Abstract

IMPORTANCE Studies focusing on genetic susceptibility of childhood Hodgkin lymphoma (HL) are limited.

OBJECTIVES To identify genetic variants associated with childhood-onset HL vs adult-onset HL.

DESIGN, SETTING, AND PARTICIPANTS This genetic association study was performed with 3 cohorts: the St Jude Lifetime Cohort Study (SJLIFE), initiated in 2007 with ongoing follow-up, and the original and expansion cohorts of the Childhood Cancer Survivor Study (CCSS), initiated in the 1990s with ongoing follow-up. Results of these genome-wide association studies (GWASs) were combined via meta-analysis. Data were analyzed from June 2021 to June 2022.

MAIN OUTCOMES AND MEASURES Childhood HL was the focused outcome. Single-nucleotide variant (SNV, formerly single-nucleotide polymorphism) array genotyping and imputation were conducted for the CCSS original cohort, and whole-genome sequencing was performed for the SJLIFE and CCSS expansion cohort.

RESULTS A total of 1286 HL cases (mean diagnosis [SD] age, 14.6 [3.9] years), 6193 non-HL childhood cancer cases, and 369 noncancer controls, all of European ancestry, were included in the analysis. Using step-wise conditional logistic regression, the odds ratios (ORs) for each of the 3 independent SNVs identified in the human leukocyte antigen (HLA) locus were 1.80 (95% CI, 1.59-2.03; \( P = 2.14 \times 10^{-21} \)) for rs28383311, 1.53 (95% CI, 1.37-1.70; \( P = 2.05 \times 10^{-14} \)) for rs3129198, and 1.51 (95% CI, 1.35-1.69; \( P = 6.21 \times 10^{-13} \)) for rs3129890. Further HLA imputation revealed 9 alleles and 55 amino acid changes that potentially conferred HL susceptibility. In addition, 5 non-HLA loci were identified: (1) rs1432297 (OR, 1.29; 95% CI, 1.18-1.41; \( P = 2.5 \times 10^{-10} \); \( r^2 = 0.55 \); \( D^2 = 0.75 \) with previously reported rs1432295, REL); (2) rs2757647 (OR, 1.30; 95% CI, 1.18-1.42; \( P = 3.5 \times 10^{-10} \); \( r^2 = 0.59 \); \( D^2 = 0.83 \) with previously reported rs6928977, AHI1); (3) rs13279159 (OR, 1.33; 95% CI, 1.20-1.47; \( P = 1.7 \times 10^{-8} \); \( r^2 = 0.75 \); \( D^2 = 1.00 \) with previously reported rs2019960, PVT1); (4) rs3824662 (OR, 1.52; 95% CI, 1.33-1.73; \( P = 3.9 \times 10^{-10} \); \( r^2 = 0.91 \); \( D^2 = 1.00 \) with previously reported rs3781093, GATA3); and (5) rs17953624 (OR, 1.98; 95% CI, 1.56-2.51; \( P = 1.5 \times 10^{-8} \); minor allele frequency, 0.02), a novel uncommon SNV mapped to PDGFD. Twelve of 18 previously reported genome-wide significant non-HLA SNVs (67%) were replicated with statistically significant results.

CONCLUSIONS AND RELEVANCE In this genetic association study, a predominantly common and potentially unique genetic etiology was found between childhood-onset and adulthood-onset HL.

Key Points

Question What genetic variants or genes are associated with the risk of childhood Hodgkin lymphoma?

Findings In this genetic association study nested within 3 childhood cancer survivor cohorts, 3 independent single-nucleotide variants (SNVs, formerly single-nucleotide polymorphisms) in the human leukocyte antigen (HLA) locus and 5 loci in non-HLA regions were found to be significantly associated with risk of childhood-onset Hodgkin lymphoma. Two-thirds of previously reported Hodgkin lymphoma–associated non-HLA SNVs were replicated in the analysis.

Meaning These findings suggest predominantly common and potentially unique genetic etiology between childhood-onset and adulthood-onset Hodgkin lymphoma.

