

Multiple Hodgkin lymphoma–associated loci within the HLA region at chromosome 6p21.3

*Loukas Moutsianas,¹ *Victor Enciso-Mora,² Yussanne P. Ma,¹ Stephen Leslie,¹ Alexander Dilthey,¹ Peter Broderick,² Amy Sherborne,² Rosie Cooke,³ Alan Ashworth,⁴ Anthony J. Swerdlow,³ Gilean McVean,^{1,5} and Richard S. Houlston²

¹Department of Statistics, University of Oxford, Oxford, United Kingdom; ²Section of Cancer Genetics, Institute of Cancer Research, Sutton, United Kingdom;

³Section of Epidemiology, Institute of Cancer Research, Sutton, United Kingdom; ⁴Breakthrough Breast Cancer Research Centre, London, United Kingdom; and

⁵Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom

Since an association between the human leukocyte antigen (HLA) region and Hodgkin lymphoma (HL) was first reported in 1967, many studies have reported associations between HL risk and both single nucleotide polymorphism (SNP) and classic HLA allele variation in the major histocompatibility complex. However, population stratification and the extent and complexity of linkage disequilibrium within the major histocompatibility com-

plex have hindered efforts to fine-map causal signals. Using SNP data to impute alleles at classic HLA loci, we have conducted an integrated analysis of HL risk within the HLA region in 582 early-onset HL cases and 4736 controls. We confirm that the strongest signal of association comes from an SNP located in the class II region, rs6903608 (odds ratio [OR] = 1.79, $P = 6.63 \times 10^{-19}$), which is unlikely to be driven by association to *HLA-DRB*, *DQA*,

or *DQB* alleles. In addition, we identify independent signals at rs2281389 (OR = 1.73, $P = 6.31 \times 10^{-13}$), a SNP that maps closely to *HLA-DPB1*, and the class II HLA allele DQA1*02:01 (OR = 0.56, $P = 1.51 \times 10^{-7}$). These data suggest that multiple independent loci within the HLA class II region contribute to the risk of developing early-onset HL. (*Blood*. 2011; 118(3):670-674)

Introduction

Hodgkin lymphoma (HL) is a common lymph node cancer of germinal center B-cell origin, which is characterized by malignant Hodgkin and Reed-Sternberg (HRS) cells mixed with a dominant background population of reactive lymphocytes and other inflammatory cells.¹ Although Epstein-Barr virus (EBV) infection may be causally related to a number of cases, there is little evidence to support the involvement of other environmental risk factors.²

Evidence for inherited genetic influence on susceptibility is provided by the increased familial risk and high concordance between monozygotic twins.³ Since an association between the human leukocyte antigen (HLA) region and HL risk was first reported in 1967,⁴ studies have subsequently identified associations between both HLA class I and class II alleles and common HL (cHL) risk.⁵⁻¹⁰ Studies to date have evaluated only specific HLA alleles and have not taken into account the existence of complex linkage disequilibrium patterns between the multiple risk loci mapping to the major histocompatibility complex (MHC) region, associations between HLA alleles at different resolution, and the need to control rigorously for population stratification. In view of the limitations of these previously published studies, we have conducted a more comprehensive analysis.

It has recently been shown that single nucleotide polymorphism (SNP) data within the 6p21 region can be used to impute alleles at key classic class I (*HLA-A*, *HLA-B*, and *HLA-C*) and class II (*HLA-DRB1*, *HLA-DQA1*, and *HLA-DQB1*) loci with accuracy that exceeds 90% at the 4-digit level.^{11,12} Using existing genotype data from a previous genome-wide association study of HL, which identified many SNPs within the HLA region that are strongly

associated with disease,¹³ we can therefore assess the relative contribution of classic HLA alleles and SNP variation to HL risk. We identify multiple independent signals across the HLA, the strongest of which is most likely not driven by any of the HLA loci whose types were statistically inferred in this study.

Methods

Samples and genotyping

We analyzed the constitutional DNA of 615 patients diagnosed with HL (International Classification of Diseases 10 codes C81.0-3) who were ascertained through the Royal Marsden Hospitals National Health Service Trust Family History study, between 2004 and 2008 (mean age at diagnosis 38 years, SD 16 years) and an ongoing national study of cHL in females (mean age at diagnosis 23 years, SD 6 years) being conducted by the Institute of Cancer Research to assess late effects of HL treatment. Both routes of patient ascertainment thus favored the acquisition of early-onset HL cases. All cases were British residents and were self-reported to be of European ancestry.

