

## Genetic Susceptibility Variants for Chronic Lymphocytic Leukemia

Susan L. Slager<sup>1</sup>, Lynn R. Goldin<sup>2</sup>, Sara S. Strom<sup>3</sup>, Mark C. Lanasa<sup>4</sup>, Logan G. Spector<sup>5</sup>, Laura Rassenti<sup>6</sup>, Jose F. Leis<sup>1</sup>, Nicola J. Camp<sup>7</sup>, Neil E. Kay<sup>1</sup>, Celine M. Vachon<sup>1</sup>, Martha Glenn<sup>7</sup>, J. Brice Weinberg<sup>4</sup>, Kari G. Rabe<sup>1</sup>, Julie M. Cunningham<sup>1</sup>, Sara J. Achenbach<sup>1</sup>, Curtis A. Hanson<sup>1</sup>, Gerald E. Marti<sup>8</sup>, Timothy G. Call<sup>1</sup>, Neil E. Caporaso<sup>2</sup>, and James R. Cerhan<sup>1</sup>

### Abstract

**Background:** There is strong and consistent evidence that a genetic component contributes to the etiology of chronic lymphocytic leukemia (CLL). A recent genome-wide association study of CLL identified seven genetic variants that increased the risk of CLL within a European population.

**Methods:** We evaluated the association of these variants, or variants in linkage disequilibrium with these variants, with CLL risk in an independent sample of 438 CLL cases and 328 controls.

**Results:** Of these seven single nucleotide polymorphisms (SNP), six had  $P$  trend < 0.05 and had estimated odds ratios (OR) that were strikingly comparable to those of the previous study. Associations were seen for rs9378805 [OR, 1.47; 95% confidence intervals (CI), 1.19-1.80;  $P$  trend = 0.0003] near *IRF4* and rs735665 near *GRAMD1B* (OR, 1.47; 95% CI, 1.14-1.89;  $P$  trend = 0.003). However, no associations ( $P > 0.05$ ) were found for rs11083846, nor were any found for any SNP in linkage disequilibrium with rs11083846.

**Conclusions:** Our results confirm the previous findings and further support the role of a genetic basis in the etiology of CLL; however, more research is needed to elucidate the causal SNP(s) and the potential manner in which these SNPs or linked SNPs function in CLL pathogenesis. *Cancer Epidemiol Biomarkers Prev*; 19(4); 1098-102. ©2010 AACR.

### Introduction

Chronic lymphocytic leukemia (CLL) is a malignant B-cell lymphoproliferative disorder and is most common in male Caucasians, with a lifetime risk of 1 in 200 (1). It has been known for decades that a genetic component affects CLL risk, with the risk on the order of 8-fold for individuals with relatives with CLL compared with the general population (2, 3). However, identification of susceptibility genes has remained elusive. Three genome-wide linkage studies have been conducted (4-6). Although regions of interest were found, no definitive susceptibility genes were identified. Furthermore, a number of candidate gene studies have been conducted (see review in ref. 7), but no consistent results were identified across these studies. Only recently, through a genome-wide association study of 517

CLL cases from the United Kingdom and 1,438 British 1958 Birth Cohort controls, researchers identified seven genetic variants that increased CLL risk (8). These findings were replicated through two internal validation samples of 520 cases and 891 controls (validation 1) and 504 cases and 786 controls (validation 2). We evaluated the association of these seven variants, or variants in linkage disequilibrium (LD) with these variants, with CLL risk in an independent sample of 438 non-Hispanic Caucasian CLL cases and 328 controls from the United States.

### Materials and Methods

#### Study population

All studies were approved by institutional review boards. Peripheral blood samples were obtained from

**Authors' Affiliations:** <sup>1</sup>Mayo Clinic College of Medicine, Rochester, Minnesota; <sup>2</sup>National Cancer Institute-Division of Cancer Epidemiology & Genetics, Bethesda, Maryland; <sup>3</sup>University of Texas M.D. Anderson Cancer Center, Houston, Texas; <sup>4</sup>Duke University Medical Center, Durham, North Carolina; <sup>5</sup>University of Minnesota, Minneapolis, Minnesota; <sup>6</sup>University of California San Diego, Moores Cancer Center, La Jolla, California; <sup>7</sup>University of Utah School of Medicine, Salt Lake City, Utah; and <sup>8</sup>U.S. Food and Drug Administration, Bethesda, Maryland

**Authorship:** S.L. Slager directed the overall study and wrote the manuscript; K.G. Rabe and S.J. Achenbach conducted the data analyses; J.M. Cunningham conducted the genotyping; L.R. Goldin, N.E. Caporaso, and G.E. Marti are the primary investigators of the NCI site for CLL family collection; S.S. Strom is the primary investigator of the M.D. Anderson site for CLL family collection; M.C. Lanasa and J.B. Weinberg are the primary investigators of the Duke University site for CLL family collection; L.G.

