Published in final edited form as:

Leuk Res. 2012 November ; 36(11): 1359–1364. doi:10.1016/j.leukres.2012.07.014.

Using Haplotype Analysis to Elucidate Significant Associations between Genes and Hodgkin Lymphoma

Anthony M. D'Amelio Jr., PhD^{1,2}, Claudia Monroy, MS², Randa El-Zein, MD PhD², and Carol J. Etzel, PhD²

¹The University of Texas Graduate School for Biomedical Sciences, Houston, TX, 77030

²Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX

Abstract

In this study, we estimated the association between the inferred haplotypes in the inflammation, DNA repair, and folate pathways, and developed risk models for Hodgkin Lymphoma. The study population consisted of 200 Hodgkin Lymphoma cases and 220 controls. A susceptible association was observed on the XPC gene with Haplotype CT (rs2228001 and rs2228000), and a protective association was observed on the IL4R gene with Haplotype TCA (rs1805012, rs1805015, and rs1801275). These results can provide the necessary tools to identify high-risk individuals after validation in large data sets.

Keywords

Single Nucleotide Polymorphism; Haplotypes; Hodgkin lymphoma; DNA Repair; Inflammation

1. Introduction

In 2012, approximately 9,060 individuals were diagnosed in the United States (US) with Hodgkin lymphoma (HL) with about 1,190 deaths [1]. Incidence of HL increased to 4.5 cases per 100,000 in individuals between the ages of 20 and 25, decreased slightly through ages 35 and 60, and then increased again to over 4 cases per 100,000 individuals above the age of 70, with a peak of 5.1 cases per 100,00 individuals between the age of 80 and 85 [2, 3]. Evidence suggests that genetic factors could be responsible for the incidence peaks in both young and old adults [4–6].

© 2012 Elsevier Ltd. All rights reserved.

Address for correspondence and reprint requests: Dr. Carol J. Etzel, The University of Texas M. D. Anderson Cancer, Department of Epidemiology, 1515 Holcombe Boulevard, Unit 1340, Houston, TX 77030, Phone: +1-713-745-2468, Fax: +1-713-792-8261, cetzel@mdanderson.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Authors' Contributions:

A.M.D. was involved in the analysis of the results, the design and performance of the statistical portion of the study, and wrote the manuscript. C.M. conducted the genotyping and critically reviewed the manuscript, R.E.Z. was involved in the overall design of the study, collected the clinical data of all involved with the study, oversaw the genotyping and critically reviewed the manuscript. C.J.E. was involved in the overall design of the study, including the statistical methodology design, and critical reviewed of the manuscript.

Conflict of interest:

All authors have no financial support or conflict of interest to disclose.

Epidemiological studies have shown associations between hundreds of single nucleotide polymorphisms (SNPs) and cancer risk [7–10]. Several studies have suggested that many of these polymorphisms result in amino acid substitutions which may alter wild-type protein function and lead to substantial changes in protein levels, thereby contributing to cancer susceptibility. This is especially true with modifications in the DNA repair, folate, and inflammation pathways [11–21]. However, the single-SNP based strategy, in which each SNP is analyzed individually, may be insufficient to explain the risk of developing cancer, as complex diseases most likely result from genetic variants in multiple genes. Recently, several studies have shown that epistatic interactions of multiple SNPs are necessary to contribute to cancer susceptibility [22–25].

Haplotype analysis is the study of a pattern of descent of a set of linked alleles occurring on the same chromosome. Several programs that infer haplotypes are Haploview[26], PHASE [27, 28], and Haplo.stats [29]. Haplo.stats also uses a joint-effects logistic model to estimate an odds ratio for the association between either a haplotype or a covariate with disease. In a joint-effects model, each haplotype is compared to the most frequent haplotype under the assumption that the most frequent haplotype represents the “normal” haplotype status [30]. Further we have shown, in previous studies, that variants among SNPs and combinations of haplotypes containing two SNPs can modulate HL risk [6, 31, 32]. In this current study, we build upon our previous findings, and hypothesize that gene-gene interactions between candidate SNPs in the DNA repair (*MGMT*, *XPC*), folate pathways (*MTHFR*), and inflammation (*IL1B*, *IL4*, *IL4R*, *IL10*) may contribute to HL susceptibility. To test this hypothesis, we used haplotype analysis to discover sets of linked SNPs that could potentially elucidate the association of multiple polymorphisms in these three pathways and the risk of HL.