DNA was extracted from samples using conventional methodologies and quantified using PicoGreen (Invitrogen). Genotyping of cHL cases was conducted using Illumina Infinium HD Human610-Quad BeadChips according to the manufacturer's protocols (Illumina). DNA samples with GenCall scores < 0.25 at any locus were considered "no calls." A SNP was deemed to have failed if < 95% of DNA samples generated a genotype at the locus. Cluster plots were manually inspected for all SNPs identified as potentially associated. Publicly available data from the Wellcome Trust Case-Control Consortium 2 study were used as controls. Specifically, this included 2420

Submitted March 1, 2011; accepted April 28, 2011. Prepublished online as *Blood* First Edition paper, May 19, 2011; DOI 10.1182/blood-2011-03-339630.

*L.M. and V.E.-M. contributed equally to this study.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2011 by The American Society of Hematology

persons from the 1958 Birth Cohort (58BC, also known as the National Child Development Study)¹⁴ and 2737 from the National Blood Service Collection (NBS), all of which had been genotyped on Illumina 1.2M arrays at the Wellcome Trust Sanger Institute. To ensure quality of genotyping in all assays, at least 2 negative controls and 1% to 2% duplicates (showing a concordance > 99.99%) were genotyped.

Collection of blood samples and clinicopathologic information from subjects was undertaken with informed consent and relevant ethical review board approval from the University of Oxford and the Institute of Cancer Research in accordance with the tenets of the Declaration of Helsinki.

Quality control

For any samples showing a high degree of relatedness, we removed the sample with the lower call rate from the analysis, following previously described protocols.¹³ This obviates the need to adjust for familial correlations and ensuring that unbiased estimates of HL risk associated with genotypes are derived. We excluded SNPs on the basis of deviation from Hardy-Weinberg equilibrium using a threshold of $P < 1.0 \times 10^{-5}$ in controls. We also removed SNPs with minor allele frequency < 0.05 and call rate < 0.95.¹³ To identify and exclude persons with non-Western European ancestry, case and control data were merged with data on persons of different ethnicities from the International HapMap Project, and dissimilarity measures were used to perform principal component analysis (supplemental Figure 1, available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article). For the control samples, we also followed the guidelines from the Wellcome Trust Case-Control Consortium and removed some additional controls.¹³ After imposing these stringent quality control measures, SNP genotypes were available on 582 cases and 4736 controls (2235 58C and 2501 NBS).

Statistical and bioinformatics analysis

For single SNP analyses, we considered the MHC to be defined by a 4.5-Mb region bordered by the *RFP* and *MLN* genes (rs209130 at 28 975 779 bp and rs1547668 at 33 883 424 bp, respectively) at the telomeric and centromeric ends of 6p21, respectively. For the HLA imputation, we made use of genome-wide association study SNP data for an extended region in 6p21 bounded by rs1165196 and rs2772372 (25 921 129 and 33 535 328 bp, respectively). All positions are with respect to build 36 of the Human Genome.

Analyses were primarily undertaken using R, Version 2 software. Initially, we analyzed the association between each SNP and the risk of HL using the Cochran-Armitage trend test. Odds ratios and associated 95% confidence intervals were calculated by unconditional logistic regression. We then conducted stepwise logistic regression analyses to interrogate the region for independent secondary signals.

From a reference database of SNP haplotypes carrying known HLA alleles, we imputed HLA genotypes for the cHL cases, the NBS controls, and the subset of the 58BC samples, for which classic HLA data were not available, as previously described.^{11,12} The reference database combines classic HLA data from the HapMap Project and the 58BC. Details of quality control have been previously described.¹² Imputation was performed for 3 class I and 3 class II loci: HLA-A ($n = 2474$ reference samples), HLA-B ($n = 3090$), HLA-C ($n = 2022$), HLA-DQA1 ($n = 175$), HLA-DQB1 ($n = 2629$), and HLA-DRB1 ($n = 2665$). Imputation accuracy was assessed through a cross-validation analysis of the training data. Training was performed on two-thirds of the reference panel and tested on the remaining third at the SNPs chosen for imputation in the current study. Thresholding calls at a posterior probability of 0.7 provided call rates of between 0.90 (HLA-DRB1) and 0.99 (HLA-C) and accuracy of > 0.95 for all loci at the 2-digit level and between 0.91 (HLA-DQA1) and 0.98 (HLA-DQB1) at the 4-digit level.

