

Mapping of the *IRF8* Gene Identifies a 3'UTR Variant Associated with Risk of Chronic Lymphocytic Leukemia but not Other Common Non-Hodgkin Lymphoma Subtypes

Susan L. Slager¹, Sara J. Achenbach¹, Yan W. Asmann¹, Nicola J. Camp², Kari G. Rabe¹, Lynn R. Goldin³, Timothy G. Call¹, Tait D. Shanafelt¹, Neil E. Kay¹, Julie M. Cunningham¹, Alice H. Wang¹, J. Brice Weinberg^{4,5}, Aaron D. Norman¹, Brian K. Link⁶, Jose F. Leis¹, Celine M. Vachon¹, Mark C. Lanasa⁴, Neil E. Caporaso³, Anne J. Novak¹, and James R. Cerhan¹

Abstract

Background: Our genome-wide association study (GWAS) of chronic lymphocytic leukemia (CLL) identified 4 highly correlated intronic variants within the *IRF8* gene that were associated with CLL. These results were further supported by a recent meta-analysis of our GWAS with two other GWAS of CLL, supporting the *IRF8* gene as a strong candidate for CLL risk.

Methods: To refine the genetic association of CLL risk, we conducted Sanger sequencing of *IRF8* in 94 CLL cases and 96 controls. We then conducted fine mapping by genotyping 39 variants (of which 10 were identified from sequencing) in 745 CLL cases and 1,521 controls. We also assessed these associations with risk of other non-Hodgkin lymphoma (NHL) subtypes.

Results: The strongest association with CLL risk was observed with a common single-nucleotide polymorphism (SNP) located within the 3' untranslated region (UTR) of *IRF8* (rs1044873, log additive OR = 0.7, $P = 1.81 \times 10^{-6}$). This SNP was not associated with the other NHL subtypes (all $P > 0.05$).

Conclusions: We provide evidence that rs1044873 in the *IRF8* gene accounts for the initial GWAS signal for CLL risk. This association appears to be unique to CLL with little support for association with other common NHL subtypes. Future work is needed to assess functional role of *IRF8* in CLL etiology.

Impact: These data provide support that a functional variant within the 3'UTR of *IRF8* may be driving the GWAS signal seen on 16q24.1 for CLL risk. *Cancer Epidemiol Biomarkers Prev*; 22(3); 461–6. ©2013 AACR.

Introduction

Chronic lymphocytic leukemia (CLL) is a B-cell malignancy and one of the most common non-Hodgkin lymphomas (NHL). It is estimated that 16,060 new cases and 4,580 deaths will occur in the United States in 2012 (1).

Our genome-wide association study (GWAS; 2) of CLL identified and validated 4 single-nucleotide polymorphism (SNP) variants in the interferon regulatory factor 8 (*IRF8*) gene on chromosome 16q24.1. These 4 SNPs

(rs305077, rs391525, rs2292982, and rs2292980) were intronic, had log-additive ORs ranging between 0.55 and 0.57, and were highly correlated with each other (all pairwise $r^2 = 0.99$, HapMap 2 CEU). These associations were further supported by a recent meta-analysis of 3 GWAS of CLL, which included ours (3).

The *IRF8* gene is a strong candidate for CLL biology. It is a transcription factor that is expressed in B cells and plays an important role in myeloid and B-cell development (4). Herein we conducted a more detailed evaluation of the *IRF8* gene to identify potentially functional genetic variants that are associated with CLL risk. Specifically, we conducted germline DNA sequencing in 94 CLL cases and 96 controls. The variants identified from sequencing and any SNPs identified from tagging the *IRF8* gene were then genotyped in 745 CLL cases and 1,521 controls. In addition, because NHL comprises a group of closely related B- and T-cell neoplasms, we explored the association of these *IRF8* variants in 1,699 patients with other fairly common NHL subtypes (586 diffuse large B-cell lymphomas [DLBCL], 588 follicular lymphomas, 230 marginal zone lymphomas [MZL], 158 T-cell lymphomas [TCL], and 137 mantle cell lymphomas [MCL]).