2. Materials and Methods

2.1. Study Population and SNP Selection

Details of the study population were published elsewhere [6]. Briefly, the study included 200 histological confirmed adult HL cases registered at the University of Texas MD Anderson Cancer Center and 220 controls accrued through random-digit dialing and frequency matched to the cases on age (± 5 years), sex, and race/ethnicity. Interviews were conducted by trained interviewers and demographic and clinical characteristics included: age at diagnosis, smoking status, sex, race/ethnicity, family history of cancer, HL histological subtypes, disease stage and presence of B symptoms. Sixty-two SNPs in DNA repair, inflammation, and folate pathway genes were genotyped using Taq-Man-based methods [6]. The location (chromosome and base-pair information), and the reference SNP (rs) number for each SNP is listed in Supplemental Table 1. For quality control and to ensure proper genotyping of samples, 5% of all samples were selected randomly for repeat analysis. The institutional review board at the University of Texas MD Anderson Cancer Center approved the study.

2.2. Haplotype Analysis

Inferred haplotype block construction was completed using the program Haploview, which was an open source program written completely in JAVA [26], using mainly the Gabriel confidence interval method option. This method stated that a block of SNPs were significant if 95% of all informative comparisons of the SNPs in question show strong linkage disequilibrium (LD) [33]. Also, for all the chromosomes for which SNPs exist, gold plots were constructed to further visualize the haplotype block used in this analysis. For chromosome 16, since there was no substantial difference in base-pair length between the two haplotypes constructed with the Gabriel method, these haplotypes were combined to

form a single haplotype with 3 SNPs: rs1805012, rs1805015, and rs1801275. Figure 1 had some examples of gold plots constructed with the Gabriel confidence interval method plus the three SNP haplotype for chromosome 16.

To calculate the differences in haplotype frequency and to conduct joint-effect modeling within each haplotype block, the program haplo.stats was used. Haplo.stats was a program developed in the programming language R that used an Expectation-Maximization algorithm to first determine the frequency of haplotypes for each individual, and then determine which haplotypes and haplotype blocks show significant differences between cases and controls with the use of a haplotype-specific score [29, 30]. This haplotype-specific score calculated a statistic that can be tested for significance using a chi-square distribution with degrees of freedom equal to the amount of inferred haplotypes that exist in each haplotype block. For this analysis, the score was measured in the presence of the demographic covariates (age, sex, smoking status, and race), conducted with the assumption of additive and dominant genetic traits separately, and included haplotypes with frequency greater than 1%.

Testing for association with HL (as measured by the odds ratio), with accounting for age, sex, smoking status, and race, was also analyzed using haplo.stats. The haplotype analyses were based on the regression-based method for binary (case-control) responses as described by Schaid et al. 2002 [29] and Lake et al. 2003 [30]. For each haplotype block, we fit the joint-effects model in which each haplotype is compared to the most frequent haplotype (which was used as the reference group) [29, 30]. Analyses were conducted with these haplotype blocks assuming both additive and dominant genetic traits, and only haplotypes with frequencies greater than 1% were included. We applied the Benjamini-Hochberg procedure [34] to account for multiple testing.

2.3. Developing a Preliminary Genetic Risk Model for HL

As a parallel analysis to the above, preliminary risk models for HL were developed to determine the effectiveness of combining the most significant haplotypes from each haplotype block. We selected the haplotypes with the lowest p-values, and had frequencies of at least 1%. In addition, all haplotypes were modeled in the presence of the demographic variables, age, sex, smoking status, and race, using SPSS. Multivariable logistic risk models with non-genetic and genetic data were built assuming additive genetic traits and dominant genetic traits, separately, and haplotypes with adjusted p-values of less than 0.25 were used in the final model. We estimated the level discriminatory power for the model using the relative risk profiles calculated by the multivariable logistic risk models in SPSS, and any differences in discriminatory power between the additive and dominant genetic models were tested with NCSS/PASS.

3. Results

Demographic analysis showed that controls (49 years) were on average two years older than cases (47 years), but this was within the ± 5 year criterion. The majority of the participants were male (54% of cases and 56% of controls, $P=0.625$ from Fisher's exact test). There was a borderline significant difference in the percentage of Non-Caucasians among cases (18%) and controls (11%, $P=0.053$). Although not a part of the frequency-matching criteria, there was no difference in ever smoking status among cases (46%) and controls (45%; $P=0.922$).

Figure 1 showed the haplotypes on chromosomes 1 and 16, as determined by the gold plot. For example, using the SNPs on chromosome 1 that were genotyped in association with this HL study, the first two SNPs, and SNPs rs1800872, rs1800871, and rs1800896 were linked

together in separate blocks. With chromosome 16, separate haplotype blocks were inferred for SNPs rs1805012 and rs1805015, and for SNPs rs1801275 and rs1805016.