The familial relative risk of HL attributable to any locus is given by the formula¹⁵:

$$\lambda^* = \frac{p(pr_2 + qr_1)^2 + q(pr_1 + q)^2}{(p^2r_2 + 2pqr_1 + q^2)^2}$$

where p is the population frequency of the minor allele, $q = 1 - p$, and r_1 and r_2 are the relative risks (approximated by the odds ratios) for heterozygotes and the rarer homozygotes, relative to the more common homozygotes. From λ^* , it is possible to quantify the impact the locus makes to the overall familial risk of HL seen in first-degree relatives. Assuming a multiplicative interaction between risk alleles, the proportion of the overall familial risk attributable to the locus is given by $\log(\lambda^*)/\log(\lambda_0)$, where λ_0 , the overall familial risk, is assumed on the basis of epidemiologic studies to be 3.93.¹⁶

To examine whether any directly typed or imputed SNPs annotate potential regulatory sequences within the 13.8-kb region of linkage disequilibrium (LD) encompassing the SNP rs6903608 (which had the highest signal of association), we performed *in silico* searches using Transfac Matrix Database Version 7.2,¹⁷ PRMod Version 1.0,¹⁸ and EEL¹⁹ software.

Results

We examined samples from 615 cases and 5157 controls and subjected these to rigorous quality control as described in “Statistical and bioinformatics analysis.” Furthermore, we evaluated the datasets for ancestral differences by principal component analysis. After quality control, data were available on 582 cases and 4736 controls. Supplemental Figure 1 shows that the cases and controls were ancestrally comparable.

To assess the relationship between specific HLA alleles and HL, we imputed key class I and II HLA alleles at 2- and 4-digit resolutions, using the algorithms described earlier.^{11,12} Comparison between observed and imputed HLA alleles for a subset of controls, for which we had experimentally determined HLA types, showed a high degree of concordance (> 90% at the 4-digit level; see “Statistical and bioinformatics analysis”). For each person and locus, the imputation methodology returns probabilities associated with the sample having each allelic type; these probabilities have been shown to be reasonably well calibrated.¹¹ Thus, we were able to incorporate these probabilities into the general logistic regression framework, to account for any uncertainty associated with imputed types.

We initially considered 1700 SNPs mapping to the 4.5-Mb region that encompasses the classic MHC region at 6p21.3. Figure 1 shows the strength of the single SNP associations across this region. A total of 25 SNPs showed evidence for an association with HL risk at $P < 10^{-7}$, most mapping to HLA class II regions. The strongest single-SNP based association was attained for rs6903608 mapping at 32 536 263 bp, about 16 kb centromeric to *HLA-DRA* ($P = 6.63 \times 10^{-19}$; Figure 1A; supplemental Table 1).

Five HLA class II alleles showed evidence of association with HL risk at the 1×10^{-7} threshold in the unconditional logistic regression analysis, namely, DQA1*02:01, DRB1*07:01, DQB1*03:03, DRB1*15:01, and DQB1*06:02 (supplemental Table 1). The associations at DQA1*02:01, DRB1*07:01, and DQB1*03:03 confer an increased risk of HL, whereas the associations at DRB1*15:01 and DQB1*06:02 confer a reduced risk. Some of these alleles, such as DRB1*15:01 and DQB1*06:02, are known to be in strong LD in European populations. The strongest association was provided by DQA1*02:01 (odds ratio [OR] = 0.45, $P = 6.17 \times 10^{-15}$). Of the class I HLA alleles, the strongest association observed was for B*07:02, although it did not reach genome-wide significance (supplemental Table 1; OR = 1.39, $P = 4.40 \times 10^{-5}$).