Introduction

Hodgkin lymphoma (HL) originates from B lymphocytes and is characterized by the presence of Reed-Sternberg cells and their variants. The remarkably high cure rate of 87% is attributable to advances in therapeutics featuring multiagent chemotherapy with or without radiotherapy. HL is most frequently diagnosed among people aged 20 to 34 years (30.7%); children and young adults (<20 years) account for 12.2% of patients with HL. Despite its low incidence in children younger than 5 years, HL is the most common cancer among adolescents and young adults aged 15 to 19 years, and it has a favorable 5-year survival, exceeding 95%. However, late effects of chemotherapy and radiotherapy are common after the cure of childhood-onset HL, resulting in accelerated aging and excess late mortality.

HL is classified into histological subtypes based on cellular composition and architecture; the 2 major subtypes are classical HL (approximately 95% of cases) and nodular lymphocyte predominant HL (approximately 5% of cases). Increased familial risk and high concordance between monozygotic twins provide evidence of genetic susceptibility to HL. In addition to genetic risk factors, environmental risk factors, such as Epstein-Barr virus (EBV) infection, are also causally related to specific HL subtypes.

The most significant genetic association for HL is mapped to the human leukocyte antigen (HLA) region. HLA genes serve as the fundamental platform for immune surveillance and responsiveness to internal and external antigens and are implicated in a number of human diseases. HLA class I molecules present antigens primarily to cytotoxic T-cells to kill target cells, and HLA class II molecules stimulate antibody production in response to specific antigens. Thus, HLA genes are crucial for the development of autoimmune diseases, infectious diseases, and cancers. HL was the first human disease associated with HLA genes, and subsequent studies revealed that HLA alleles along with specific single-nucleotide variants (SNVs, formerly single-nucleotide polymorphisms; eg, rs6903608) are significantly associated with HL risk. In addition to the strong association between HLA and HL, genetic variants in non-HLA regions have been reported, including rs4459895 (LPP [OMIM 600700]) and rs1432295 (REL [OMIM 164910]), among others.

Although the genetic etiology of HL has been investigated, studies among patients with childhood-onset HL are lacking. Research using cancer survivors to investigate cancer etiology is subject to potential survival bias, but the very favorable survival for pediatric HL makes this a reasonable population for consideration of such investigation. In a recent study, we demonstrated that germline pathogenic variants in BRCA2 are associated with risk of pediatric or adolescent non-HL. Here, we followed a similar paradigm to leverage genetic data of HL survivors to identify new genetic variants in HLA and non-HLA regions that are associated with the risk of childhood-onset HL.

Methods

Study Populations

Three large cohorts of survivors of childhood cancer were included in the current genetic association study. The St Jude Lifetime Cohort Study (SJLIFE) was approved by the St Jude institutional review board (IRB), and the Childhood Cancer Survivor Study (CCSS) was approved by the IRB of each participating center. This study is covered by the IRB approval for the 2 cohort studies. The SJLIFE is a retrospective cohort study initiated in 2007 with prospective clinical follow-up of 5-year survivors of childhood cancer treated at the St Jude Children's Research Hospital. The CCSS is a retrospective cohort study of 5-year survivors of childhood cancer diagnosed and treated at 31 institutions in North America. The CCSS includes an original cohort diagnosed between 1970 and 1986 and an expansion cohort diagnosed between 1987 and 1999. In both studies, self-report or proxy-report questionnaires were used to assess demographic characteristics; clinical information was abstracted from medical records.
Genotype Data

SNV-array genotyping was conducted using the HumanOmni5M-4 version 1 (Illumina) for 5739 childhood cancer survivors in the CCSS original cohort, followed by genome-wide imputation using the Human Reference Consortium Panel as the reference on Michigan imputation server (dbGaP accession: phs001327.v2.p1). Whole-genome sequencing (WGS) with 30 folds was performed on DNA derived from peripheral blood samples from 4853 SJLIFE participants, including 4402 childhood cancer survivors and 451 noncancer controls, and buccal/saliva samples from 2998 survivors in the CCSS expansion cohort. The entire collection of WGS data for the SJLIFE and CCSS expansion cohorts is accessible at the St Jude Cloud. All participants providing biological material for DNA analysis provided written informed consent.

