A novel risk locus at 6p21.3 for Epstein-Barr virus positive Hodgkin lymphoma

Manon Delahaye-Sourdeix¹, Kevin Y. Urayama², Valérie Gaborieau¹, Rianne Veenstra³, Matthieu Foll¹, Amelie Chabrier¹, Yolanda Benavente⁴,⁵, Alexandra Nieters⁶, Nikolaus Becker⁷, Lenka Foretova⁸, Marc Maynadié⁹, Anthony Staines¹⁰, Karin Ekstrom Smedby¹¹, Ingrid Glimelius¹²,¹³, Tracy Lightfoot¹⁴, Pierluigi Cocco¹⁵, Pilar Galan¹⁶, Lars J. Vatten¹⁷, Eric J. Duell¹⁶, Lambertus Kiemeney¹⁸, Eve Roman¹⁴, Silvia de Sanjosé⁴,⁵, Mark Lathrop¹⁹, Mads Melbye²⁰, Paul Brennan¹, Arjan Diepstra³, Anke van den Berg³, Henrik Hjalgrim²⁰, Ruth F. Jarrett²¹, James D. McKay¹∗

*Corresponding author full information
James McKay
International Agency for Research on Cancer (IARC/WHO)
Genetic Cancer Susceptibility group (GCS)
150 cours Albert Thomas
69372 Lyon CEDEX 08
France
Phone: +33 (0) 4 72 73 8093
Fax: +33 (0) 4 72 73 8388
gcs@iarc.fr

Author affiliations
¹ International Agency for Research on Cancer (IARC), Lyon, France
² Department of Human Genetics and Disease Diversity, Tokyo Medical and Dental University, Tokyo, Japan
³ Department of Pathology and Medical Biology, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands
⁴ Unit of Infections and Cancer. Cancer Epidemiology Research Programme. IDIBELL. Catalan Institute of Oncology. Barcelona, Spain
⁵ CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
⁶ Centre of Chronic Immunodeficiency Freiburg, University Medical Centre Freiburg, Freiburg, Germany
⁷ Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany
⁸ Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic
⁹ Registry of Hematological Malignancies, EA4184, University of Burgundy, University Hospital of Dijon, Dijon, France
¹⁰ School of Nursing, Dublin City University, Glasnevin, Dublin 9, Ireland
¹¹ Department of Medicine Solna, Clinical Epidemiology Unit, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden
Running title: Hodgkin Lymphoma, HLA and EBV status

Word count:
Abstract: 219
Main text: 1855
Nb of tables: 0
Nb of figures: 2
Nb of supplementary tables: 6
Nb of supplementary figures: 1

Funding

This work was supported by Institut National du Cancer, France (P. Galan); Spanish Ministry of Health grant CIBERESP [PI11/0181008-1555, 06/02/0073] and Agència de Gestió d’Ajuts Universitaris i de Recerca–Generalitat de Catalunya [Catalonian Government 2014SGR756] (Y. Benavente, E.J. Duell, S. de Sanjosé); Ministry of Health of the Czech Republic [MZ0 MOU2005] (L. Foretova); German José Carreras Leukemia Foundation [DJCLS_R04/08] (N. Becker); Federal Office for Radiation Protection [StSch4261, StSch4420] (A. Nieters); EC 5th Framework Program Quality of Life [QLK4-CT-2000-00422] (Y. Benavente, A. Nieters, S. de Sanjosé); EC 6th
Disclosure of Conflicts of Interest

The authors declare no conflict of interest.
Abstract

Background: A proportion of the genetic variants involved in susceptibility to Hodgkin lymphoma (HL) differ by the tumor's Epstein–Barr virus (EBV) status, particularly within the major histocompatibility complex (MHC) region.

Materials and Methods: We have conducted a single nucleotide polymorphism (SNP) imputation study of the MHC region, considering tumor EBV status in 1,200 classical HL (cHL) cases and 5,726 control subjects of European origin. Notable findings were genotyped in an independent study population of 468 cHL cases and 551 controls.