Comparisons of haplotype frequencies were summarized in Table 1. Only three inferred haplotypes blocks resulted in significant differences between cases and controls at the 10% level. These were the XPC haplotype block (rs2228001 and rs2228000) on chromosome 3 (p-value= 0.06 assuming additive genetic traits and p-value = 0.04 assuming dominant genetic traits), the IL4 haplotype block (rs2243250 and rs2070874) on chromosome 5 (p-value=0.07 assuming dominant genetic traits), and the IL4R haplotype block (rs1805012, rs1805015, and rs1801275) on chromosome 16 (p-value=0.08 assuming additive genetic traits and p-value = 0.04 assuming dominant genetic traits). On chromosomes 3, 5, and 16, the largest differences in haplotype frequency between cases and controls were observed with haplotype CT, haplotype CC, and haplotype TTG, respectively.

Associations between inferred haplotypes and HL are summarized in Table 2 assuming both additive and dominant genetic traits. In these set of analyses, each haplotype is modeled separately adjusting for age, sex, race, and smoking status. Significant associations (unadjusted p-value <0.05) were observed for haplotypes on chromosomes 3 and 16, and borderline significant associations (unadjusted p-value<0.10) were observed for haplotypes on chromosomes 1. A significantly increased risk of HL was observed with haplotype CT (XPC: rs2228001 and rs2228000) assuming either dominant genetic traits (Odds Ratio= 1.75, 95% CI = 1.15–2.65, unadjusted p-value = 0.009) or additive genetic traits (Odds Ratio = 1.49, 95% CI = 1.03–2.16, unadjusted p-value = 0.037). A significant protective association between inferred haplotypes and HL was calculated for chromosome 16 with haplotype TCA (IL4R: rs1805012, rs1805015, and rs1801275) assuming either additive genetic traits (Odds Ratio = 0.53, 95% CI = 0.31–0.93, unadjusted p-value = 0.026) or dominant genetic traits (Odds Ratio = 0.42, 95% CI = 0.23–0.79, unadjusted p-value = 0.007). Another possible protective association was observed assuming additive genetic traits with haplotype CC (MTHFR gene: rs1801131 and rs1801133), (Odds Ratio = 0.73, 95% CI = 0.52–1.04, unadjusted p-value = 0.079). When accounting for multiple testing, haplotypes TCA on chromosome 16 and CT on chromosome 3 remained at least borderline significant (adjusted p-value=0.059 for both) assuming dominant genetic traits.

We then included those haplotypes that had an adjusted p-value <0.25 into one multivariable genetic risk model for HL, adjusting for age, sex, smoking status, and race (Table 3). Non-Caucasian status for race was the most significant non-genetic variable for models constructed assuming additive genetic traits for haplotypes (Odds Ratio = 2.01, 95% CI = 1.14–3.54, p-value = 0.017) and dominant genetic traits for haplotypes (Odds Ratio = 2.07, 95% CI = 1.16–3.67, p-value = 0.013). All of the haplotypes listed in the risk models were more significant in terms of p-value when assuming dominant genetic traits. When comparing these models, the dominant genetic model had a significantly ($P=0.004$) higher discriminatory power (AUC = 0.631, 95% CI = 0.575–0.681) compared to the additive genetic model (AUC = 0.613, 95% CI = 0.556–0.663), although the absolute difference in improvement was small (0.018). This significant result from a small improvement was due to the high level of covariance between the two models. We also compared these two genetic models to the one including only the demographic variables and observed that the non-genetic model had an AUC=0.586 (95% CI = 0.529–0.638). Although the absolute differences in AUCs between the non-genetic model and the additive model (diff=0.027, $P=0.303$) and the dominant model (diff=0.045, $P=0.091$) are smaller than the difference between the AUCs of the two genetic models, the higher p-values are due to the lower level of covariance between the non-genetic model and both genetic models.

4. Discussion

The purpose of this study was to use haplotype analysis to elucidate associations in HL that could not be obtained solely with single SNP based strategies and to develop risk models to examine HL risk. Although differences in frequency between case and control haplotypes were only borderline significant (p-value < 0.10) in the overall HL set assuming additive genetic traits (i.e., *XPC* gene on chromosome 3, and *IL4R* gene on chromosome 16), there were some haplotypes which showed either susceptibility to HL development or protection from HL development. A haplotype on the *IL4R* gene (rs1805012, rs1805015, and rs1801275) showed a potential significant protective association, and a haplotype on the *XPC* gene from chromosome 3 (rs2228001 and rs2228000) showed a potential susceptible association with HL. Also, non-Caucasians had approximately twice the risk of HL compared to Caucasians.

