and a BH2 region. Cell Death Differ 10:185–192. https://doi.org/10.1038/sj.cdd.4401204

Cozen W, Li D, Best T et al (2012) A genome-wide meta-analysis of nodular sclerosing hodgkin lymphoma identifies risk loci at 6p21.32. Blood 119:469–475. https://doi.org/10.1182/blood-2011-03-343921

Cozen W, Timofeeva MN, Li D et al (2014) A meta-analysis of hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. Nat Commun 5:3856. https://doi.org/10.1038/ncomms4856

Dai Q, Zhou G, Zhao H et al (2023) OTTERS: a powerful TWAS framework leveraging summary-level reference data. Nat Commun 14:1271. https://doi.org/10.1038/s41467-023-36862-w

de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015) MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 11:e1004219. https://doi.org/10.1371/journal.pcbi.1004219

Enciso-Mora V, Broderick P, Ma Y et al (2010) A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). Nat Genet 42:1126–1130. https://doi.org/10.1038/ng.696

Frampton M, da Silva Filho MI, Broderick P et al (2013) Variation at 3p24.1 and 6q23.3 influences the risk of Hodgkin's lymphoma. Nat Commun 4:2549. https://doi.org/10.1038/ncomms3549

Ge T, Chen CY, Ni Y, Feng YA, Smoller JW (2019) Polygenic prediction via bayesian regression and continuous shrinkage priors. Nat Commun 10:1776. https://doi.org/10.1038/s41467-019-09718-5

Gerring ZF, Mina-Vargas A, Gamazon ER, Derks EM (2021) E-MAGMA: an eQTL-informed method to identify risk genes using genome-wide association study summary statistics. Bioinformatics 37:2245–2249. https://doi.org/10.1093/bioinformatics/btab115

Greaves P, Clear A, Owen A et al (2013) Defining characteristics of classical hodgkin lymphoma microenvironment T-helper cells. Blood 122:2856–2863. https://doi.org/10.1182/blood-2013-06-508044

Horton R, Wilming L, Rand V et al (2004) Gene map of the extended human MHC. Nat Rev Genet 5:889–899. https://doi.org/10.1038/nrg1489

Kharazmi E, Fallah M, Pukkala E et al (2015) Risk of Familial classical hodgkin lymphoma by relationship, histology, age, and sex: a joint study from five nordic countries. Blood 126:1990–1995. https://doi.org/10.1182/blood-2015-04-639781

Küppers R (2009) The biology of Hodgkin's lymphoma. 9:15–27. https://doi.org/10.1038/nrc2542

Kurki MI, Karjalainen J, Palta P et al (2023) FinnGen provides genetic insights from a well-phenotyped isolated population. Nature 613:508–518. https://doi.org/10.1038/s41586-022-05473-8

Kushekhar K, van den Berg A, Nolte I, Hepkema B, Visser L, Diepstra A (2014) Genetic associations in classical hodgkin lymphoma: a systematic review and insights into susceptibility mechanisms. Cancer Epidemiol Biomarkers Prev 23:2737–2747. https://doi.org/10.1158/1055-9965.Epi-14-0683

Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X (2019) ACAT: A fast and powerful P value combination method for Rare-Variant analysis in sequencing studies. Am J Hum Genet 104:410–421. https://doi.org/10.1016/j.ajhg.2019.01.002

Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC (2017) Polygenic scores via penalized regression on summary statistics. Genet Epidemiol 41:469–480. https://doi.org/10.1002/gepi.22050

Morris P, Shaman J, Attaya M et al (1994) An essential role for HLA-DM in antigen presentation by class II major histocompatibility molecules. Nature 368:551–554. https://doi.org/10.1038/368551a0

Neefjes J, Jongsma ML, Paul P, Bakke O (2011) Towards a systems Understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11:823–836. https://doi.org/10.1038/nri3084

Pavlou M-AS, Kontos CKJhAo (2012) BCL2L15 (BCL2-like 15). Atlas Genet Cytogenet Oncol Haematol 16:115–118. https://doi.org/10.4267/2042/46940

Perini GF, Ribeiro GN, Pinto Neto JV, Campos LT, Hamerschlak N (2018) BCL-2 as therapeutic target for hematological malignancies. J Hematol Oncol 11:65. https://doi.org/10.1186/s13045-018-0608-2

Privé F, Vilhjálmsson BJ, Aschard H, Blum MGB (2019) Making the most of clumping and thresholding for polygenic scores. Am J Hum Genet 105:1213–1221. https://doi.org/10.1016/j.ajhg.2019.11.001

Ribeiro AG, Vermeulen R, Cardoso MRA, Latorre M, Hystad P, Downward GS, Nardocci AC (2021) Residential traffic exposure and Lymphohematopoietic malignancies among children in the City of São Paulo, Brazil: an ecological study. Cancer Epidemiol 70:101859. https://doi.org/10.1016/j.canep.2020.101859

Sud A, Thomsen H, Law PJ et al (2017a) Genome-wide association study of classical hodgkin lymphoma identifies key regulators of disease susceptibility. Nat Commun 8:1892. https://doi.org/10.1038/s41467-017-00320-1

Sud A, Hemminki K, Houlston RS (2017b) Candidate gene association studies and risk of hodgkin lymphoma: a systematic review and meta-analysis. Hematol Oncol 35:34–50. https://doi.org/10.1002/hon.2235

Sud A, Thomsen H, Orlando G et al (2018) Genome-wide association study implicates immune dysfunction in the development of hodgkin lymphoma. Blood 132:2040–2052. https://doi.org/10.1182/blood-2018-06-855296

Sudlow C, Gallacher J, Allen N et al (2015) UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 12:e1001779. https://doi.org/10.1371/journal.pmed.1001779

Urayama KY, Jarrett RF, Hjalgrim H et al (2012) Genome-wide association study of classical hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. J Natl Cancer Inst 104:240–253. https://doi.org/10.1093/jnci/djr516

Wang G, Sarkar A, Carbonetto P, Stephens M (2020) A simple new approach to variable selection in regression, with application to genetic fine mapping. J Royal Stat Soc Ser B Stat Methodol 82:1273–1300. https://doi.org/10.1111/rssb.12388

Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26:2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Yang J, Ferreira T, Morris AP et al (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet 44:369–375 s361-363. https://doi.org/10.1038/ng.2213

Yu G, Wang LG, Han Y, He QY (2012) ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16:284–287. https://doi.org/10.1089/omi.2011.0118

Zhou G, Zhao H (2021) A fast and robust bayesian nonparametric method for prediction of complex traits using summary statistics. PLoS Genet 17:e1009697. https://doi.org/10.1371/journal.pgen.1009697

Zhou W, Nielsen JB, Fritsche LG et al (2018) Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 50:1335–1341. https://doi.org/10.1038/s41588-018-0184-y

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.