Results: We identified and subsequently replicated a novel association between a common genetic variant rs6457715 and cHL. While strongly associated with EBV-positive cHL (OR=2.33 [1.83 – 2.97] P=7x10-12)), there was little evidence for association between rs6457715 and the EBV negative subgroup of HL (OR=1.06 [0.92 – 1.21]), indicating that this association was specific to the EBV positive subgroup (Phet< P=10-8). Furthermore, the association was limited to EBV-positive cHL subgroups within mixed cell (MCHL) and nodular sclerosis subtypes (NSHL), suggesting that the association is independent of histological subtype of cHL.

Conclusions: rs6457715, located near the HLA-DPB1 gene, is associated with EBV positive cHL and suggests this region as a novel susceptibility locus for cHL.

Impact: This expands the number of genetic variants that are associated with cHL, and provides additional evidence for a critical and specific role of EBV in the etiology of this disease.
Introduction

Hodgkin lymphoma (HL) is a cancer of the lymphatic system characterized by the presence of B-cell derived Hodgkin Reed-Sternberg (HRS) tumor cells (1). HL is relatively rare, but contributes substantially to worldwide disease burden, totalling 66,000 new cases in 2012 and 25,000 deaths (2). It affects mainly young adults aged 15 to 35 years and older adults aged 55 years and over. Classical HL (cHL) is the major form and comprises four histological subtypes of which the nodular sclerosis HL (NSHL) subtype is most common followed by mixed cellularity HL (MCHL).

In approximately one-third of cHL cases in industrialised countries, the HRS cells are a clonally expanded population of Epstein–Barr virus (EBV)-infected cells. Young adult and NSHL cases are less likely to be EBV-positive than older adult and MCHL cases (3); EBV-positive cases are more likely to be male in contrast to more equal gender representation among EBV-negative cases. EBV-positive cases are associated with recent EBV infection, as evidenced by an association with infectious mononucleosis, and also with weakened immunity, e.g. HIV infection or iatrogenic immunosuppression after organ transplantation (4-6). The human leukocyte antigen (HLA) genes, located within the major histocompatibility complex (MHC) region, have been implicated in the etiology of cHL (7-11). Some associations show heterogeneity by tumor EBV status (9, 11-14) reinforcing the importance of acknowledging EBV status in examining the etiology of cHL.

Here, we expanded our investigation of the MHC region based on data from a previous cHL genome-wide association study (GWAS) and performed EBV-status specific analyses using imputation procedures for both SNPs and HLA alleles.
Materials and Methods

Imputation of 791,716 SNPs in human chromosome 6 was conducted using MACH v1.0 (15) and Minimac (version 2010.12.13) (16). The August 2010 release of the 1000 Genomes Project European (CEU) data set was used as the reference panel to impute genotypes for 1,200 cHL cases (of which 265 were EBV positive) from the EPILYMPH study, the Scotland and Newcastle Lymphoma Group (SNLG) studies, the Young Adult Hodgkin’s Disease Case-Control Study (YHHCCS), the Scandinavian Lymphoma Etiology Study, and the Northern Dutch Hodgkin Lymphoma study, and 5,726 controls (a large subset that were part of a GWAS of cHL reported previously (11)) (Supplementary Table S1). IRB: the study was approved by the IARC Ethics Committee (project no. 08-21). Poorly imputed SNPs defined by an R^2 < 0.80 and a quality < 0.90 were excluded from the analyses. As a technical validation, we compared imputation dosages and direct genotypes in a subset of 562 individuals. For this, we classified subjects based on imputation dosages as homozygous wild-type (less than 0.3), heterozygous (between 0.7 and 1.3), or homozygous variant (greater than 1.7).

HL EBV status was ascertained through in situ hybridization for EBERs and/or through immunohistochemical staining for EBV LMP-1 protein on formalin-fixed paraffin-embedded tumor samples as described previously (11).

Test of association between imputed SNPs and cHL as well as by subtypes (i.e. NSHL, MCHL, EBV-positive cHL, and EBV-negative cHL) was performed using a probabilistic dosage model in ProbABEL v0.4.3 (17), adjusted for sex and the first informative eight eigenvectors from a principal components analysis (EIGENSOFT 3.0, Broad Institute, Cambridge, MA) derived using a subseries of 11,029 SNPs across the genome to control for potential population stratification (18).
Additionally, we adjusted for the effect of rs2734986 (HLA-A), rs6904029 (HCG9), rs2248462 (MICB), rs2395185 (HLA-DRA) and rs6903608 (HLA-DRA), five MHC region SNPs previously associated with cHL risk.