Authors' Affiliations: ¹College of Medicine, Mayo Clinic, Rochester, Minnesota; ²University of Utah School of Medicine, Salt Lake City, Utah; ³Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; ⁴Duke University Medical Center; ⁵Durham VA Medical Center, Durham, North Carolina; and ⁶University of Iowa College of Medicine, Iowa City, Iowa

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: Susan Slager, Mayo Clinic, 200 First Street SW, Rochester, MN 55905. Phone: 507-284-5965; Fax: 507-284-9542. E-mail: slager@mayo.edu

doi: 10.1158/1055-9965.EPI-12-1217

©2013 American Association for Cancer Research.

Materials and Methods

Study participants

Participants from 2 studies were included. The first is the Genetic Epidemiology of CLL (GEC) Consortium. GEC Consortium has the overall aim of investigating the genetic basis of CLL through the collection of CLL families (i.e., families with 2 or more relatives with CLL), and includes researchers from 7 institutions (Duke University/Veterans Administration (VA) Medical Center, the Mayo Clinic, the National Cancer Institute, the University of Minnesota/Minneapolis VA Medical Center, the University of Texas M.D. Anderson Cancer Center, CancerCare Manitoba, and the University of Utah). Family recruitment at each site occurs through hematology clinics or through the Internet. Medical records (including pathology reports) of CLL patients were reviewed when available to confirm that each case met the 1996 criteria (5) for CLL diagnosis, the criteria that was in effect through the study interval. Medical records were available for 95% of the familial CLL cases.

The Mayo Clinic Case–Control study of Lymphoma is a clinic-based study of incident cases and frequency-matched controls (based on age, sex, and residence; 6). Cases were newly diagnosed NHL/CLL patients seen at Mayo Clinic Rochester who were aged 18 years or older; a resident of Minnesota, Iowa, or Wisconsin; and HIV negative at the time of diagnosis. Controls were ascertained from patients visiting the General Internal Medicine clinic at Mayo Clinic; eligibility requirements included age 18 years or older and a resident of Minnesota, Iowa, or Wisconsin; controls were excluded if they had prior diagnoses of lymphoma, leukemia, or HIV infection. We additionally included newly diagnosed NHL/CLL patients from the University of Iowa and Mayo Clinic Lymphoma SPORE; these patients had the same eligibility requirements as the other cases in the Mayo Clinic Case–Control study, except that they could be a resident of any U.S. state. This analysis included cases and controls enrolled from 2002 through 2009. The NHL diagnoses were confirmed by study pathologists and classified according to the World Health Organization classification (7).

The institutional review boards at each study center approved these studies; all participants provided written informed consent.

Sanger sequencing

To identify potential functional variants, we sequenced the exons of the *IRF8* gene. DNA was extracted from buccal mucosal cells of 94 familial CLL cases from Mayo Clinic CLL pedigrees and from peripheral blood of 96 controls from the Mayo Clinic Case–Control Study. Purified DNA was amplified by PCR using primer pairs that span the *IRF8* exons. PCR fragments were sequenced at the Mayo Clinic DNA Sequencing Core Facility and analyzed using Mutation Surveyor software (Softgenetics). Sequencing was read in both

directions and scored manually by 2 independent reviewers; we observed 100% concordance between the 2 reviewers.

SNP identification and selection

We identified tagged SNPs using HapMap version 2 CEU samples. Our region was 5Kb up- and downstream of *IRF8*. Linkage disequilibrium (LD) blocks were defined within the region by $r^2 \geq 0.8$ using LDSelect program (8). We selected SNPs from each block that had minor allele frequency (MAF) of $\geq 5\%$. A total of 32-tagged SNPs were selected that provided 100% coverage of the gene.

Genotyping

A total of 39 *IRF8* variants (tagged SNPs and SNP identified by sequencing) were successfully genotyped in the GEC/Mayo Clinic Case–Control/SPORE samples as part of a larger genotyping project using a custom Illumina Infinium array (Illumina). Standard genotyping quality control procedures were conducted and included duplicate genotyping, dropping samples, or SNPs with call rates $<95\%$, and testing for Hardy–Weinberg equilibrium (HWE). We found $>99.9\%$ genotyping concordance among the 3,502 samples with duplicate genotypes. All of the *IRF8* variants had HWE $P > 0.01$ and SNP call rates $>99.9\%$.