To evaluate the independence of associations, we conducted stepwise logistic regression, jointly on SNPs and imputed HLA

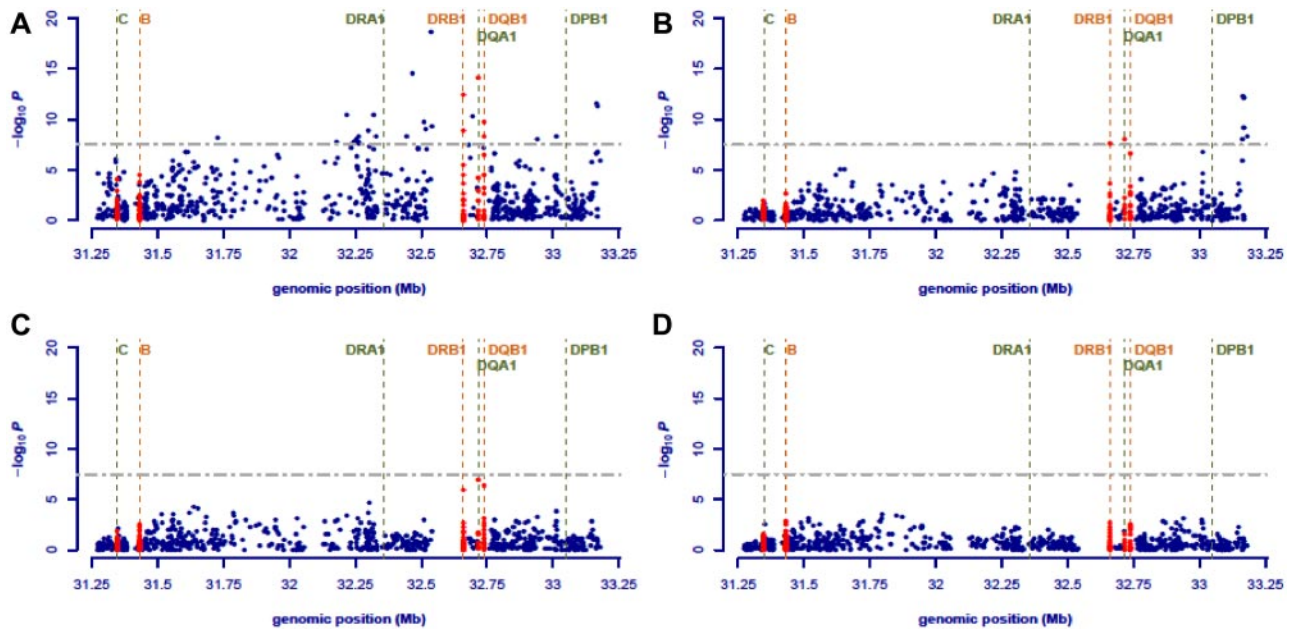


Figure 1. Associations in the MHC region and conditional logistic regression analysis. Plots of the evidence for association between the HLA and cHL, as covariates, are added to the model in a stepwise manner. (A-D) The horizontal axis indicates the location of each SNP along chromosome 6, and the vertical axis, the $-\log_{10} P$ values of each SNP (blue) and HLA allele (red), obtained using logistic regression models with different numbers of covariates. (A) Evidence for association obtained from the 1-parameter logistic regression analysis. (B) The remaining signal when rs6903608 is included as a covariate. (C) The signal when rs2281389 is also included in the model. (D) The remaining signal when DQA1*02:01 is also added as a covariate. Multiple red dots on the same locus indicate the $-\log_{10} P$ values for different allelic subtypes at the respective locus. The positions of common HLA loci have been marked with vertical dashed lines and annotated accordingly. Note that some loci (DRA1, DPB1) for which imputation results were not available have also been marked in the figure. For all loci, the vertical lines indicate the mean position of the longest transcript. The yellow horizontal line indicates the commonly used 5×10^{-8} threshold for genome-wide significance.

alleles, initially including rs6903608 as a covariate in our model. Conditional analysis showed that most, but not all, of the class II variation defined by the SNP genotype could be explained for by rs6903608 (Figure 1B; supplemental Table 2). The strongest secondary signal appeared to be for SNPs rs9277565 (OR = 1.70, $P = 4.57 \times 10^{-13}$) and rs2281389 (OR = 1.73, $P = 6.31 \times 10^{-13}$), which map closely to *HLA-DPB1* and one of which, rs9277565, has previously been shown to be in LD ($r^2 = 0.37$) with *HLA-DPB1**03:01.²⁰ Note that classic alleles at *HLA-DPB1* were not imputed here because of lack of training data. Inclusion of rs2281389 in the logistic model accounted for most of the remaining signal (Figure 1C; supplemental Table 3), except for the protective effect of *HLA-DQA1**02:01 (OR = 0.56, $P = 1.51 \times 10^{-7}$). *DQA1**02:01 is nearly perfectly correlated with *DRB1**07:01, and it is unclear which of the 2 alleles drives the observed signal; here for simplicity, we only refer to *DQA1**02:01. When rs6903608, rs2281389, and *DQA1**02:01 were all included as covariates to the model, no additional loci showed genome-wide evidence for association (Figure 1D; supplemental Table 4).