In previous studies, variant alleles on the Methylene tetrahydrofolate reductase (*MTHFR*) gene had been shown to be protective for non HL risk [14] and relapse cancer events when both variant alleles were present for *MTHFR* 677 C>T SNP (rs1801133) and the *MTHFR* 1298 A>C SNP (rs1801131) [35]. These variant alleles resulted in decreased *MTHFR* gene activity, which led to rises in methyl donors that limited possible breaks in cell mitosis [35]. A variation of genes that catalyzed the conversion of 5–10-methyl-tetrahydrofolate (THF) into 5-methyl-THF, the more commonly circulated form of folate, was shown to affect the risk of non-HL development [14]. Further, the alteration of the (*MTHFR*) gene has been shown to decrease the enzyme needed to catalyze the conversion into 5-methyl-THF by 30 to 60 percent [36, 37]. In this study, the decrease in risk of HL was only borderline significant in the presence of haplotypes inferred from these two SNPs while assuming additive genetic traits, but when assuming dominant genetic traits, the effect of *MTHFR* was minimal. More studies will be needed to elucidate the relationship between genes that regulate folate, and Hodgkin lymphoma.

SNPs in DNA repair genes, such as *XPC* and *XPB* genes, have been linked to many cancers [6, 38, 39]. DNA damage and subsequent repair are critical for maintaining genomic integrity. Modifications with SNPs in DNA repair genes may modulate the DNA repair phenotype, particularly when these SNPs are located within coding or regulating regions, leading to alterations in protein expression and repair enzymes [12, 13]. In our study, study subjects with both variant alleles for the *XPC* haplotype had an increased risk for HL. Interestingly enough, haplotypes with only the 1st *XPC* SNP as a variant allele showed little association with HL, but the combination of the 2nd variant allele with the 2nd *XPC* SNP caused increased cancer risk. This finding also highlighted the strength of haplotype approaches to find potential genetic associations that may not be evident from single SNP analysis. For example, if SNP rs2228001 would have been tested alone, the association between DNA damage modification on the *XPC* gene and HL would have been missed.

In chronic inflammation, the cell death process is stopped, and cell growth is uncontrolled [16]. With HL, the inflammatory response is initiated by the presence of the malignant Hodgkin and Reed Sternberg cells [17] and is characterized by the massive presence of reactive inflammatory cells in response to these cells [20, 21]. Inflammation genes have been strongly associated with cancer, including stomach, breast, and HL [16–19, 31]. In this study, haplotypes consisting of variant alleles from SNP rs1801275 and SNP rs1805016 from gene *IL4R* gene show decreased risk of HL. A possible explanation for these results is that interleukin 4 receptors, IL4Rα and IL-13Rα1, are expressed in Hodgkin tumor cells [40, 41]. Therefore the common polymorphisms in promoter regions of genes related to the pro- and anti-inflammatory response may contribute to HL susceptibility and serve as plausible candidates for further study.

One interesting aspect of this study was that even though the interleukin 4 receptor gene had two haplotypes significant for protection from HL, the interleukin 4 gene did not have any significant haplotypes. One major reason for this lack of significance was the differences in association with HL when haplotypes were inferred in the different racial groups. For Caucasians, the TC haplotype led to a very significant susceptible association for HL (Odds Ratio = 6.73, 95% CI = 1.44–31.48, p-value = 0.001), while for non-Caucasians, the TC haplotype was very protective against HL development (Odds Ratio = 0.03, 95% CI = 0.0047–0.198, p-value = 0.001). Due to the small size of the non-Caucasian population (62 individuals), these results could be spurious. Therefore, future statistical and biological analysis needs to be conducted to examine whether the inflammation pathway works differently with HL development in Caucasians and non-Caucasians.

Risk models have been used to estimate risk in a wide variety of complex diseases, from cardiovascular disease to lung cancer. Recently, there have been attempts to add genetic markers to risk models already published in the literature. For example, SNPs that have been validated in Genome Wide Association Studies (GWAS) have been added to the Gail model for breast cancer, and the Spitz model for lung cancer in an attempt to improve discriminatory power of the original models [42, 43]. After adding 7 SNPs to the Breast Cancer Risk Assessment Tool (BCRAT), discriminatory power increased from 60.7% to 63.2%, and after adding 3 SNPs in the original Spitz lung cancer risk models, discriminatory power increased from 66.1% to 67.3%. In this study, we attempt to create the first risk model for Hodgkin lymphoma by adding both non-genetic and genetic factors together, and we obtain a discriminatory power result of 63.1% assuming dominant genetic traits, which is a borderline significant increase of 4.5% from a model with just non-genetic risk factors (58.6%). However, it should be noted that this discriminatory power is modest even with the inclusion of genetic factors. This increase of discriminatory power shows the potential strength of genetic-based risk models for HL, but more work needs to be done to develop stronger models.