Statistical analysis

Main analyses used SAS version 9.2 (SAS Institute Inc.). Tests for HWE were done using either the Pearson goodness-of-fit test or the Fisher's exact test, where appropriate among our controls. The association between each SNP and CLL/NHL risk was assessed by the Cochran–Armitage trend test. OR and 95% confidence intervals (CI) were calculated using logistic regression with and without adjustment for age and sex covariates. A conservative Bonferroni-corrected P -value threshold for the 39 variants tested was used for statistical significance (i.e., $0.05/39 = 0.0013$). We assessed for independence of association by jointly modeling additive effects of rs1044873 with all other variants in a logistic regression analyses. Tests for association between genotypes and *IRF8* mRNA expression were done using linear regression and publicly available expression and genotyped data (9) from the 60 unrelated CEU HapMap samples. LD metric (r^2) among variants in the *IRF8* gene was calculated using Haploview (10).

Bioinformatics

We used *phastCons* (11) and *phyloP* (12) from the PHAST package (<http://compgen.bscb.cornell.edu/phast/>) to identify conserved regions across all vertebrate species, as well as within 2 species subsets, primate and placental mammals. Because the 3' untranslated region (UTR) are common binding sites for miRNA genes, we used Poy-miRTS (13) to search for putative SNPs that may affect miRNA targeting in human.

Table 1. Sample characteristics

Controls (n = 1521)	
Mean age (SD), yrs	62 (13.7)
Number male (%)	777 (51.1)
CLL (n = 745)	
Mean age (SD), yrs	62 (11.1)
Number male (%)	510 (68.5)
Follicular lymphoma (n = 588)	
Mean age (SD), yrs	60 (13.1)
Number male (%)	314 (53.4)
DLBCL (n = 586)	
Mean age (SD), yrs	61 (15.6)
Number male (%)	313 (53.4)
MZL (n = 230)	
Mean age (SD), yrs	63 (12.4)
Number male (%)	102 (44.3)
MCL (n = 137)	
Mean age (SD), yrs	65 (12.4)
Number male (%)	105 (76.6)
T cell (n = 158)	
Mean age (SD), yrs	59 (15.6)
Number male (%)	93 (58.9)

Abbreviations: CLL, chronic lymphocytic leukemia; DLBCL, diffuse large B-cell lymphoma; MZL, marginal zone lymphoma; MCL, mantle cell lymphoma; T cell, T-cell lymphoma.

Results

All participants were non-Hispanic Caucasians. The age and sex characteristics for each NHL subtype and controls are described in Table 1. Among the CLL subtype, 10% of the cases were from CLL families that had multiple members with confirmed CLL. Only one CLL case per family was included in analyses.

To identify possible functional SNPs within the *IRF8* gene, we sequenced all 9 exons including 50 basepairs flanking each exon, 100 basepairs of the promoter, and 480 basepairs of the 3'UTR, in 94 familial CLL cases and 96 controls. We detected 13 variants, of which 7 were known SNPs with MAF > 0.05, 5 were known SNPs with 0.005 < MAF ≤ 0.05, and 1 novel variant seen in only one control individual. Variant call rates were all >99.4%, and significance for HWE tests were all $P > 0.2$. Association results with CLL risk for these variants are shown in Table 2. The most significant finding was with rs1044873 ($P = 2.82 \times 10^{-4}$). Of note, rs1044873 had the highest LD with the four GWAS SNPs based on HapMap version 2 CEU samples (r^2 ranging between 0.53 and 0.54 for each of the 4 SNPs).