Although genetic variation defined by rs6903608 provides the strongest association signal, collectively these data provide evidence for 2 additional independent disease loci, defined by rs2281389 and *HLA-DQA1**02:01 (Figure 1; supplemental Tables 2-3). rs2281389 lies in close proximity to *HLA-DPB1*, variation in which has been associated with HL risk in a number of previously published studies,^{5,21} with *DPB1**03:01 appearing to confer susceptibility (OR = 1.42) and *DPB1**02:01 resistance (OR = 0.49) to HL.²¹

Discussion

To examine the relationship between HLA and the risk of developing HL, we have conducted a large case-control study and

have systematically examined the relationship between genetic variants and HL risk for the entire 4.5-Mb 6p21.3 region that encodes the MHC.

Using genome-wide association SNP data in principal components analysis has allowed us to minimize the potential problem of population stratification as formal statistical analysis provided no significant evidence that population substructure is a confounding factor in our study. The large dataset and dense set of informative markers in our study have also made possible the detection of independent effects at class II loci. Imputation of HLA alleles from SNP data has allowed us to also analyze systematically the relationship between HLA types and HL risk, at both the 2- and 4-digit levels of significance.

Our study provides unequivocal evidence for an association between the MHC and risk of developing HL. Moreover, and specifically, it provides evidence of a relationship between HLA class II genotypes and HL risk. Although we acknowledge that we did not impute all of the HLA alleles, the analysis we have conducted on a dense set of single SNPs in the region will have been sufficient to recover strong associations at untyped loci. Moreover, our imputations cover the alleles carried by 97.4% of the European population, according to dbMHC (averaged across all loci for which imputations are made). Our analysis found no evidence that class I variation also contributes to disease development, as the signals from the class I region were of moderate significance ($P > 10^{-5}$) and disappeared in the conditional analysis.

Although our study represents the most comprehensive interrogation of the relationship between genetic variation in MHC and HL risk to date, the analysis has limitations. A hallmark of HL epidemiology is a bimodal age-specific incidence, and it has been argued that the disease in young and older adults is etiologically different²²; in particular, the prevalence of EBV in tumors is higher in older HL patients ($\sim 50\%$ age ≥ 40 years compared with

~ 30% age < 40 years).²³ HLA class I polymorphisms have been reported to be associated with development of infectious mononucleosis on primary EBV infection,²⁴ and HLA-A*02 and HLA*01 alleles have been linked to the risk of developing EBV-positive HL.¹⁰ Such observations are in keeping with predominant class I restricted T-cell response to EBV infection. The majority of HL cases we have studied were young (93% age < 35 years at diagnosis); hence, our findings are more likely to apply to EBV-negative HL. This suggestion is reinforced by the lack of association between HLA-A*02 or HLA-A*01 alleles and HL in the current study ($P > .05$). Our results therefore suggest, albeit individually, substantially different genetic contributions to the risk of EBV-positive and EBV-negative HL within the MHC.

The neoplastic multinuclear HRS cells that typify cHL are derived from germinal center B cells that express HLA class II antigens.¹ Several observations suggest that antigen presentation is involved in the pathobiology of HL. HRS cells stimulate T-cell proliferation in mixed lymphocyte reactions and induce cytotoxic activity.^{25,26} Lack of membranous expression of HLA class II by HRS cells of HL is common at primary diagnosis and is associated with extranodal disease, lack of EBV in HRS cells, and absence of HLA class I expression.²⁷⁻³⁰ Expression of HLA class II by HRS cells is retained more frequently in EBV-associated cHL, and in HLA class I-positive HRS cells.²⁷⁻³⁰ Hence, a possible explanation for the association of HLA class II expression status might be that HLA class II directly activates T-helper 1 cells, thereby inducing and maintaining cytotoxic antitumor immune responses. Lack of HLA class II expression results in diminished activation of T-helper 2 and/or Treg cells. Therefore, the class II associations we identified may reflect differential binding affinity of the HLA class II molecule with immunogenetic antigenic peptides modifying class II down-regulation. This postulate is in keeping with class II variation being primarily associated with development of EBV-negative HL, something fully supported by our findings of strong class II associations.

