Brief report

Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia

*Susan L. Slager,1 *Christine F. Skibola,2 Maria Chiara Di Bernardo,3 Lucia Conde,3 Peter Broderick,3 Shannon K. McDonnell,1 Lynn R. Goldin,4 Naomi Croft,3 Amy Holroyd,9 Shelley Harris,3 Jacques Riby,2 Daniel J. Serie,1 Neil E. Kay,1 Timothy G. Call,1 Paige M. Bracci,5 Eran Halperin,6 Mark C. Lanasa,7 Julie M. Cunningham,1 Jose F. Leis,1 Vicki A. Morrison,8,9 Logan G. Spector,9 Celine M. Vachon,1 Tait D. Shanafelt,1 Sara S. Strom,10 Nicola J. Camp,11 J. Brice Weinberg,7,12 Estella Matutes,13 Neil E. Caporaso,4 Rachel Wade,14 Martin J. S. Dyer,15 Claire Dearden,13 James R. Cerhan,1 Daniel Catovsky,13 and Richard S. Houlston3

1Mayo Clinic College of Medicine, Rochester, MN; 2Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, CA; 3Molecular and Population Genetics, Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, United Kingdom; 4Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 5Department of Epidemiology and Biostatistics, School of Medicine, University of California, San Francisco, San Francisco, CA; 6Blavatnik School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel; 7Duke University Medical Center, Durham, NC; 8Minneapolis Veterans Administration Medical Center, Minneapolis, MN; 9University of Minnesota, Minneapolis, MN; 10The MD Anderson Cancer Center, Houston, TX; 11University of Utah School of Medicine, Salt Lake City, UT; 12Durham Veterans Administration Medical Center, Durham, NC; 13Section of Haemato-Oncology, Division of Pathology, Institute of Cancer Research, Sutton, United Kingdom; 14Clinical Trial Service Unit, University of Oxford, Oxford, United Kingdom; and 15Department of Cancer Studies and Molecular Medicine, Leicester University, Leicester, United Kingdom

We performed a meta-analysis of 3 genome-wide association studies to identify additional common variants influencing chronic lymphocytic leukemia (CLL) risk. The discovery phase was composed of genome-wide association study data from 1121 cases and 3745 controls. Replication analysis was performed in 861 cases and 2033 controls. We identified a novel CLL risk locus at 6p21.33 (rs210142; intronic to the BAK1 gene, BCL2 antagonist killer 1; \(^{*} P = 9.47 \times 10^{-18}\)). A strong relationship between risk genotype and reduced BAK1 expression was shown in lymphoblastoid cell lines. This finding provides additional support for polygenic inheritance to CLL and provides further insight into the biologic basis of disease development. (Blood. 2012;120(4):843-846)

Introduction

Chronic lymphocytic leukemia (CLL) is the most common form of lymphoid malignancy in Western countries.1 Although CLL shows a strong familial risk,2 the genetic basis of inherited predisposition to CLL is largely unknown. Recent genome-wide association studies (GWAS) of CLL have provided evidence that the coinheritance of multiple low-risk variants located on chromosomes 2q37.1, 2q37.3, 2q13, 6p21.3, 6p25.3, 8q24.21, 11q24.1, 15q21.3, 15q23, 15q25.2, 16q24.1, and 19q13.32 contributes to the heritability of CLL.3-6 The statistical power of individual GWAS has been limited by the modest effect sizes of individual genetic variants, the need to establish stringent statistical significance thresholds, and financial constraints on the number of variants that can be followed up. Meta-analysis of existing GWAS data therefore offers the opportunity to discover additional CLL susceptibility loci.

In this study, we conducted a meta-analysis of GWAS data, followed by validation in an independent case-control series, enabling us to identify a novel susceptibility locus for CLL at 6p21.33.

Methods

Participants

All data collection from study participants were approved by the respective institutional review boards, and all participants provided written informed consent in accordance with the Declaration of Helsinki. For all cases, the diagnosis of CLL had been pathologically confirmed in accordance with the World Health Organization guidelines.7

Discovery datasets

The discovery phase was composed of 3 previously described GWAS conducted in the United Kingdom (UK-GWAS)3,5 with 503 cases and 2699 controls, in the San Francisco Bay Area (SF-GWAS)8 with 211 cases and 750 controls, and in the Genetic Epidemiology of CLL (GEC) consortium (GEC-GWAS)3 with 407 cases and 398 controls (supplemental Methods, available on the Blood Web site; see the Supplemental Materials link at the top of the online article).