One of our study limitations is the relatively small sample size, with 420 individuals available for analysis. Despite the small sample size, we believe this study has elucidated several genes with potential associated to HL worthy of further follow-up. Another limitation of this study is the lack of Epstein-Barr Virus (EBV) information in this dataset. There is evidence to suggest that HL development could depend on EBV status as males, and older non-white adults are more susceptible to the HL virus if they are EBV positive [44, 45]. Also, the literature suggests that rs6903608 is associated with HL among EBV-negative [45, 46] individuals. In conclusion, the interactions between specific SNPs in both the inflammation and the DNA repair pathway have potential to be viable markers for HL development. Future analysis could include conducting external validations to confirm the results obtained in this study. We feel that development of this model is a good first step in finding individuals that could have elevated risk of Hodgkin lymphoma.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Contents of this manuscript have been published as part of Anthony M. D'Amelio Jr's dissertation entitled "Development of a Bayesian Joint Logistic Model to Better Study the Association between Haplotypes and Disease". We like to thank Dr. Michelle McHugh for helping to edit this manuscript.

Funding Sources:

Our research was supported by a cancer prevention fellowship funded by National Cancer Institute grant K07CA093592, National Cancer Institute grants CA123. Anthony M. D'Amelio Jr. was also funded through the Marilyn and Frederick R. Lummis Jr, M.D., fellowship in Biomedical Sciences.

References

1. American Cancer Society. Cancer Facts & Figures 2012. Atlanta: American Cancer Society; 2012. <http://www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-031941.pdf>
2. Horner, MJ.; Ries, LAG.; Krapcho, M.; Neyman, N.; Aminou, R.; Howlader, N., et al., editors. SEER Cancer Statistics Review, 1975–2006. Bethesda, MD: National Cancer Institute; 2009. http://seer.cancer.gov/csr/1975_2006/, based on November 2008 SEER data submission, posted to the SEER web site,
3. Leukemia and Lymphoma Society. Facts 2012. White Plains, New York: Leukemia and Lymphoma Society; 2012. <http://www.lls.org/content/nationalcontent/resourcecenter/freeeducationmaterials/generalcancer/pdf/facts.pdf>
4. Diehl V, Tesch H. Hodgkin's Disease – Environment or Genetic? *N Engl J Med*. 1996; 332:461–462. [PubMed: 7824019]
5. Kamel OW, van de Rijn M, Weiss LM, Del Zoppo GJ, Hench PK, Robbins BA, et al. Reversible lymphomas associated with Epstein-Barr virus occurring during methotrexate therapy for rheumatoid arthritis and dermatomyositis. *N Engl J Med*. 1993; 328:1317–1321. [PubMed: 8385742]
6. El-Zein R, Monroy CM, Etzel CJ, Cortes AC, Xing Y, Collier AL, et al. Genetic Polymorphisms in DNA Repair Genes as Modulators of Hodgkin Disease Risk. *Cancer*. 2009; 115:1651–1659. [PubMed: 19280628]
7. Berwick M, Vineis P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J Natl Cancer Inst*. 2000; 92:874–897. [PubMed: 10841823]
8. Shen M, Purdue MP, Krick A. Polymorphisms in DNA repair genes and risk of non-Hodgkin's lymphoma in New South Wales, Australia. *Haematologica*. 2007; 92:1180–1185. [PubMed: 17666372]
9. Pakakasama S, Sirirat T, Kanchanachumpol S, Udomsubpayakul U, Mahasirimongkol S, Kitpoka P, et al. Genetic polymorphisms and haplotypes of DNA repair genes in childhood acute lymphoblastic leukemia. *Pediatr Blood Cancer*. 2007; 48:16–20. [PubMed: 16435384]
10. Brenner AV, Butler MA, Wang SS, Ruder AM, Rothman N, Schulte PA, et al. Single-nucleotide polymorphisms in selected cytokine genes and risk of adult Glioma. *Carcinogenesis*. 2007; 28:2543–2547. [PubMed: 17916900]
11. Shen H, Sturgis EM, Khan SG, Qiao Y, Shahlavi T, Eicher SA, et al. An intronic Poly(AT) polymorphism of the DNA repair gene XPC and risk of squamous cell carcinoma of the head and neck: a case-control study. *Cancer Res*. 2001; 61:3321–3325. [PubMed: 11309287]
12. Manuguerra M, Saletta F, Karagas MR, Berwick M, Vegila F, Vineis P, et al. XRCC3 and XPD/ERCC2 single nucleotide polymorphisms and the risk of cancer: a HuGE review. *Am J Epidemiol*. 2006; 164:297–302. [PubMed: 16707649]
13. Park DJ, Stoehlmacher J, Zhang W, Tsao-Wei D, Groshen S, Lenz HJ. A Xeroderma pigmentosum group D gene polymorphism predicts clinical outcome to platinum-based chemotherapy in patients with advanced colorectal cancer. *Cancer Res*. 2001; 61:8654–8658. [PubMed: 11751380]
14. Kim KN, Lee IK, Kim YK, Tran HT, Yang DH, Lee JJ, et al. Association between folate-metabolizing pathway polymorphism and non-Hodgkin lymphoma. *Br J Haematol*. 2008; 140:287–294. [PubMed: 18042267]
15. Frosst P, Blom HJ, Milos R, Goyette P, Sheppard CA, Matthews RG. A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat Genet*. 1995; 10:111–113. [PubMed: 7647779]
16. Coussen LM, Werb S. Inflammation and Cancer. *Nature*. 2002; 420:860–867. [PubMed: 12490959]