Quality Control

Of the 5739 survivors in the CCSS original cohort, 4671 remained after exclusion of 416 individuals of non-European ancestry whose admixture coefficient for 1000 Genomes CEU population was less than 80% and 652 who were also enrolled in SJLIFE. The final CCSS original cohort set consisted of 631 HL cases and 4040 non-HL cases (ie, survivors who had any childhood cancer diagnosis except HL) (eFigure 1 in the Supplement). Only variants (n = 8.4 million) with minor allele frequency (MAF) greater than 0.01, Hardy-Weinberg Equilibrium (HWE) \( P > 1 \times 10^{-6} \), and having passed other filtering (ie, applying--keep-filtered = PASS setting in vcftools command) were retained for association analysis.

Among the 4402 SJLIFE survivors, 467 were diagnosed with HL. A total of 451 community controls with no history of childhood cancer were included and frequency matched to survivors by age, sex, and race. We subsequently excluded individuals (HL cases and controls) per the following criteria: (1) non-European ancestry with admixture coefficient for the 1000 Genomes CEU population less than 80% (86 case participants; 58 control participants); (2) incorrect or incomplete clinical characteristics (5 control participants); (3) individuals showing cryptic relatedness with another study participant (1 case participant; 19 control participants). Finally, a total of 380 HL cases and 369 controls remained for further analysis (eFigure 1 in the Supplement).

Of the 2998 survivors in the CCSS expansion cohort, 2428 were available for further analyses based on the following exclusions: (1) failure of the coverage and mapping quality control (50 individuals); (2) unexpected duplicates (12 individuals); (3) samples with excessive heterozygosity (97 individuals); and (4) individuals of non-European ancestry with admixture coefficient for 1000 Genomes CEU population less than 80% (411 individuals). Among the 2428 survivors, 275 were diagnosed with HL and considered cases; the 2153 survivors with other diagnoses (non-HL) were considered controls for comparison (eFigure 1 in the Supplement).

Additional quality control was performed to filter low-quality variants derived from WGS data, including the following criteria: (1) minimum genotype quality score of 20; (2) minimum depth of 5; (3) minimum mean depth of 10; (4) HWE \( P > 1 \times 10^{-5} \); (5) maximum missing rate of 10% across all samples; (6) MAF greater than 0.01; and (7) exclusion of nonbiallelic loci. A total of 6.9 million SNVs and small insertions and deletions were in common with 8.4 million genotyped or imputed variants from the CCSS original cohort.

Statistical Analysis

Genome-wide association study (GWAS) analyses were performed in each of the 3 cohorts (SJLIFE, CCSS original, and CCSS expansion). A logistic regression model was analyzed for associations between HL cancer status and allelic dose of each variant (0, 1, or 2), with adjustments for sex and the top principal components derived from genotypes. Fixed-effects meta-analysis was used to combine the summary statistics from the 3 GWAS analyses. Heterogeneity between data sets was assessed using \( \chi^2 \) as well as \( P \) for heterogeneity, calculated from the Cochran Q statistic. Variants with a meta-GWAS \( P < 5 \times 10^{-8} \) were deemed genome-wide significant.
To refine associations between HL and HLA regions, step-wise conditional logistic regression was used to identify independent SNVs. First, the top (most significant) SNV from meta-GWAS was added to the list of covariates, and conditional association analysis was performed in each cohort. Then, results of the 3 conditional analyses were combined via meta-analysis to identify the next top SNV (independent secondary signal). This process was repeated until no SNV in the meta-analysis had a $P < 5 \times 10^{-8}$.

PLINK (version 1.90b) was used for association analysis. Other statistical analyses were performed with R version 3.6.1 (R Project for Statistical Computing), and 2-sided $P \leq .05$ was considered statistically significant. Manhattan plots, regional SNV association results, and quantile-quantile plots were generated with LocusZoom. Pairwise linkage disequilibrium (LD) values were obtained by choosing individuals of European ancestry from the 1000 genomes reference panel using LDlink.