To evaluate these sequencing variants in a larger sample and to fine map the *IRF8* gene, we genotyped the sequencing variants, along with tagged SNPs (see Materials and Methods), for a total 39 variants in 745 CLL cases and 1,521 controls. There were 156 subjects (89 controls and 67 CLL cases) who were genotyped by both Sanger sequencing and Illumina iSelect for 10 SNPs. Among the 1,560 duplicate observations, 99.8% were concordant. Significant associations ($P < 1.3 \times 10^{-3}$) were observed in 18 variants, which included the 4 GWAS SNPs that were previously identified, 4 of the sequencing SNPs (including rs1044873), and 10-tagged SNPs (Table 3). The 4 previously reported GWAS SNPs (rs305077, rs391525, rs2292982, and rs2292980) had statistical significance ranging between 1.9×10^{-5} and 4.7×10^{-5} . The most significant SNP is rs1044873, identified by sequencing and in moderately high LD (Supplementary Fig. S1) with the GWAS SNPs (OR = 0.72; 95% CI = 0.63–0.82; $P = 1.81 \times 10^{-6}$). The second most significant SNP was an intronic SNP, rs11649318 (OR = 0.72; 95% CI = 0.63–0.83; $P = 3.04 \times 10^{-6}$). It was moderately correlated with rs1044873 ($r^2 = 0.54$, Supplementary Fig. S1) and highly correlated with the 4 GWAS SNPs (all pairwise r^2 between 0.84 and 0.85)

Table 2. Sequencing results for the *IRF8* gene located on chromosome 16q24

SNP	Position	Risk allele	MAF		OR (95% CI)	P
			94 cases	96 controls		
rs8052064	85945231	T	0.0851	0.2135	0.33 (0.17–0.62)	4.22E-04
rs16939945	85945249	T	0.0319	0.0260	1.24 (0.37–4.21)	7.29E-01
rs17444416	85948098	A	0.0585	0.0625	0.94 (0.42–2.09)	8.78E-01
rs61995933	85952315	T	0.0000	0.0104	0 (0.00–0.00)	1.59E-01
rs28368116	85955194	C	0.0053	0.0052	1.02 (0.06–16.57)	9.88E-01
rs10514611	85955242	T	0.1436	0.2865	0.4 (0.23–0.68)	6.04E-04
rs1568391	85955304	T	0.3511	0.5260	0.49 (0.32–0.75)	7.95E-04
chr16_85955318	85955318	C	0.0000	0.0052	0 (0.00–0.00)	3.21E-01
rs305072	85955663	C	0.0213	0.0156	1.38 (0.30–6.33)	6.79E-01
rs1044873	85955671	T	0.2394	0.4115	0.43 (0.27–0.68)	2.82E-04
rs71714786	85955730	Indel	0.3656	0.5156	0.54 (0.36–0.82)	3.71E-03
rs28368119	85955948	T	0.0053	0.0000	0 (0.00–0.00)	3.11E-01
rs6638	85956044	T	0.3617	0.5104	0.55 (0.36–0.83)	4.11E-03

Table 3. Association of CLL risk with *IRF8* sequencing and tagged SNPs located on 16q24. 745 CLL cases and 1,521 controls were successfully genotyped