17. Ernst PB, Gold BD. The disease spectrum of *Helicobacter pylori*: the immunopathogenesis of gastroduodenal ulcer and gastric cancer. *Annu Rev Microbiol.* 2000; 54:615–640. [PubMed: 11018139]
18. Khan G. Epstein-Barr virus, cytokines, and inflammation: A cocktail for the pathogenesis of Hodgkin's lymphoma? *Exp Hematol.* 2006; 34:399–406. [PubMed: 16569586]
19. Balkwill F, Coussens LM. Cancer: An inflammatory link. *Nature.* 2004; 431:405–406. [PubMed: 15385993]
20. Skinnider BF, Mak TW. The role of cytokines in classical Hodgkin lymphoma. *Blood.* 2002; 99:4283–4287. [PubMed: 12036854]
21. Yung L, Linch D. Hodgkin's lymphoma. *Lancet.* 2003; 36:943–951. [PubMed: 12648984]
22. Vineis P, Anttila S, Benhamou S, Spinola M, Hirvonen A, Kiyohara C, et al. Evidence of gene gene interactions in lung carcinogenesis in a large pooled analysis. *Carcinogenesis.* 2007; 28:1902–1905. [PubMed: 17307802]
23. Galvan A, Loannidis JP, Dragani TA. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet.* 2010; 26:132–141. [PubMed: 20106545]
24. Ribas G, Gonzalez-Neira A, Salas A, Milne RL, Vega A, Carracedo B, et al. Evaluating Hap-Map SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. *Human Genet.* 2006; 118:669–679. [PubMed: 16323010]
25. Hartman M, Loy EY, Ku CS, Chia KS. Molecular epidemiology and its current clinical use in cancer management. *Lancet Oncol.* 2010; 11:383–390. [PubMed: 20359664]
26. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21:263–265. [PubMed: 15297300]
27. Stephens M, Smith N, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 2001; 68:978–989. [PubMed: 11254454]
28. Stephens M, Scheet P. Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *Am J Hum Genet.* 2005; 76:449–462. [PubMed: 15700229]
29. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet.* 2002; 70:425–434. [PubMed: 11791212]
30. Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, et al. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Human Hered.* 2003; 55:56–65.
31. Monroy CM, Cortes AC, Lopez MS, D'Amelio AM, Etzel CJ, Younes A, et al. Hodgkin disease risk: role of genetic polymorphisms and gene-gene interactions in inflammation pathway genes. *Mol Carcinog.* 2011; 50:36–46. [PubMed: 21061265]
32. Monroy CM, Cortes AC, Lopez M, Rourke E, Etzel CJ, Younes A, et al. Hodgkin lymphoma risk: Role of genetic polymorphisms and gene-gene interactions in DNA repair pathways. *Mol Carcinog.* 2011; 50:825–834. [PubMed: 21374732]
33. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The Structure of Haplotype Blocks in the Human Genome. *Science.* 2002; 296:2225–2229. [PubMed: 12029063]
34. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J.R. Stat. Soc. Ser. B.* 1995; 57:289–300.
35. Kuzelicki NK, Milek M, Jazbec J, Mlinaric-Rascan I. 5,10-Methylenetetrahydrofolate reductase (MTHFR) low activity genotypes reduce the risk of relapse-related acute lymphoblastic leukemia (ALL). *Leuk Res.* 2009; 33:1344–1348. [PubMed: 19178944]
36. van der Put NM, Gabreëls F, Stevens EM, Smeitink JA, Trijbels FJ, Eskes TK, et al. A second common mutation in the methylenetetrahydrofolate reductase gene: an additional risk factor for neural-tube defects? *Am J Hum Genet.* 1998; 62:1044–1051. [PubMed: 9545395]
37. Luccock M. Folic acid: nutritional biochemistry, molecular biology, and role in disease processes. *Mol Genet Metab.* 2000; 71:121–138. [PubMed: 11001804]
38. Vogel U, Overvad K, Wallin H, Tjønneland A, Nexø BA, Raaschou-Nielsen O. Combinations of polymorphisms in XPD, XPC and XPA in relation to risk of lung cancer. *Cancer Lett.* 2005; 222:67–74. [PubMed: 15837542]