For the replication series, 468 cHL cases and 551 study-specific controls from EPILYMPH, SNLG, YHHCCS, and the Epidemiology and Genetics Lymphoma Case-Control Study had complete genotype data for the 6 MHC loci of interest (5 known loci and rs6457715) (Supplementary Table S1 and S2). 106 of the EBV-typed cases were classified as EBV-positive cHL. Genotyping of rs6457715 was performed using Taqman Pre-Designed SNP Genotyping Assays (Applied Biosystems, Carlsbad, CA). The genotype distribution was in accordance with that expected by Hardy-Weinberg equilibrium and the assay had duplicate genotyping concordance rate of >99%.

From the GWAS genotyped data, classical HLA loci A, C, B, DRB1, DRB3-5, DQA1, DQB1, and DPB1 were imputed using HLA*IMP:02 (19). We confirmed the robustness of the imputation process by comparing HLA loci imputation (except DRB3-5) with directly genotyped HLA data (locus-specific PCR followed by sequence specific oligonucleotide hybridization (20)) in 334 UK and 284 Dutch individuals. The average concordance rate between HLA imputed and directly genotyped data was of 93.8% (ranging from 82.8% to 99.3%) (Supplementary Table S3).

LCLs were generated from blood samples from 95 healthy controls and 70 post-therapy cHL patients (from blood samples collected at least 1 year after completion of all therapies) by infection with EBV. Genotyping of the LCLs was carried out using a TaqMan SNP assay. Association between HLA-DPB1 gene expression levels and genotype was assessed by linear regression using R.
Results

We performed SNP imputation of 1,200 classical HL (cHL) cases and 5,726 control subjects of European origin to undertake a comprehensive evaluation of the MHC region in total cHL and by EBV tumor status of cHL, while controlling for the effects of previously described susceptibility variants (11). We did not identify any novel signal ($P < 10^{-7}$) in total cHL or the EBV-negative subgroup, in addition to previously described associations in MHC class I and class II regions (Supplementary Figure 1a and 1b). However, three imputed genetic variants (rs6457715, rs6457714 and rs6457711) were associated with the EBV-positive subgroup at genome wide significance levels ($P < 10^{-7}$) (Figure 1). The rs6457715, rs6457714 and rs6457711 variants showed evidence of linkage disequilibrium (LD) and conditioning on rs6457715 (A/G) was consistent with a single association signal (risk allele frequencies: EBV-positive cHL=0.87; controls=0.79). We directly genotyped rs6457715 within a subset (N=562, methods) of the sample in which we undertook imputation, and the concordance rate for rs6457715 between categorized dosages and direct genotypes was 99.47%, confirming the high accuracy of the imputation process. Within this discovery set, the rs6457715 major allele (A) was strongly associated with an increased risk of EBV-positive cHL (OR=2.39 [1.80 – 3.18], $p=1\times10^{-9}$), with no evidence for association in EBV-negative (p=0.791) or overall cHL (p=0.622) (Supplementary Table S2).

We subsequently directly genotyped rs6457715 in an independent series of 468 cHL cases and 551 controls of European origin and observed similar patterns of association as in the discovery series, including a statistically significant increased risk of EBV-positive cHL (OR=2.17 [1.36 – 3.49], $p = 0.0013$), but not EBV-negative (p=0.791) or overall cHL (p=0.622) (Supplementary Table S2). The association was more pronounced in the model adjusted for known MHC cHL susceptibility variants in both the validation and replication cohorts (Supplementary Table S2).
Combining the imputation-based GWAS and replication results (Figure 2) demonstrated a 2.3-fold increased risk of EBV-positive cHL associated with rs6457715 (allele A relative to allele G) \((p=7.53\times 10^{-12})\) but no effect within the EBV-negative subgroup \((p=0.405)\) and marked evidence for heterogeneity between the EBV-positive and EBV-negative groups \((p_{\text{het}}=3\times 10^{-8})\). While the predominant cHL histological subtypes tend to be correlated with the tumor EBV status (NSHL and MCHL more likely to be EBV negative and positive, respectively \((3)\)), the heterogeneity of this association remained when considering the tumor EBV status within the MCHL and NSHL histological subgroups, implying that EBV status is the main source of heterogeneity (Figure 2).