SNP	Position	Risk allele	MAF		OR (95% CI)	P	OR ^a (95% CI ^a)	P ^a
			Cases	Controls				
rs391023	85927814	A	0.2873	0.3419	0.75 (0.65–0.86)	5.24E-05	0.87 (0.73–1.02)	8.88E-02
rs375288	85927871	A	0.2604	0.3100	0.75 (0.65–0.87)	1.28E-04	0.87 (0.74–1.04)	1.20E-01
rs191022	85932132	C	0.2128	0.2377	0.83 (0.71–0.97)	2.01E-02	0.93 (0.79–1.10)	4.08E-01
rs2270501	85932988	A	0.1384	0.1716	0.75 (0.63–0.90)	1.86E-03	0.91 (0.74–1.12)	3.71E-01
rs2270502	85933038	A	0.0403	0.0487	0.80 (0.58–1.08)	1.48E-01	0.70 (0.51–0.96)	2.79E-02
rs305084	85934168	G	0.0946	0.0888	1.11 (0.89–1.38)	3.38E-01	0.99 (0.79–1.23)	9.05E-01
rs16882	85935573	G	0.1456	0.1795	0.77 (0.65–0.92)	3.55E-03	0.96 (0.78–1.19)	7.34E-01
rs12924316	85936263	A	0.1389	0.1736	0.75 (0.62–0.89)	1.34E-03	0.91 (0.74–1.12)	3.77E-01
rs3794661	85939666	A	0.0262	0.0289	0.87 (0.59–1.29)	4.98E-01	1.03 (0.69–1.53)	8.83E-01
rs305080	85941774	A	0.2705	0.3307	0.73 (0.63–0.84)	1.39E-05	0.90 (0.71–1.15)	4.16E-01
rs305079	85942496	G	0.0262	0.0309	0.82 (0.56–1.21)	3.17E-01	0.96 (0.65–1.42)	8.27E-01
rs305077	85943466	G	0.2765	0.3329	0.75 (0.65–0.86)	4.73E-05	0.97 (0.75–1.24)	8.02E-01
rs391525	85944439	G	0.2745	0.3324	0.74 (0.64–0.85)	2.66E-05	0.95 (0.74–1.22)	6.79E-01
rs2292982	85944823	C	0.2732	0.3324	0.74 (0.64–0.85)	1.93E-05	0.93 (0.72–1.19)	5.63E-01
rs2292980	85945076	G	0.2738	0.3330	0.74 (0.64–0.85)	2.06E-05	0.93 (0.73–1.20)	5.89E-01
rs8052064	85945231	A	0.1161	0.1492	0.75 (0.62–0.91)	2.76E-03	0.93 (0.75–1.15)	4.99E-01
rs16939945	85945249	A	0.0175	0.0250	0.67 (0.43–1.06)	8.70E-02	0.59 (0.37–0.94)	2.56E-02
rs12923978	85946324	G	0.1557	0.1811	0.81 (0.69–0.97)	1.79E-02	1.03 (0.84–1.27)	7.69E-01
rs11649318	85946481	G	0.3027	0.3705	0.72 (0.63–0.83)	3.04E-06	0.84 (0.68–1.03)	9.72E-02
rs903202	85947779	G	0.3987	0.4474	0.80 (0.70–0.91)	7.29E-04	0.99 (0.83–1.18)	9.00E-01
rs403038	85949071	A	0.1175	0.1483	0.77 (0.64–0.92)	5.29E-03	0.96 (0.77–1.19)	7.06E-01
rs305071	85949271	A	0.1013	0.1243	0.79 (0.65–0.97)	2.27E-02	0.99 (0.79–1.25)	9.61E-01
rs16939967	85949473	A	0.1315	0.1473	0.86 (0.72–1.03)	1.05E-01	1.11 (0.90–1.38)	3.33E-01
rs11117415	85950686	G	0.0349	0.0529	0.64 (0.46–0.89)	6.96E-03	0.77 (0.55–1.08)	1.35E-01
rs4843860	85950921	G	0.2430	0.2403	1.00 (0.86–1.15)	9.67E-01	1.09 (0.94–1.27)	2.46E-01
rs8058904	85951682	G	0.1470	0.1880	0.75 (0.64–0.89)	1.08E-03	0.95 (0.77–1.17)	6.16E-01
rs8064189	85951796	C	0.1946	0.2544	0.72 (0.62–0.84)	1.77E-05	0.85 (0.70–1.04)	1.14E-01
rs13338943	85952951	A	0.1081	0.0930	1.16 (0.95–1.43)	1.47E-01	1.03 (0.83–1.27)	7.99E-01
rs28368116	85955194	C	0.0067	0.0049	1.14 (0.51–2.58)	7.45E-01	1.06 (0.47–2.40)	8.89E-01
rs10514611	85955242	A	0.1826	0.2387	0.72 (0.62–0.84)	3.68E-05	0.88 (0.71–1.10)	2.58E-01
rs1568391	85955304	A	0.4362	0.4898	0.79 (0.70–0.90)	2.81E-04	1.02 (0.84–1.25)	8.43E-01
chr16_85955318	85955318	G	0.0000	0.0003	0.00 (0.00–0.00)	9.72E-01	0.00 (0.00–0.00)	9.73E-01
rs305072	85955663	G	0.0134	0.0148	0.88 (0.51–1.52)	6.54E-01	0.81 (0.47–1.40)	4.55E-01
rs1044873	85955671	A	0.3134	0.3840	0.72 (0.63–0.82)	1.81E-06	0.82 ^b (0.67 ^b –1.00 ^b)	5.30E-02 ^b
rs28368119	85955948	A	0.0040	0.0013	3.61 (1.00–13.09)	5.06E-02	3.10 (0.84–11.34)	8.81E-02
rs6638	85956044	T	0.4309	0.4869	0.78 (0.69–0.89)	1.66E-04	1.00 (0.82–1.23)	9.85E-01
rs880365	85959362	A	0.1946	0.2561	0.71 (0.61–0.83)	1.04E-05	0.84 (0.69–1.02)	7.63E-02
rs11648480	85960279	A	0.1812	0.2393	0.71 (0.61–0.83)	2.04E-05	0.86 (0.69–1.06)	1.55E-01
rs1472235	85960967	A	0.2248	0.2768	0.74 (0.64–0.86)	9.84E-05	0.95 (0.76–1.19)	6.40E-01