The strongest association identified in our study was attained with rs6903608. Inherited susceptibility to HL is likely to be mediated through the coinheritance of multiple risk alleles. Under a multiplicative model of interaction between risk loci on the basis of the risks associated with heterozygosity and homozygosity for the C-risk allele of rs6903608 (supplemental Table 1), this variant accounts for approximately 6% of the increased risk of HL seen in first-degree relatives of patients. rs6903608 maps to an LD block of 13.8 kb from 32 503 014 to 32 516 820 bp on chromosome 6. *In silico* analysis of this region using EEL,¹⁹ PreMod,¹⁸ and Transfac¹⁷ software provided no evidence that the SNP is directly functional in terms of impacting on a regulatory sequence. As rs6903608 localizes within 16 kb of *HLA-DRA*, it is possible that the SNP tags

this locus. Unfortunately, because of the lack of training data, we were not able to impute alleles at this locus and thus examine this possibility. Nonetheless, our analysis of the joint SNP and imputed HLA dataset indicates that none of the classic class I HLA alleles or class II genes *HLA-DRB*, *HLA-DQA*, *HLA-DQB* are likely to be driving this association.

In conclusion, our analysis provides unequivocal evidence that MHC variation is a major determinant of HL risk and specifically implicates class II variation at multiple loci in disease etiology. Given that our analysis is based on a high proportion of early-onset HL cases, further studies will allow the generalizability of our findings to all forms of HL to be addressed.

Acknowledgments

The authors thank the patients and their clinicians who participated in this collection.

This work was supported by the Leukemia and Lymphoma Research (United Kingdom), Cancer Research United Kingdom (C1298/A8362, supported by the Bobby Moore Fund), and the Wellcome Trust (awards 076113 and 085475). Sample and data acquisition was supported by Breakthrough Breast Cancer and the European Union. National Health Service funding was provided to the National Institute for Health Research Biomedical Research Center. This study made use of control genotyping data generated by the Wellcome Trust Case-Control Consortium. Genotype data were from the British 1958 Birth Cohort DNA collection (supported by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02).

Authorship

Contribution: L.M., V.E.-M., Y.P.M., S.L., A.D., and A.S. performed statistical and bioinformatic analyses; P.B. oversaw sample coordination and laboratory analyses; A.J.S., A.A., and R.C. provided samples and data from a study conducted at the Institute of Cancer Research; R.S.H., G.M., V.E.-M., and L.M. drafted the manuscript; and R.S.H. designed the study and obtained funding for the parent project.

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

A complete list of the investigators who contributed to the generation of the data is available at www.wtccc.org.uk.

Correspondence: Richard S. Houlston, Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, United Kingdom; e-mail: richard.houlston@icr.ac.uk.

References

- Kuppers R. The biology of Hodgkin's lymphoma. *Nat Rev Cancer*. 2009;9(1):15-27.
- Hjalgrim H, Engels EA. Infectious aetiology of Hodgkin and non-Hodgkin lymphomas: a review of the epidemiological evidence. *J Intern Med*. 2008;264(6):537-548.
- Mack TM, Cozen W, Shibata DK, et al. Concordance for Hodgkin's disease in identical twins suggesting genetic susceptibility to the young-adult form of the disease. *N Engl J Med*. 1995;332(7):413-418.
- Amiel J. Study of the leukocyte phenotypes in Hodgkin's disease. In: Teraski PI, ed. *Histocompatibility Testing*. Copenhagen, Denmark: Munksgaard; 1967:79-81.
- Klitz W, Aldrich CL, Fildes N, Horning SJ, Begovich AB. Localization of predisposition to Hodgkin disease in the HLA class II region. *Am J Hum Genet*. 1994;54(3):497-505.
- Oza AM, Tonks S, Lim J, Fleetwood MA, Lister TA, Bodmer JG. A clinical and epidemiological study of human leukocyte antigen-DPB alleles in Hodgkin's disease. *Cancer Res*. 1994;54(19):5101-5105.
- Tonks S, Oza AM, Lister TA, Bodmer JG. Association of HLA-DPB with Hodgkin's disease. *Lancet*. 1992;340(8825):968-969.
- Marshall WH, Barnard JM, Buehler SK, Crumley J, Larsen B. HLA in familial Hodgkin's disease: results and a new hypothesis. *Int J Cancer*. 1977;19(4):450-455.
- Niens M, van den Berg A, Diepstra A, et al. The human leukocyte antigen class I region is associated with EBV-positive Hodgkin's lymphoma: HLA-A and HLA complex group 9 are putative candidate genes. *Cancer Epidemiol Biomarkers Prev*. 2006;15(11):2280-2284.
- Niens M, Jarrett RF, Hepkema B, et al. HLA-A*02 is associated with a reduced risk and HLA-A*01 with an increased risk of developing EBV+ Hodgkin lymphoma. *Blood*. 2007;110(9):3310-3315.
- Leslie S, Donnelly P, McVean G. A statistical method for predicting classical HLA alleles from SNP data. *Am J Hum Genet*. 2008;82(1):48-56.