We estimated HLA allele genotypes using imputation techniques. Multivariate regression analysis of rs6457715 adjusting for imputed alleles at each HLA locus \((i.e. A, C, B, DRB1, DRB3, DRB4, DRB5, DQA1, DQB1, \text{ and } DPB1)\) resulted in little attenuation of effect for rs6457715, suggesting that the relationship is independent of these HLA alleles (Supplementary Table S4).

rs6457715 has been suggested to be an eQTL for \(HLA-DPB1\) \((21)\), however, we were unable to detect a clear association \((p=0.15)\) between \(HLA-DPB1\) gene expression levels and rs6457715 within 165 individuals (70 cHL patients and 95 controls) where cells have been cultured and \(HLA-DPB1\) gene expression levels subsequently assayed by Taqman real time PCR.

Modelling a genetic risk score of the three independent MHC alleles associated with EBV-positive cHL \((rs2734986, rs6904029 (11))\) and rs6457715) suggested that there was a per-allele increased odds of 1.86 for EBV positive cHL, with the highest quartile of allele carriers having a \(~7\) fold increased odds relative to the lowest quartile (Supplementary Table S5).
Discussion

We have identified a novel group of highly correlated genetic variants, located near the MHC class-II *HLA-DPB1* gene, which are associated with genetic susceptibility to cHL. Statistically robust observations were made in both discovery and validation cohorts, and cross validation of the genotyping methods confirmed their technical fidelity.

The association was restricted to the subgroup of cHL in which EBV was detected within the tumor cells by EBV EBER *in-situ* hybridization and/or LMP-1 immunohistochemistry. This implies that these alleles are only relevant to the subset of cHL that are “EBV positive cHL”, i.e. where the HRS cells appear originate from a clonally expanded population of EBV-infected cell. Most MCHL cases are EBV positive, however as the heterogeneity by tumor EBV status was also observed following stratification by NSHL and MCHL subtypes of cHL. This strong heterogeneity implies a biological interaction between EBV and the genetic alleles leading to susceptibility to Hodgkin lymphoma irrespective of the histological subtype.

The association with the genetic variants was statistically independent of the previously described cHL susceptibility alleles found within the MHC region. A proportion of these other susceptibility alleles also demonstrate heterogeneity in their effects when considering the EBV status of the tumor (9, 11, 12). The class I genetic variants rs2734986 and rs6904029, which correlated with HLA-A gene alleles A*02 and A*01 respectively, are associated with EBV-positive cHL, whereas a genetic variant located in the class II region, rs6903608 has been associated uniquely with EBV-negative cHL (9, 11, 12). As noted above for rs6457715, heterogeneity in EBV stratified cHL risk was observed even following stratification for the NSHL and MCHL subtypes, for the rs2734986 and rs6904029 and EBV positive cHL and rs6903608 and EBV negative cHL. All of these observations taken together imply a remarkable relationship between tumor EBV status and the risk conferred by genetic variants within the MHC region and reinforces the importance of acknowledging EBV status when examining the etiology of HL.
We were unable to establish a link between rs6457715 and a particular HLA allele (Supplementary Table S4), albeit within the limitation inferring HLA genotypes by imputation. rs6457715 resides within an intronic region of the pseudogene *HLA-DPB2* and is suggested to be an *HLA-DPB1* expression quantitative trait locus (eQTL) ([21]). However, multiple additional independent genetic variants at this locus, not associated with EBV-positive cHL, are also suggested to be *HLA-DPB1* eQTLs (Supplementary Table S6), implying that differences in *HLA-DPB1* expression levels are unlikely to clearly explain the association. Furthermore, we were unable to replicate this eQTL association with rs6457715 and *HLA-DPB1*. As with the rs6903608 effect on EBV-negative cHL ([9]), the functional mechanism through which the association with rs6457715 is mediated remains ambiguous.