^aAssociations adjusted for rs1044873, age and sex.^bAssociation adjusted for rs11649318, age and sex.

based on our 1,521 control samples. As shown in Table 3, when we conditioned on our top hit (rs1044873) in the logistic analyses, the associations for the other 38 SNPs attenuated greatly (all $P > 0.02$), especially rs11649318 ($P = 0.097$). When we conditioned on rs11649318, rs1044873 remained borderline significant, $P = 0.053$ (Table 3). These data suggest that rs1044873 is capturing the associations in the region.

We next assessed association of the 39 *IRF8* SNPs across other NHL subtypes (Supplementary Tables S1–S5). No association with Bonferoni-corrected P -values ($P < 1.3 \times 10^{-3}$) was observed for follicular lymphomas, DLBCL, MZL, and MCL. However, for TCL risk, we observed a single association (rs305072, OR = 2.94; 95% CI = 1.54–5.60; $P = 0.001$). This result will need further replication. Of note, the rs1044873 and rs11649318 SNPs had all $P >$

0.05 and OR approximately 1.00 for all of the NHL subtypes. These results clearly indicate that the *IRF8* gene is a candidate gene specific to CLL risk.

We also evaluated the association of rs1044873 SNP with *IRF8* mRNA expression from lymphocytes using publicly available data and found no association with mRNA expression across the 3 genotype levels ($P = 0.27$). Finally, through bioinformatics, we assessed the functional significance of rs1044873. Rs1044873 is located in the 3'UTR of the *IRF8* gene and therefore is potentially located within a target region for miRNA. However our bioinformatics analysis does not support this. According to *PolymiRTS*, rs1044873 is not within any validated miRNA target nor is rs1044873 located within any conserved elements according to *phastCons* and *phyloP*.

Discussion

Our CLL GWAS identified the *IRF8* gene as a strong candidate for CLL risk with 4 intronic, highly correlated SNPs. To identify more significant variants than these intronic SNPs, we previously imputed genotypes in our GWAS cases and controls using HapMap version 2 CEU samples. One other intronic SNP (rs11649318) was identified that had greater association (OR = 0.54, $P = 3.2 \times 10^{-7}$) than that of our observed 4 genotyped SNPs and was also found to be highly correlated with our 4 GWAS SNPs. We were also able to impute 7 of our sequenced variants (including rs1044873) with high quality. However, they were not as significant as the 4 GWAS SNPs (e.g., rs1044873, OR = 0.62, $P = 3.2 \times 10^{-5}$). As a result, sequencing and fine mapping was needed to further refine the basis of association and potentially identify functional variants.