Direct imputing summary association statistics of HLA variants (DISH) was used to infer statistical associations between HL risk and HLA alleles, amino acids, and SNVs based on all the variants cataloged in the Type 1 Diabetes Genetics Consortium reference panel comprising 5225 individuals of European ancestry. Biallelic SNVs (n = 28 455) with meta-GWAS summary statistics in the HLA region (Genome Reference Consortium Human Build 38, chr6:28510120-33480577) were fed into the program. Stringent quality control was applied to filter the imputed variants (MAF $\geq 0.005$; imputation score $r^2 \geq 0.5$) after the DISH imputation.

**Results**

**Meta-analysis of 3 GWAS Data Sets**

HL cases were diagnosed between 3.0 and 22.7 years of age (mean [SD] age, 14.6 [3.9] years). The summary statistics from the 3 GWASs were combined with a fixed-effects meta-analysis encompassing 1286 HL cases, 369 noncancer controls, and 6824 other childhood cancer cases, for a total of 6.88 million SNVs. We observed little evidence of genomic inflation by quantile-quantile plot of the meta-GWAS results ($\lambda = 1.021$) (eFigure 2 in the Supplement), suggesting little population stratification. The overall landscape of genome-wide associations and significant peaks are illustrated in Figure 1.

**Figure 1. Manhattan Plot for Genome-Wide Associations With Hodgkin Lymphoma**

The overall landscape for meta-analysis of 3 genome-wide association studies are shown. The $-\log_{10} P$ values were plotted against the genomic position (x-axis), ie, chromosome and chromosomal position based on Genome Reference Consortium Human Build 38. The black horizontal dotted line represents the genome-wide significance threshold of $P = 5 \times 10^{-8}$. Genome-wide significant peaks were marked at 2p16.1 (REL), 6q21.3 (HLA-DQA1), 6q23.3 (AHI1), 8q24.21 (PVT1), 10p14 (GATA3), and 11q22.3 (PDGFD).
SNVs in the HLA Region

A total of 691 SNVs in the HLA region were significantly associated with HL risk. The strongest association was marked by rs28383311 (OR, 1.80; 95% CI, 1.59-2.03; \( P = 2.14 \times 10^{-21} \)) (Table 1), in the neighborhood of HLA-DRB1 (OMIM 142857), HLA-DQA1 (OMIM 146880), and HLA-DQB1 (OMIM 604305) genes (Figure 2A), which had a moderate LD \( (r^2 = 0.46, D' = 1.00) \) with previously reported SNV rs2858870 near HLA-DRA gene in adolescent and young adult-onset HL.29

After conditioning on rs28383311, the secondary strongest association peak was marked by rs3129198 (OR, 1.53; 95% CI, 1.37-1.70; conditional \( P = 2.05 \times 10^{-14} \)) (Table 1), near HLA-DPA1 (OMIM 142880) and HLA-DPB1 (OMIM 142858) genes (Figure 2B). There was no LD \( (r^2 = 0.00, D' = 0.07) \) between rs3129198 and rs28383311 (eFigure 3 in the Supplement). However, rs3129198 had a perfect LD \( (r^2 = 1.0, D' = 1.00) \) with rs2281389 (HLA-DPB1) in a previous study comprising mostly adulthood-onset HL.17