39. Hu Z, Wang Y, Wang X, Liang G, Miao X, Xu Y, et al. DNA repair gene XPC genotypes/haplotypes and risk of lung cancer in a Chinese population. *Int J Cancer*. 2005; 115:478–483. [PubMed: 15700316]
40. Kawakami M, Kawakami K, Kioi M, Leland P, Puri RK. Hodgkin lymphoma: therapy with interleukin-4 receptor–directed cytotoxin in an infiltrating animal model. *Blood*. 2005; 105:3707–3713. [PubMed: 15626735]
41. Murata T, Obiri NI, Puri RK. Structure of and signal transduction through interleukin-4 and interleukin-13 receptors. *Int J Mol Med*. 1998; 1:551–557. [PubMed: 9852261]
42. Gail MH. Discriminatory Accuracy from Single-Nucleotide Polymorphisms in Models to Predict Breast Cancer Risk. *J Natl Cancer Inst*. 2008; 100:1037–1041. [PubMed: 18612136]
43. Spitz MR, Amos CI, D'Amelio A Jr, Dong Q, Etzel C. Re: Discriminatory Accuracy from Single-Nucleotide Polymorphisms in Models to Predict Breast Cancer Risk. *J Natl Cancer Inst*. 2009; 101:1731–172. [PubMed: 19903803]
44. Glaser SL, Lin RJ, Stewart SL, Ambinder RF, Jarrett RF, Brousset P, et al. Epstein-Barr virus-associated Hodgkin's disease: epidemiologic characteristics in international data. *Int J Cancer*. 1997; 70:375–382. [PubMed: 9033642]
45. Urayama KY, Jarrett RF, Hjalgrim H, Diepstra A, Kamatani Y, Chabriert A, et al. Genome-wide association study of classical Hodgkin Lymphoma and Epstein-Barr virus status-defined groups. *J Natl Cancer Inst*. 2012; 104:240–253. [PubMed: 22286212]
46. Enciso-Mora V, Broderick P, Ma Y, Jarrett RF, Hjalgrim H, Hemminki K, 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; 42:1126–1130. [PubMed: 21037568]

a

b

Figure 1.

Gold Plots constructed from Single Nucleotide Polymorphisms on Chromosome 1 (1a) and Chromosome 16 (1b). The rs numbers for each SNP are listed at the top of each figure, and those SNPs bolded are the SNPs which forms a haplotype. Diamond values represent the D' for the 2 SNPs being examined. For example, the D' value for SNPs rs1801131 and rs1801133 is 0.95.

Table 1

Estimated frequencies for haplotypes on chromosomes 1–3, 5, 10, and 16

Gene (Chr.)	SNPs	bp	Haplo.	Haplotype Frequency: Cases (%)	Haplotype Frequency: Controls (%)	P-Value	
						Add	Dom
<i>MTHFR</i> (1)	rs1801131 rs1801133	11788742 11790644	AC	43.35	36.65	0.11	0.48
			AT	30.89	32.75		
			CC	25.76	29.82		
<i>IL10</i> (1)	rs1800872 rs1800871 rs1800896	203334802 203335029 203335292	ACA	31.39	30.46	0.57	0.66
			ACG	41.73	44.09		
			CTA	26.36	25.45		
<i>IL1B</i> (2)	rs1143627 rs16944	113310618 113311098	CT	36.17	36.10	1.00	0.94
			TC	62.95	62.79		
<i>XPC</i> (3)	rs2228001 rs2228000	14162450 14174889	CC	33.96	37.19	0.06	0.04
			CT	27.51	20.76		
			AC	38.53	42.05		
<i>IL4</i> (5)	rs2243250 rs2070874	132037053 132037609	CC	82.50	79.08	0.14	0.07
			TC	3.00	4.10		
			TT	14.50	16.82		
<i>MGMT</i> (10)	rs2308321 rs2308327	131455054 131455160	AA	89.50	90.45	0.45	0.46
			GG	10.50	9.55		
<i>IL4R</i> (16)	rs1805012 rs1805015 rs1801275	27281465 27281681 27281901	TTA	6.25	8.00	0.08	0.04
			TTG	79.46	74.05		
			TCA	5.61	8.86		
			CCA	8.40	9.09		