Spector is the primary investigator of the University of Minnesota/Minneapolis Veteran Administration Medical Center site for CLL family collection; L. Rassenti is the primary investigator of the UCSD site for CLL family collection; J.F. Leis, T.G. Call, N.E. Kay, C.A. Hanson, J.R. Cerhan, and C.M. Vachon contributed to CLL family recruitment at Mayo Clinic; N.J. Camp and M. Glenn is the primary investigator of the University of Utah study for CLL family recruitment; J.R. Cerhan is the primary investigator for the Mayo NHL/CLL case-control study; all authors contributed to study design, and reviewed and provided revisions to the manuscript.

**Corresponding Author:** Susan L. Slager, Mayo Clinic College of Medicine, 200 1st Street Southwest, Rochester, MN 55905. Phone: 507-284-5965; Fax: 507-284-9542. E-mail: slager@mayo.edu

doi: 10.1158/1055-9965.EPI-09-1217

©2010 American Association for Cancer Research.

two ongoing studies: the Genetic Epidemiology of CLL Consortium and the Mayo Clinic non-Hodgkin lymphoma/CLL study. The Genetic Epidemiology of CLL consortium is a collaboration of researchers from seven institutions with the overall aim of investigating the genetic basis of CLL through the collection of CLL families (i.e., families with two or more relatives with CLL). A total of 110 Caucasian CLL patients from 110 families were available at the time of genotyping. These families were collected from Duke University, Mayo Clinic, M.D. Anderson Cancer Center, National Cancer Institute, University of Minnesota/Minneapolis Veteran Administration Medical Center, University of California San Diego, and University of Utah. The second study is the Mayo Clinic non-Hodgkin lymphoma/CLL case-control study, which is an ongoing, clinic-based study conducted in Rochester, MN (9). Briefly, newly diagnosed non-Hodgkin lymphoma/CLL patients who were ages 20 y or older, HIV negative, and residents of the Midwest United States at the time of diagnosis were enrolled. Clinic-based controls were ascertained from patients visiting the General Internal Medicine clinic. Eligibility requirements included age 20 y or older and being a resident of Minnesota, Iowa, or Wisconsin; patients were excluded if they had prior diagnoses of lymphoma, leukemia, or HIV infection. From this study, genotype data were available from 328 CLL cases and 475 controls.

The diagnosis of all CLL cases across both studies were reviewed and confirmed by a hematopathologist, and classified according to the WHO criteria (10).

### Genotyping

This replication used genotype data from two larger ongoing genotyping projects from the Genetic Epidemiology of CLL Consortium and the Mayo Clinic case-control studies (Fig. 1). For the first project, we genotyped 438 CLL cases (110 familial CLL and 328 sporadic CLL) and 328 controls using the Affymetrix 6.0 SNP Array. Rigorous quality control measures were implemented including dropping single nucleotide polymorphisms (SNP) or individuals with call rates <95%, and evaluating for subject relatedness, evidence of population stratification, and sex discrepancies. Concordancy of genotype calls across duplicates was >99.7%. Four of the seven SNPs (rs17483466, rs13397985, rs9378805, and rs735665) identified by Di Bernardo et al. (8) were directly available on the Affymetrix 6.0 platform on 407 CLL cases and 296 controls who passed quality control criteria. From the second genotyping project, 129 sporadic CLL cases and 475 controls were genotyped on the ParAllele Immune and Inflammation SNP panel. Similar quality control measures as listed above were implemented in this project; for full details see ref. (9). One (rs872071) of the seven SNPs identified by Di Bernardo et al. (8) was available from this platform in all 129 CLL cases and 468 controls who were available at the time of genotyping. A total of 223 subjects and 1,974 SNPs overlapped these two projects, providing

additional assessment of genotype concordancy (>99.5% concordant).