In the next iteration of step-wise conditional logistic regression, we added rs3129198 and rs28383311 as covariates for adjustments in the model. A tertiary association signal emerged and was marked by the strongest associated SNV rs3129890 (OR = 1.51; 95% CI, 1.35-1.69; \( P = 6.21 \times 10^{-13} \)) (Table 1) in the neighborhood of HLA-DRA (OMIM 142860) (Figure 2C). The rs3129890 had a very low LD with the primary SNV rs28383311 \( (r^2 = 0.06, D' = 0.89) \) or the secondary SNV rs3129198 \( (r^2 = 0.01, D' = 0.12) \) (eFigure 3 in the Supplement) and a low LD with rs2395185 \( (r^2 = 0.15, D' = 1.00) \) near the HLA-DRA gene, which was reported previously by a study among HL cases diagnosed across a wide age range (16-80 years).30 Notably, rs28383311 and rs3129198 are mapped to HLA class II, whereas rs3129890 is mapped to the juncture between HLA class II and III. Previously reported HL-associated SNVs in the HLA region are provided in the eTable 1 in the Supplement. In addition, 10 of 14 previously reported genome-wide significant HLA SNVs (71%) were replicated with \( P < .05 \). Among the replicated SNVs, 8 of 10 (80%) were mapped to HLA class II; rs2248462 (MICB [OMIM 602436]) is mapped to the HLA-I, and rs204999 (PRRT1 [OMIM 618297]) is mapped to the HLA III. Most (3 of 4 [75%]) of the remaining insignificant SNVs were mapped to the HLA class I region.

Significant HLA Alleles and Amino Acid Changes

We identified 9 HLA alleles across 3 genes (6 for DQB1 [OMIM 604305], 2 for DPB1 [OMIM 142858], and one for DRB1 [OMIM 142857]) that had \( P \) values exceeding genome-wide significance (ie, <5 \times 10^{-8}); Seven alleles were positively associated with HL risk, and 2 alleles were negatively associated with it (Table 2). The HLA allele with the strongest association was HLA-DQB1_0603 \( (P = 9.35 \times 10^{-11}; \text{imputed } Z = 6.48) \). An additional set of 55 amino acid changes in HLA genes was significantly associated with HL risk (eTable 2 in the Supplement). Previously reported HLA alleles associated with HL susceptibility are provided in eTable 3 in the Supplement. All 6 HLA alleles in class II genes were significant, including HLA-DQB1_0602 and HLA-DRB1_1501, at the genome-wide significance level, whereas 2 HLA alleles in class I genes were not significant in our study. Regarding the 4 previously reported HL-associated amino acid changes, our analysis replicated the association for AA_DPB1_55_33156577_A \( (P = 3.6 \times 10^{-3}; \text{imputed } Z = 2.91) \) but not for the other 3 amino acid residues in HLA-DPB1 (ie, leucine 35, alanine 69, and valine 84).31

<table>
<thead>
<tr>
<th>SNV</th>
<th>Position</th>
<th>Allele</th>
<th>Unconditional logistic regression OR (95% CI)</th>
<th>Conditional logistic regression OR (95% CI)</th>
<th>Conditional logistic regression OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs28383311</td>
<td>32619234</td>
<td>A</td>
<td>1.80 (1.59-2.03)</td>
<td>2.14 \times 10^{-11}</td>
<td>NA</td>
</tr>
<tr>
<td>rs3129198</td>
<td>33103045</td>
<td>A</td>
<td>1.54 (1.39-1.72)</td>
<td>2.64 \times 10^{-15}</td>
<td>1.53 (1.37-1.70)</td>
</tr>
<tr>
<td>rs3129890</td>
<td>32446496</td>
<td>T</td>
<td>1.23 (1.10-1.37)</td>
<td>1.71 \times 10^{-4}</td>
<td>1.48 (1.32-1.65)</td>
</tr>
</tbody>
</table>

Abbreviations: NA, not applicable; OR, odds ratio; SNV, single-nucleotide variant.

\( ^a \) Conditional on rs28383311 and rs3129198.

\( ^b \) Conditional on rs28383311.