Through our comprehensive sequencing and fine-mapping efforts, we refined our CLL association of *IRF8* to rs1044873, located within the 3'UTR of *IRF8*. The 3'UTRs are sequences on the 3' end of mRNA that are not translated into protein and are common binding targets of miRNAs. We did not find this variant to be associated with *IRF8* mRNA expression based on publicly available data nor did we find it to be located within a target region of known miRNA. These *in silico* findings do not exclude the possibility of other and more complex associations between *IRF8* and CLL risk. For example, given that this gene is in the interferon family of transcription factors and is often associated with immune lymphocyte population activation, it is possible that multiple microenvironmental factors modify the *IRF8* expression levels (4).

This is the first study to evaluate associations between *IRF8* variants and other NHL subtypes. All but one of the subtype associations was nonsignificant at our Bonferroni threshold of significance. Bonferroni correction has been known to be conservative, especially when variants are correlated with each other. Relaxing our threshold to $P < 0.05$, we would detect 2 additional potential risk SNPs for follicular lymphomas (Supplementary Table S2). However, by and large, the results do not support association of

IRF8 with other NHL subtypes. We are aware however that the sample size for MCL and TCL are small (with 135 and 156 cases, respectively) allowing us only 64% and 70% statistical power, respectively, to detect a protective effect of 0.73 assuming a 0.05 error rate.

IRF8 is a strong candidate for CLL risk. *IRF8* has been recently shown to regulate the expression of *MDM2* (14), which inhibits p53 function. The inhibition of p53 is necessary to allow B cells to undergo DNA double-strand breaks, somatic hypermutation, and class switch recombination in response to specific pathogens. Furthermore, *IRF8* activates *BCL6*, a gene critical for B-cell development in the germinal center (15).

Our Sanger sequencing effort yielded only 1 novel rare variant identified in a control that was previously not reported in HapMap. This variant was subsequently reconfirmed via different genotyping platform. As the 1000 Genome Project (16) expands with more subjects sequenced with greater coverage, the need for sequencing individual genes within a particular study samples will diminish.

Strengths of our study include carefully designed studies with high level of confidence in our sequencing and genotyping, rigorous pathology review and classification, large sample size, and an unbiased sequencing effort of the coding regions of *IRF8*. That is, we did not limit our fine mapping to those variants that were correlated ($r^2 \geq 0.8$) with the original 4 GWAS SNPs. Had we done so, we would have missed rs1044873. A limitation of our study is that our samples consisted of only non-Hispanic Caucasians; this is mostly because of the fact that NHL is rare in other ethnicities. However, this limitation does allow us to minimize the effect of population stratification. As we have seen in our CLL GWAS (2) and our prior genotyping studies (17), we have very little evidence of population stratification in our work. One other point, our Sanger sequencing experiment had DNA extracted from buccal mucosal cells in the CLL cases and from peripheral blood lymphocytes in the controls. This difference of cell type for DNA extraction between cases and controls did not have any apparent effect on our results given the high concordance of genotype results between our Sanger sequencing and Illumina iSelect genotyping.

In conclusion, we provide strong evidence that rs1044873 within the *IRF8* gene accounts for the initial GWAS signal for CLL risk. Importantly, this association appears to be unique to CLL, with little support for association across the other common NHL subtypes, and suggests distinct etiologic pathways across the NHL subtypes. The location of rs1044873 within the 3'UTR region supports the hypothesis that *IRF8* mRNA expression would be altered; however, future work will be needed to assess this in more detail.

Disclosure of Potential Conflicts of Interest

N.E. Kay has commercial research support from Celgene and Gilead. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: S.L. Slager, N.J. Camp, T.G. Call, T.D. Shanafelt, J.R. Cerhan

Development of methodology: S.L. Slager, J.R. Cerhan

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S.L. Slager, N.J. Camp, L.R. Goldin, T.G. Call, T.D. Shanafelt, J.M. Cunningham, A.H. Wang, J.B. Weinberg, B.K. Link, J.F. Leis, M.C. Lanasa, N.E. Caporaso, A.J. Novak, J.R. Cerhan

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.L. Slager, S.J. Achenbach, Y.A. Asmann, K.G. Rabe, A.H. Wang, C.M. Vachon, N.E. Caporaso, J.R. Cerhan