### Statistical analysis

Tests for Hardy-Weinberg equilibrium were done using a Pearson goodness-of-fit test. We used logistic regression to estimate odds ratios (OR) and corresponding 95% confidence intervals (CI). A *P* trend was calculated assuming an ordinal genotypic relationship. We also evaluated whether the associations between CLL risk and each SNP varied by family history status. We used polytomous logistic regression with three outcome categories (familial CLL cases, sporadic CLL cases, and controls), and we tested for heterogeneity of the ORs across the two case groups.

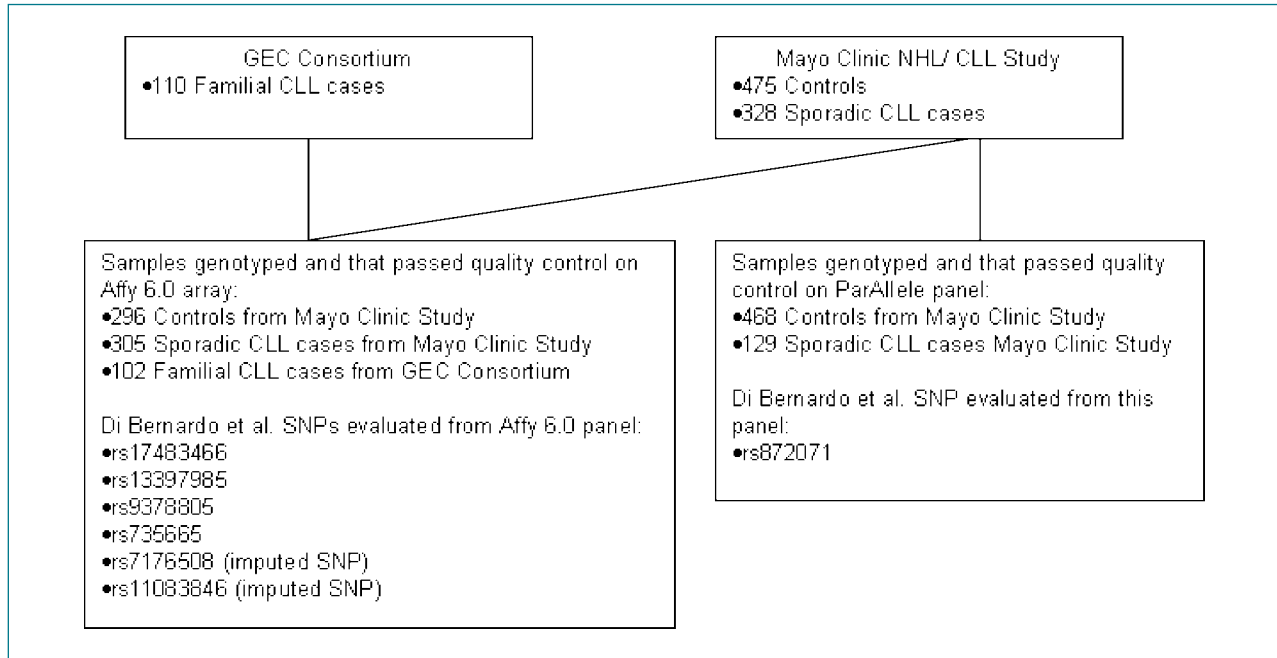
There were two SNPs (rs7176508 and rs11083846) that were not genotyped in either of the two genotyping platforms. For these, we reported statistical results for all neighboring SNPs that were in high LD ( $r^2 > 0.9$ ) with these SNPs based on HapMap, NCBI build 36. Pairwise LD measure ( $r^2$ ) was estimated using the program Haploview. Furthermore, we imputed genotypes for these two untyped SNPs using MACH 1.0.14 (11). The 60 unrelated HapMap CEU samples were used to obtain the phased chromosomes, and the expected genotype was estimated based on the posterior probability. We imputed genotypes for only the 407 cases and 296 controls who were genotyped on Affymetrix 6.0 platform because the regions around the two untyped SNPs had no genotype data available on the ParAllele Immune and Inflammation SNP panel.

### Results

Gender and age distribution were similar across the two genotyping projects. Among the cases, 64% were male, and the mean age at diagnosis was 61.4 years ( $\pm 11.1$ ). Among the controls, 57% were male and the mean age of recruitment was 61.8 years ( $\pm 12.5$ ).