\( ^c \) Position according to human reference Genome Reference Consortium Human Build 38.
Figure 2. Regional Plot of Association Results and Recombination Rates for the Hodgkin Lymphoma Risk Loci

The $-\log_{10} P$-values (y-axis) are plotted against chromosomal position (x-axis). The black horizontal dotted line represents the genome-wide significance threshold $P = 5 \times 10^{-8}$. The most significant single-nucleotide variant (SNV) in each analysis is marked at the center of the 1-megabase (Mb) region nearby. The color intensity of each SNV represents the extent of linkage disequilibrium (LD) with the lead genetic variant, dark blue ($r^2 = 0$) to red ($r^2 = 1.0$). The light blue line at the bottom of each plot shows genetic recombination rates by using 1000 Genomes Project LD. The bottom tracks show nearby genes mapping to the region of association. Association hits, which have been reported in the EBI genome-wide association study (GWAS) catalog, are also annotated. Diamond-shaped markers indicate the lead SNV in each region, and circles indicate other SNVs in variable extent of LD with the lead SNV. Unconditional logistic regression results were plotted for A, and step-wise conditional logistic regression results were plotted for B and C.
SNVs in the Non-HLA Region

Our analysis identified the following genome-wide significant SNVs in non-HLA loci (Table 3). First, we identified rs1432297 (OR, 1.29; 95% CI, 1.18-1.41; \( P = 2.5 \times 10^{-8} \)) at 2p16.1, which is in moderate LD (\( r^2 = 0.52, D' = 0.73 \)) with previously reported rs1432295,\(^{18} \) mapping to the REL gene. REL encodes c-Rel, which belongs to the Rel family of genes that regulate inflammation and immune response. REL is normally expressed in germinal B-cells and immunoblasts. Moreover, serving as a proto-oncogene, REL also functions in B-lymphocyte survival and proliferation\(^{32} \) and has been identified as a candidate gene for HL risk.\(^{33} \) Second, we identified rs2757647 (OR, 1.30; 95% CI, 1.18-1.42; \( P = 3.5 \times 10^{-8} \)) at 6q23.3, which is in moderate LD (\( r^2 = 0.59, D' = 0.83 \)) with previously reported rs6928977,\(^{18} \) mapping to the AHI1 (OMIM 608894) gene. AHI1 is a prerequisite for development of the cerebellum and cortex in humans, and perturbations in AHI1 expression may result in specific subtypes of human leukemia.\(^{34} \) Third, we identified rs13279159 (OR, 1.33; 95% CI, 1.20-1.47; \( P = 1.7 \times 10^{-6} \)) at 8q24.21, with a high LD (\( r^2 = 0.91; D' = 1.00 \)) with previously reported rs2019960,\(^{19} \) located at a 82-kb LD block telomeric to PVT1 (OMIM 165140), which represents a long noncoding RNA associated with acute myeloid leukemia and HL. Fourth, we identified rs3824662 (OR, 1.52; 95% CI, 1.33-1.73; \( P = 3.9 \times 10^{-10} \)) at 10p14, with a high LD (\( r^2 = 0.91; D' = 1.00 \)) with previously reported rs3781093,\(^{18} \) mapping to an intronic region of GATA3 (OMIM 131320); GATA3 expression is aberrant in Reed-Sternberg cells and their variants. In addition, a novel uncommon SNV with a large effect size, rs117953624 (OR, 1.98; 95% CI, 1.56-2.51; \( P = 1.5 \times 10^{-8} \); MAF, 0.02), was identified and mapped to PDGFD (platelet-derived growth factor D; OMIM 609673) (efigure 4 in the Supplement). Collective evidence supports a role for PDGF in HL pathogenesis,\(^{35} \) and PDGF is a transforming, angiogenic growth factor.\(^{10} \) In addition, 12 of 18 previously reported genome-wide significant non-HLA SNVs (67%) were replicated with \( P < .05 \) (eTable 4 in the Supplement).