Chr = Chromosome
bp = base pair location
Haplo = Haplotypes
Add = Assuming Additive Genetic Trait
Dom = Assuming Dominant Genetic Trait
Significant results are bolded

Determining associations between haplotypes and HL using the joint effect logistic model controlling for age, sex, race, and smoking status

Table 2

Gene (Chm.)	Haplo.	Genetic Traits					
		Additive			Dominant		
		Odds Ratio (95% CI)	p-value	Adj. p-value	Odds Ratio (95% CI)	p-value	Adj. p-value
<i>MTHFR</i> (1)	AC AT CC	1.00 0.87(0.64–1.20) 0.73(0.52–1.04)	NA 0.414 0.079	NA 0.728 0.299	1.00 0.78(0.51–1.20) 0.80(0.53–1.22)	NA 0.226 0.300	NA 0.588 0.650
<i>IL10</i> (1)	ACA ACG CTA	1.09(0.77–1.54) 1.00 1.08(0.76–1.52)	0.617 NA 0.675	0.798 NA 0.798	1.20(0.79–1.82) 1.00 1.10(0.73–1.66)	0.385 NA 0.643	0.686 NA 0.727
<i>IL1B</i> (2)	CT TC	1.00(0.76–1.32) 1.00	0.992 NA	0.992 NA	0.97(0.65–1.45) 1.00	0.879 NA	0.879 NA
<i>XPC</i> (3)	CC CT AC	0.98(0.72–1.32) 1.49(1.03–2.16) 1.00	0.873 0.037 NA	0.946 0.241 NA	0.91(0.60–1.38) 1.75(1.15–2.65) 1.00	0.6710.009 NA 1.00	0.727 0.059 NA
<i>IL4</i> (5)	CC TC TT	1.00 0.53(0.25–1.16) 0.80(0.55–1.16)	NA 0.114 0.244	NA 0.299 0.485	1.00 0.59(0.25–1.37) 0.83(0.54–1.30)	NA 0.223 0.422	NA 0.588 0.686
<i>MGMT</i> (10)	AA GG	1.00 1.19(0.75–1.89)	NA 0.448	NA 0.728	1.00 1.14(0.69–1.87)	NA 0.607	NA 0.727
<i>IL4R</i> (16)	TTA TTG TCA CCA	0.65(0.39–1.11) 1.00 0.53(0.31–0.93) 0.89(0.54–1.46)	0.115 NA 0.026 0.636	0.299 NA 0.241 0.798	0.60(0.33–1.11) 1.00 0.42(0.23–0.79) 0.85(0.50–1.43)	0.107 NA 0.007 0.537	0.464 NA 0.059 0.727

NA = Not Applicable
 Significant results are bolded
 Chm. = Chromosome
 Haplo. = Haplotype
 Adj. P-Value = P-Value adjusted by using the Benjamini-Hochberg FDR procedure

Table 3

Model results without using any genetic factors, and model results using haplotypes inferred with haplo.stats assuming additive genetic traits and dominant genetic traits, separately.

Model Without Genetic Factors					
Covariates		Odds Ratio (95% Confidence Interval)	P-Value		
Constant		1.34	0.429		
Age (years)		0.99 (0.98–1.00)	0.187		
Smoke (Ever vs. Never)		1.10 (0.74–1.63)	0.637		
Sex (Male vs. Female)		0.88 (0.60–1.30)	0.515		
Race (Non-Caucasian vs. Caucasian)		1.79 (1.03–3.12)	0.038		
Models with Genetic Factors					
		Additive		Dominant	
Covariates		Odds Ratio (95% Confidence Interval)	p-value	Odds Ratio (95% Confidence Interval)	p-value
Constant		1.18	0.678	1.10	0.793
Age (years)		0.99 (0.98–1.01)	0.224	0.99 (0.98–1.01)	0.248
Smoke (Ever vs. Never)		1.05 (0.71–1.58)	0.796	1.05 (0.70–1.57)	0.831
Sex (Male vs. Female)		0.87 (0.59–1.29)	0.488	0.87 (0.59–1.30)	0.499
Race (Non-Caucasian vs. Caucasian)		2.01 (1.21–3.84)	0.017	2.07 (1.16–3.67)	0.013
Gene (Chromosome)	Haplotype	Odds Ratio (95% Confidence Interval)	p-value	Odds Ratio (95% Confidence Interval)	p-value
<i>XPC</i> (3)	CT	1.48 (1.05–2.09)	0.026	1.78 (1.19–2.66)	0.005
<i>IL4R</i> (16)	TCA	0.59 (0.34–1.01)	0.053	0.46 (0.25–0.85)	0.013

Significant results are bolded