Our associations for the seven SNPs identified by Di Bernardo et al. (8) and CLL risk are shown in Table 1. For all of the five typed SNPs, the observed *P* values were < 0.05. Our strongest finding was with rs9378805 SNP on chromosome 6p25.3, with a 47% increased risk of CLL (OR, 1.47; 95% CI, 1.19-1.80; *P* trend = 0.0003) with each allele. This effect size is very similar to the findings of Di Bernardo et al., who reported a 51% increased risk (95% CI, 1.38-1.65). For rs872071, which was in LD ( $r^2 = 0.74$ ) with rs9378805, we also observed an association (OR, 1.35; 95% CI, 1.02-1.78; *P* trend = 0.037), but the effect was slightly attenuated compared with the results of Di Bernardo et al. (OR, 1.54; 95% CI, 1.41-1.69). Our second strongest finding of the genotyped SNPs was rs735665 on chromosome 11q24.1 (OR, 1.47; 95% CI, 1.14-1.89; *P* trend = 0.003), followed by the two SNPs on chromosome 2q13 (rs17483466 and rs13397985). The magnitude of these associations was also comparable to that of Di Bernardo et al.

For the remaining two SNPs identified by Di Bernardo et al. located on 15q23 (rs7176508) and 19q13 (rs11083846), we did not have genotype data available. However, we



**Figure 1.** Schema of samples and genotype projects.

evaluated genotyped SNPs that were in high LD ( $r^2 > 0.9$ ) with rs7176508 and rs11083846. As shown in Table 1, the four SNPs in high LD with rs7176508 on chromosome 15q23 were associated ( $P$  trend  $< 0.001$ ) with effect sizes comparable to that of Di Bernardo et al., who reported an OR of 1.37 (95% CI, 1.26-1.50). We also imputed genotypes for rs7176508, with a resulting

imputed OR of 1.50 (95% CI, 1.19-1.90). The quality score for the imputed genotypes was 0.99, indicating that our imputed genotypes are accurate. We did not replicate the findings on chromosome 19 based on two neighboring typed SNPs ( $P$  trend  $> 0.05$ ) nor based on imputing genotypes for rs11083846. The quality score for the imputed genotypes was 0.96.

**Table 1.** ORs and 95% CIs for observed and imputed SNPs with risk of CLL

SNP	Chr	Nearest gene	Position	Genotyping platform	Risk allele	Case		Control		OR* (95% CI)	$P_{\text{trend}}$
						MAF	<i>n</i>	MAF	<i>n</i>		
<b>rs17483466</b>	<b>2q13</b>	<b>ACOXL, BCL2L11</b>	<b>111,513,929</b>	Affy 6.0	G	0.26	405	0.2	294	1.43 (1.11-1.85)	6.30E-03
<b>rs13397985</b>	<b>2q13</b>	<b>SP140, SP110</b>	<b>230,799,467</b>	Affy 6.0	C	0.23	401	0.18	293	1.36 (1.04-1.79)	2.40E-02
<b>rs872071<sup>†</sup></b>	<b>6p25</b>	<b>IRF4</b>	<b>356,064</b>	ParAllele	G	0.59	129	0.52	468	1.35 (1.02-1.78)	3.70E-02
<b>rs9378805</b>	<b>6p25</b>	<b>IRF4</b>	<b>362,727</b>	Affy 6.0	G	0.57	407	0.47	296	1.47 (1.19-1.80)	3.00E-04
<b>rs735665</b>	<b>11q24</b>	<b>GRAMD1B</b>	<b>122,866,607</b>	Affy 6.0	T	0.28	406	0.21	293	1.47 (1.14-1.89)	2.80E-03
rs11072110	15q23	—	67,774,936	Affy 6.0	T	0.48	407	0.38	296	1.54 (1.24-1.92)	1.00E-04
rs10220831	15q23	—	67,779,168	Affy 6.0	G	0.47	406	0.36	296	1.56 (1.25-1.95)	9.00E-05
rs35707742	15q23	—	67,791,406	Affy 6.0	G	0.47	401	0.37	291	1.52 (1.22-1.90)	2.10E-04
rs4777184	15q23	—	67,794,377	Affy 6.0	T	0.47	407	0.37	296	1.52 (1.22-1.89)	2.10E-04
<b>rs7176508</b>	<b>15q23</b>	—	<b>67,806,044</b>	Imputed	A	0.46	407	0.37	296	1.50 (1.19-1.90)	6.20E-04
<b>rs11083846</b>	<b>19q13</b>	<b>PRKD2, STRN4</b>	<b>51,899,494</b>	Imputed	A	0.29	407	0.27	296	1.11 (0.86-1.43)	0.44
rs11668987	19q13	<i>PRKD2, STRN4</i>	51,919,661	Affy 6.0	G	0.3	407	0.28	296	1.11 (0.89-1.38)	0.36
rs4802322	19q13	<i>PRKD2, STRN4</i>	51,934,832	Affy 6.0	G	0.29	407	0.26	296	1.12 (0.89-1.42)	0.33

NOTE: Values in boldface have been reported in Di Bernardo et al.