### Table 2. Association of HLA Alleles With Hodgkin Lymphoma Susceptibility

<table>
<thead>
<tr>
<th>Marker ID</th>
<th>Position ( ^a )</th>
<th>Allele</th>
<th>Imputed ( Z ) ( ^b )</th>
<th>( r^2 ) ( ^c )</th>
<th>Imputed ( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLA_DQB1_0603</td>
<td>32631061</td>
<td>Presence</td>
<td>Absence</td>
<td>6.48</td>
<td>0.82</td>
</tr>
<tr>
<td>HLA_DQB1_0602</td>
<td>32631061</td>
<td>Presence</td>
<td>Absence</td>
<td>6.08</td>
<td>0.88</td>
</tr>
<tr>
<td>HLA_DQB1_06</td>
<td>32631061</td>
<td>Presence</td>
<td>Absence</td>
<td>5.71</td>
<td>0.92</td>
</tr>
<tr>
<td>HLA_DQB1_0501</td>
<td>32631061</td>
<td>Presence</td>
<td>Absence</td>
<td>-5.54</td>
<td>0.97</td>
</tr>
<tr>
<td>HLA_DQB1_05</td>
<td>32631061</td>
<td>Presence</td>
<td>Absence</td>
<td>-5.87</td>
<td>0.92</td>
</tr>
<tr>
<td>HLA_DQB1_0301</td>
<td>32631061</td>
<td>Presence</td>
<td>Absence</td>
<td>5.81</td>
<td>0.88</td>
</tr>
<tr>
<td>HLA_DRB1_1501</td>
<td>32552064</td>
<td>Presence</td>
<td>Absence</td>
<td>5.84</td>
<td>0.89</td>
</tr>
<tr>
<td>HLA_DPB1_10</td>
<td>33049368</td>
<td>Presence</td>
<td>Absence</td>
<td>5.49</td>
<td>0.51</td>
</tr>
<tr>
<td>HLA_DPB1_1001</td>
<td>33049368</td>
<td>Presence</td>
<td>Absence</td>
<td>5.49</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Abbreviation: HLA, human leukocyte antigen.

\( ^a \) Postion refers to the genomic coordinates based on Genome Reference Consortium Human Build 37.

\( ^b \) \( Z \) scores represent association with Hodgkin lymphoma.

\( ^c \) The \( r^2 \) \( ^c \) values represent the assessment of the imputation reliability at each HLA variant.

### Table 3. Association of SNVs in Non–Human Leukocyte Antigen Regions With Hodgkin Lymphoma Susceptibility Based on Meta–Genome-Wide Association Studies

<table>
<thead>
<tr>
<th>SNV</th>
<th>Position ( ^a )</th>
<th>Neighboring genes</th>
<th>Allele</th>
<th>Risk Other</th>
<th>OR (95% CI)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1432297</td>
<td>chr:2:60843517</td>
<td>REL</td>
<td>G</td>
<td>A</td>
<td>1.29 (1.18-1.41)</td>
<td>( 2.50 \times 10^{-06} )</td>
</tr>
<tr>
<td>rs2757647</td>
<td>chr:6:135446416</td>
<td>AHI1</td>
<td>T</td>
<td>C</td>
<td>1.30 (1.18-1.42)</td>
<td>( 3.52 \times 10^{-06} )</td>
</tr>
<tr>
<td>rs13279159</td>
<td>chr:8:128150801</td>
<td>PVT1</td>
<td>G</td>
<td>A</td>
<td>1.33 (1.20-1.47)</td>
<td>( 1.74 \times 10^{-06} )</td>
</tr>
<tr>
<td>rs3824662</td>
<td>chr:10:8062245</td>
<td>GATA3</td>
<td>C</td>
<td>A</td>
<td>1.52 (1.33-1.73)</td>
<td>( 3.86 \times 10^{-10} )</td>
</tr>
<tr>
<td>rs117953624</td>
<td>chr:11:104438259</td>
<td>PDGFD</td>
<td>C</td>
<td>T</td>
<td>1.98 (1.56-2.51)</td>
<td>( 1.45 \times 10^{-06} )</td>
</tr>
</tbody>
</table>

Abbreviations: OR, odds ratio; SNV, single-nucleotide variant.

\( ^a \) Position according to human reference Genome Reference Consortium Human Build 38.
Discussion

HL epidemiology is complex, with cases developing across an age spectrum, including children, adolescents, young adults, and older adults. Survival of HL is more favorable among children, adolescents, and young adults compared with older adults. Previous studies have suggested that a difference in HL etiology between children and adolescents vs older adults could explain this difference in survival. To investigate the genetic etiology of HL arising during childhood, adolescence, or young adulthood, we conducted a 3-way meta-GWAS analysis by leveraging preexisting data from the SJLIFE and CCSS.