12. Dillthey A, Moutsianas L, Leslie S, McVean G. HLA*IMP: an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics*. 2011;27(7):968-972.
13. Enciso-Mora V, Broderick P, Ma Y, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). *Nat Genet*. 2010;42(12):1126-1130.
14. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol*. 2006;35(1):34-41.
15. Houlston RS, Ford D. Genetics of coeliac disease. *QJM*. 1996;89(10):737-743.
16. Goldin LR, Bjorkholm M, Kristinsson SY, Turesson I, Landgren O. Highly increased familial risks for specific lymphoma subtypes. *Br J Haematol*. 2009;146(1):91-94.
17. Matys V, Kel-Margoulis OV, Fricke E, et al. TRANSFAC and its module TRANSCmpel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res*. 2006;34(Database issue):D108-D110.
18. Ferretti V, Poitras C, Bergeron D, Coulombe B, Robert F, Blanchette M. PReMod: a database of genome-wide mammalian cis-regulatory module predictions. *Nucleic Acids Res*. 2007;35(Database issue):D122-D126.
19. Hallikas O, Paliin K, Sinjushina N, et al. Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell*. 2006;124(1):47-59.
20. Field J, Browning SR, Johnson LJ, et al. A polymorphism in the HLA-DPB1 gene is associated with susceptibility to multiple sclerosis. *PLoS ONE*. 5(10):e13454.
21. Taylor GM, Gokhale DA, Crowther D, et al. Further investigation of the role of HLA-DPB1 in adult Hodgkin's disease (HD) suggests an influence on susceptibility to different HD subtypes. *Br J Cancer*. 1999;80(9):1405-1411.
22. MacMahon B. Epidemiology of Hodgkin's disease. *Cancer Res*. 1966;26(6):1189-1201.
23. Glaser SL, Lin RJ, Stewart SL, et al. Epstein-Barr virus-associated Hodgkin's disease: epidemiologic characteristics in international data. *Int J Cancer*. 1997;70(4):375-382.
24. McAulay KA, Higgins CD, Macsween KF, et al. HLA class I polymorphisms are associated with development of infectious mononucleosis upon primary EBV infection. *J Clin Invest*. 2007;117(10):3042-3048.
25. Lee SP, Constantinou CM, Thomas WA, et al. Antigen presenting phenotype of Hodgkin Reed-Sternberg cells: analysis of the HLA class I processing pathway and the effects of interleukin-10 on Epstein-Barr virus-specific cytotoxic T-cell recognition. *Blood*. 1998;92(3):1020-1030.
26. Murray PG, Constantinou CM, Crocker J, Young LS, Ambinder RF. Analysis of major histocompatibility complex class I, TAP expression, and LMP2 epitope sequence in Epstein-Barr virus-positive Hodgkin's disease. *Blood*. 1998;92(7):2477-2483.
27. Miller TP, Lippman SM, Spier CM, Slymen DJ, Grogan TM. HLA-DR (Ia) immune phenotype predicts outcome for patients with diffuse large cell lymphoma. *J Clin Invest*. 1988;82(1):370-372.
28. Rimsza LM, Roberts RA, Miller TP, et al. Loss of MHC class II gene and protein expression in diffuse large B-cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival regardless of other prognostic factors: a follow-up study from the Leukemia and Lymphoma Molecular Profiling Project. *Blood*. 2004;103(11):4251-4258.
29. Roberts RA, Wright G, Rosenwald AR, et al. Loss of major histocompatibility class II gene and protein expression in primary mediastinal large B-cell lymphoma is highly coordinated and related to poor patient survival. *Blood*. 2006;108(1):311-318.
30. Amiot L, Onno M, Lamy T, et al. Loss of HLA molecules in B lymphomas is associated with an aggressive clinical course. *Br J Haematol*. 1998;100(4):655-663.