Abbreviations: Chr, chromosome; MAF, minor allele frequency.

\*OR is based on per copy of risk allele.

<sup>†</sup>At the time of genotyping on this platform, only 129 CLL cases and 468 controls were available to genotype.

**Table 2.** ORs and 95% CIs for risk of CLL by familial CLL status for the typed SNPs on Affy 6.0 platform

SNP	Chr	Familial CLL, <i>n</i> = 102, OR* (95% CI)	Sporadic CLL, <i>n</i> = 305, OR* (95% CI)	<i>P</i> <sub>heterogeneity</sub>
rs17483466	2	1.51 (1.04-2.18)	1.40 (1.07-1.84)	0.69
rs13397985	2	1.34 (0.90-1.99)	1.37 (1.03-1.83)	0.91
rs872071	6	ND	ND	ND
rs9378805	6	1.80 (1.31-2.48)	1.37 (1.10-1.71)	0.09
rs735665	11	1.77 (1.24-2.53)	1.38 (1.05-1.80)	0.15
rs11072110	15	1.96 (1.41-2.71)	1.43 (1.13-1.80)	0.05
rs10220831	15	1.96 (1.40-2.73)	1.45 (1.14-1.84)	0.07
rs35707742	15	1.94 (1.39-2.69)	1.40 (1.11-1.77)	0.05
rs4777184	15	1.90 (1.37-2.65)	1.41 (1.11-1.78)	0.07
rs11668987	19	1.11 (0.79-1.52)	1.11 (0.88-1.40)	0.95
rs4802322	19	1.06 (0.74-1.50)	1.15 (0.89-1.47)	0.64

Abbreviations: Chr, chromosome; ND, not done.

\*ORs based on per copy of risk allele.

Because 102 of the CLL cases had a family history of CLL, we evaluated CLL risk by family history status for the genotyped SNPs. Risk estimates were generally higher among the familial CLL cases compared with that of the sporadic CLL cases (Table 2), but none had  $P < 0.05$ .

## Discussion

We found evidence for six out of the seven previously implicated SNPs for CLL risk in an independent sample of Caucasians from the United States. Even after using a Bonferroni correction of six independent regions (i.e., correcting for six independent tests), four SNPs still remained significant in our sample. There was suggestive evidence that the associations of the SNPs differed by family history status of CLL. A next step would be to evaluate whether these variants cosegregate with relatives who have CLL.

We were unable to replicate the association between CLL risk and rs11083846 located on 19q13. Limited statistical power may be an explanation. Assuming a 5% type I error rate and a 1.35 effect size (reported in Di Bernardo et al.), we have between 67% and 74% power to find this effect size with allele frequency between 0.22 and 0.28. Alternatively, this could be related to inherent differences in the populations under study. Our allele frequency of this SNP was greater in the controls compared with that of Di Bernardo et al., resulting in less allele frequency difference between cases and controls.

As reported in Di Bernardo et al., three of the replicated SNPs were located within genes, including rs872071 within the 3' untranslated region of *IRF4*, rs17483466 located in intron 10 of *ACOXL*, and rs13397985 located in intron 1 of *SPI40*. The SNP rs9378805 is ~10 kb from the 3' end of *IRF4*. *IRF4* is of great interest because of its role in lymphocyte development and previous associations with other lymphomas (12-15).