Writing, review, and/or revision of the manuscript: S.L. Slager, S.J. Achenbach, Y.A. Asmann, N.J. Camp, K.G. Rabe, L.R. Goldin, T.G. Call, T.D. Shanafelt, N.E. Kay, J.M. Cunningham, A.H. Wang, J.B. Weinberg, A.D. Norman, B.K. Link, C.M. Vachon, M.C. Lanasa, N.E. Caporaso, J.R. Cerhan

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S.L. Slager, N.E. Kay, A.H. Wang, A.D. Norman, N.E. Caporaso

Study supervision: S.L. Slager, N.J. Camp, A.D. Norman

Acknowledgments

The authors thank the study participants and the study coordinators for work in recruitment.

Grant Support

In the GEC Consortium and Mayo Clinic SPORE Lymphoma case-control study, the work was supported in part by NIH grants CA118444 (S. L. Slager), CA148690 (S.L. Slager), CA97274 (J.R. Cerhan) and CA92153 (J. R. Cerhan). The genotyping at the Mayo Clinic Genotyping Core is supported, in part, by CA15083 (J.M. Cunningham). Support was also obtained by the Veterans Affairs Research Service, and by NIH CA134919 (M.C. Lanasa).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 31, 2012; revised December 14, 2012; accepted December 28, 2012; published OnlineFirst January 10, 2013.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012;62:10–29.
2. Slager SL, Rabe KG, Achenbach SJ, Vachon CM, Goldin LR, Strom SS, et al. Genome-wide association study identifies a novel susceptibility locus at 6p21.3 among familial CLL. *Blood* 2011;117:1911–6.
3. Slager SL, Skibola CF, Di Bernardo MC, Conde L, Broderick P, McDonnell SK, et al. Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. *Blood* 2012;120:843–6.
4. Wang H, Morse HC III. IRF8 regulates myeloid and B lymphoid lineage diversification. *Immunol Res* 2009;43:109–17.
5. Cheson BD, Bennett JM, Grever M, Kay N, Keating MJ, O'Brien S, et al. National Cancer Institute-sponsored Working Group guidelines for chronic lymphocytic leukemia: revised guidelines for diagnosis and treatment. *Blood* 1996;87:4990–7.
6. Cerhan JR, Fredericksen ZS, Wang AH, Habermann TM, Kay NE, Macon WR, et al. Design and validity of a clinic-based case-control study on the molecular epidemiology of lymphoma. *Int J Mol Epidemiol Genet* 2011;2:95–113.
7. Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* 2011;117:5019–32.
8. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106–20.
9. Kwan T, Benovoy D, Dias C, Gurd S, Provencher C, Beaulieu P, et al. Genome-wide analysis of transcript isoform variation in humans. *Nat Genet* 2008;40:225–31.
10. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–5.
11. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;15:1034–50.
12. Siepel A, Pollard K, Haussler D, editors. New methods for detecting lineage-specific selection. In *Proceedings of the 10th International Conference on Research in Computational Molecular Biology (RECOMB)* ed; Italy: Springer; 2006.
13. Bao L, Zhou M, Wu L, Lu L, Goldowitz D, Williams RW, et al. PolymiRTS database: linking polymorphisms in microRNA target sites with complex traits. *Nucleic Acids Res* 2007;35(Database issue):D51–4.
14. Zhou JX, Lee CH, Qi CF, Wang H, Naghashfar Z, Abbasi S, et al. IFN regulatory factor 8 regulates MDM2 in germinal center B cells. *J Immunol* 2009;183:3188–94.
15. Lee CH, Melchers M, Wang H, Torrey TA, Slota R, Qi CF, et al. Regulation of the germinal center gene program by interferon (IFN) regulatory factor 8/IFN consensus sequence-binding protein. *J Exp Med* 2006;203:63–72.
16. Consortium GP. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
17. Cerhan JR, Ansell SM, Fredericksen ZS, Kay NE, Liebow M, Call TG, et al. Genetic variation in 1253 immune and inflammation genes and risk of non-Hodgkin lymphoma. *Blood* 2007;110:4455–63.