rs6457715 represents the third independent genetic loci within the MHC region (represented by rs2734986, rs6904029 and rs6457715) exclusively associated with EBV-positive cHL. Each of these alleles shows a strong association with cHL, with the risk allele conferring about a two-fold increased risk per allele of developing the disease. This is an unusually pronounced effect compared to the genetic risks usually conferred by common genetic variants. When considered as a genetic risk score, in a multivariate model, there was a ~7 fold difference between the top and bottom 25% of allele carriers. Such a model makes a number of important assumptions, for example the absence of LD which is a particularly complex in this region of the genome. Nevertheless, the magnitude of this association implies some potential utility for risk prediction of EBV-positive cHL, particularly in context of other described HL EBV positive risk factors, such as Infectious Mononucleosis.
Acknowledgements

We thank G. Thomas and Synergy Lyon Cancer (Lyon France) for high performance computing support.
References


Figure Legends

Figure 1
Association between genetic variants and Epstein–Barr virus (EBV)–positive HL within an approximately 6.5 Mb region of the extended major histocompatibility complex located at 6p21.
Overlaid plots of the −log₁₀ (P values) for 25,961 single-nucleotide polymorphisms (SNPs) by their chromosomal position. Results of the EBV-positive cHL analysis unadjusted for previously associated SNPs in the region are plotted in grey. Plotted in black are the results for EBV-positive cHL adjusted by other associated genetic loci including rs2734986, rs6904029, rs2248462, rs2395185 and rs6903608. The horizontal dotted line represents the genome-wide significance threshold level (P < 1 × 10⁻⁷). The arrow highlights the new EBV-positive associated variant rs6457715. There are three SNPs (rs6457715, rs6457714, and rs6457711) exceeding the p-value threshold at the 33Mb position, although results of two (rs6457715 and rs6457714) appear overlapping.

Figure 2
Summary of results for rs6457715 and risk of Hodgkin lymphoma by cHL subgroups.
Results of analyses combining the GWAS and replication stages are shown for total cHL and by histology, tumor cell EBV status, age of onset, and study. EBV-positive cHL results are also shown stratified by study. Odds ratios (ORs) and 95% confidence intervals (CIs) were derived using multiple logistic regression assuming a log-additive genetic model and adjusting for the 5 known MHC loci (i.e. rs2734986, rs2248462, rs2395185, rs6903608 and rs3823355), sex and eight principal components analysis eigenvectors (or country in the replication analysis). The risk allele is the major allele (A) of rs6457715. Combined GWAS and replication results were generated using inverse variance weighting meta-analysis.
a Study-specific results for EPILYMPH and UK Studies included GWAS and replication data, while results for SCALE and the Netherlands studies are those from the GWAS.
Chr 6: rs6457715  Cases  Controls  OR  95%CI

Total cHL             1668  6277  1.22  1.10-1.35

By histology (p_het=0.955)
NSHL                  1195  6277  1.20  1.06-1.36
MCHL                  309   6277  1.38  1.10-1.73

By EBV status (p_het<0.001)
EBV- cHL               910  6203  1.06  0.92-1.22
EBV+ cHL               371  6203  2.33  1.83-2.97

By histology & EBV (p_het<0.001)
MCHL/EBV+             119   6203  2.95  1.89-4.60
MCHL/EBV-              94   6203  1.06  0.73-1.54
NSHL/EBV+             214   6203  1.99  1.48-2.68
NSHL/EBV-             741   6203  1.08  0.93-1.25

By age of diagnosis (p_het=0.700)
15-35 years             921  6277  1.20  1.04-1.38
36-85 years             746  6277  1.25  1.07-1.46

EBV+ cHL by study (p_het=0.130)
EPILYMPH              13   147  1.35  0.42-4.34
UK Studies            207  3236  1.98  1.44-2.72
SCALE                  83   189  1.70  0.98-2.95
Netherlands            68  1804  4.33  2.20-8.52