The strength of this study includes the well-characterized CLL cases and controls, as well as a number of CLL cases that are fairly comparable with each of the three phases of Di Bernardo et al. This study has several limitations. First, genotyped data were pulled from two ongoing separate genotype studies. As a result, not all subjects were genotyped across the five typed SNPs. The effect of this, however, was minor because even with fewer subjects genotyped, we were still able to identify all the SNPs but one with comparable effect sizes as those previously reported, yet with reduced statistical power. Second, two previously implicated SNPs were not typed in our study. To overcome this limitation, we imputed genotypes at these SNPs; we also evaluated the associations of neighboring genotyped SNPs in high LD with the implicated SNPs. Both the observed genotype data at neighboring SNPs and the imputed data at the previously implicated SNPs were consistent with each other for each of the two regions on chromosomes 15 and 19. Furthermore, our imputation results had high accuracy. Thus, although we did not have actual data for these SNPs, surrogate information was available and reliable.

In conclusion, this study confirms the previously reported associations of CLL risk and SNPs located on chromosomes 2q13, 2q37.1, 6p25.3, 11q24.1, and 15q23. Candidate genes have been identified in or near the regions of interest, so further work is needed to identify the causal SNPs, as well as the biological mechanisms by which CLL risk is increased.

## Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

## Grant Support

NIH grants CA118444 and CA92153; Intramural Research Program of the NIH, National Cancer Institute; and CLL Research Consortium. Data collection in Utah was made possible by the Utah Population

Database and the Utah Cancer Registry. Partial support for all data in the Utah Population Database was provided by the University of Utah Huntsman Cancer Institute. The Utah Cancer Registry is funded by contract N01-PC-35141 from the NCI's SEER program with additional support from the Utah State Department of Health and the University of Utah.

## References

1. Horner M, Ries L, Krapcho M, et al. SEER cancer statistics review 1975-2006. Bethesda (MD): NCI; 2009.
2. Goldin LR, Pfeiffer RM, Li X, Hemminki K. Familial risk of lymphoproliferative tumors in families of patients with chronic lymphocytic leukemia: results from the Swedish Family-Cancer Database. *Blood* 2004;104:1850-4.
3. Goldin L, Bjorkholm M, Kristinsson S, Turesson I, Landgren O. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. *Haematologica* 2009;94:606-9.
4. Goldin LR, Ishibe N, Sgambati M, et al. A genome scan of 18 families with chronic lymphocytic leukaemia. *Br J Haematol* 2003;121:866-73.
5. Sellick GS, Webb EL, Allinson R, et al. A high-density SNP genome-wide linkage scan for chronic lymphocytic leukemia-susceptibility loci. *Am J Hum Genet* 2005;77:420-9.
6. Sellick GS, Goldin LR, Wild RW, et al. A high-density SNP genome-wide linkage search of 206 families identifies susceptibility loci for chronic lymphocytic leukemia. *Blood* 2007;110:3326-33.
7. Slager SL, Kay NE, Fredericksen ZS, et al. Susceptibility genes and B-chronic lymphocytic leukaemia. *Br J Haematol* 2007;139:762-71.
8. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat Genet* 2008;40:1204-10.
9. Cerhan JR, Ansell SM, Fredericksen ZS, et al. Genetic variation in 1253 immune and inflammation genes and risk of non-Hodgkin lymphoma. *Blood* 2007;110:4455-63.
10. Jaffe E, Harris N, Stein H, Vardiman J. World Health Organization classification of tumours pathology and genetics, tumours of hematopoietic and lymphoid tissues. Lyon: IARC Press; 2001.
11. Li Y, Ding J, Abecasis GR. Mach 1.0: rapid haplotype reconstruction and missing genotype inference. *Am J Hum Genet* 2006;79:S2290.
12. Feldman AL, Law M, Remstein ED, et al. Recurrent translocations involving the IRF4 oncogene locus in peripheral T-cell lymphomas. *Leukemia* 2009;23:574-80.
13. Heintel D, Zojer N, Schreder M, et al. Expression of MUM1/IRF4 mRNA as a prognostic marker in patients with multiple myeloma. *Leukemia* 2008;22:441-5.
14. Valsami S, Pappa V, Rontogianni D, et al. A clinicopathological study of B-cell differentiation markers and transcription factors in classical Hodgkin's lymphoma: a potential prognostic role of MUM1/IRF4. *Haematologica* 2007;92:1343-50.
15. Wang SS, Purdue MP, Cerhan JR, et al. Common gene variants in the tumor necrosis factor (TNF) and TNF receptor superfamilies and NF- $\kappa$ B transcription factors and non-Hodgkin lymphoma risk. *PLoS One* 2009;4:e5360.

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 12/03/2009; revised 01/13/2010; accepted 02/04/2010; published OnlineFirst 03/23/2010.