Using step-wise conditional logistic regression methods, we identified 3 independent SNVs (rs28383311, rs3129198, rs3129890) in the HLA locus; most previous studies reported only 1 SNV. The primary and secondary SNVs were in moderate or high LD with previously reported SNVs, but the tertiary SNV had a low correlation with all established HLA SNVs, suggesting both the commonality and difference for genetic factors at the HLA locus underpinning HL risk during childhood and adolescence compared with other ages at onset.

We also identified 5 SNVs at non-HLA loci with genome-wide significance. Four of these SNVs were in moderate or high LD with previously reported SNVs, further validating these findings mapped to genes, such as REL, AHI1, PVT1, and GATA3, in studies comprising mostly adult-onset HL. Collectively, these findings implicate multiple biological processes, including germinal center dysfunction, disrupted T-cell function, and nuclear factor-κB (NF-κB) activation, in HL pathogenesis across age groups. Our identification of rs117361561 (maps to PDGFD), a novel uncommon SNV with a large effect size, is intriguing because PDGFD has a fundamental role in cell growth, differentiation, survival regulation, invasion, and angiogenesis, supporting its potential association with cancer progression. Considering that NF-κB may moderate the expression of PDGF-inducible genes, the finding of a novel association between rs117361561 and HL risk provides additional evidence to support the role of NF-κB activation in HL pathogenesis, particularly in HL developing in younger people. Future mechanistic studies are needed to demonstrate how these genes function in HL carcinogenesis.

Limitations

This study has limitations. First, we used case-case comparisons in 2 GWAS analyses and a case-control comparison in 1 GWAS analysis. Using other cases as a control set is not ideal and may result in either false-negative or false-positive results. However, we evaluated the heterogeneity across 3 GWASs for reporting genome-wide significant findings to minimize false-positive results. Second, the HL cases included in this study all had 5 or more years of survival; patients who died before 5-year survival were not included. Third, we did not have detailed histologic subtype information, so stratified analysis based on histologic subtype was not carried out. Fourth, EBV infection status was unavailable for all patients with HL in CCSS. Among 380 patients with HL in SJLIFE, 65 (17%) had EBV status, with 15 (23%) EBV positive and 50 (77%) EBV negative. Previous studies have identified low prevalence of EBV-positive HL in adolescent and young adults compared with children younger than 10 years or adults older than 80 years, as well as the different etiology between patients with EBV-negative and EBV-positive HL. This appears to be consistent with our current results, which showed that none of the 3 HLA SNVs we identified were in the HLA I class implicated in EBV-positive HL. Fifth, the study was conducted among individuals of European ancestry, so the findings may not be generalizable to other racial and ethnic groups.

Conclusions

In summary, we refined HL associations in the HLA locus by identifying 3 independent SNVs. In non-HLA regions, we found 1 novel SNV and replicated the majority of previous findings, including 4 SNVs exceeding genome-wide significance. These findings suggest mostly common, as well as...
potentially unique, genetic factors underpinning HL risk across age groups. Future replication and subsequent functional investigations are warranted to elucidate the genetic architecture and biological mechanisms of HL susceptibility.


SUPPLEMENT.

eFigure 1. Study Population Flow Diagram

eFigure 2. Quantile-Quantile Plot for the P Values in the -log10 Scale for the Discovery GWAS Meta-analysis

eFigure 3. Pair-wise Linkage Disequilibrium Among 3 Independent SNPs in HLA Region

eTable 1. Meta-GWAS Results for Previously Published HL-Associated SNPs in HLA Region

eTable 2. Significant Amino Acid Changes Associated With HL in HLA Genes
eTable 3. Previously Reported HLA Alleles Associated With HL Susceptibility

eTable 4. Meta-GWAS Results for Previously Published HL-Associated SNPs in Non-HLA Region

eFigure 4. SNV rs117953624 near PDGFD