Replication series

The replication series was composed of 861 CLL cases (565 men; mean age at diagnosis, 61.9 years), ascertained through United Kingdom hematology clinics. Controls were 2033 healthy persons recruited to the National Cancer Research Network genetic epidemiologic studies, the National Study of Colorectal Cancer,9 the Genetic Lung Cancer Predisposition Study,10 and the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry. These controls were the spouses or unrelated friends of persons with malignancies. Both cases and controls were British residents and of European ancestry. Genotyping was conducted using competitive allele-specific PCR KASPar chemistry (KBiosciences).


*S.L.S. and C.F.S. contributed equally to this study as co–first authors.

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.
To confirm genotyping accuracy, duplicate samples were genotyped, along with direct sequencing of subsets of samples. For all single nucleotide polymorphisms (SNPs), more than 99% concordant results were obtained.

**Statistical analysis**

Main analyses used R Version 2.5, Stata Version 11, and PLINK Version 1.07 software. The association between each SNP and CLL risk was assessed by the Cochran-Armitage trend test. Odds ratios (ORs) and 95% confidence intervals were calculated by unconditional logistic regression. Meta-analysis was conducted under a fixed-effects model. Cochran Q statistic (to test for heterogeneity) and the F statistic (to quantify the proportion of the total variation because of heterogeneity) were calculated. Associations by sex, age, and clinic-pathologic phenotypes were examined by logistic regression in case-only analyses. The familial risk attributable to SNPs was calculated as previously described. Linkage disequilibrium metrics were based on Data Release 27/phase 3 (February 2009) on NCBI B36 assembly.

**Results and discussion**

The combined GWASs provided genotype data on 1121 cases and 3745 controls; imputation based on data from the HapMap project allowed association testing for more than 1,500,000 SNPs. We verified the previously identified risk loci (all \( P < 0.05 \)) in the meta-analysis (supplemental Table 1) and further identified 15 SNPs that mapped to 6 novel loci.

To validate these findings, we genotyped the top 10 informative SNPs in an additional series of 861 CLL cases and 2033 controls (Table 1; supplemental Table 2).

The 2 SNPs localizing to 6p21.33, rs210134 and rs210142, showed convincing evidence of an association in the replication series \( \text{OR} = 1.35, \ P = 4.87 \times 10^{-6} \) and \( \text{OR} = 1.47, \ P = 2.41 \times 10^{-8} \), respectively; Table 1). In the combined analysis, both rs210142 \( \text{OR} = 1.37, \ P = 9.47 \times 10^{-16} \) and rs210134 \( \text{OR} = 1.40, \ P = 1.03 \times 10^{-12} \) provided evidence for an association with CLL at genome-wide significance (ie, \( P < 5.0 \times 10^{-8} \); Table 1). None of the other SNPs replicated.

rs210142 maps to chromosome 6p21.33, within intron 1 of the BAK1 gene (BCL2 antagonist killer 1; MIM: 600516), and rs210134 maps 100 kb telomeric to BAK1 (supplemental Figure 1). Both SNPs map to the same region of linkage disequilibrium and are highly correlated \( (r^2 = 0.8, D^2 = 0.9) \). BAK1 promotes apoptosis by binding to and antagonizing the apoptosis repressor activity of BCL2 and other antiapoptotic proteins. Somatic rearrangements of the immunoglobulin heavy chain locus of BCL2 that result in constitutive BCL2 overexpression are found in both CLL and follicular lymphomas. The expression of BAK1 is essential for the maintenance of B-cell homeostasis; in mice that are conditionally deficient in BAK1, there is an accumulation of immature and mature follicular B cells with defective cell cycling in response to B-cell receptor stimulation.

To explore whether the 6p21.33 association reflects cis-acting regulatory effects on BAK1, we analyzed publicly available mRNA expression (Figure 1). A strong relationship between rs210134 risk genotype and reduced BAK1 expression was shown in both datasets \( (P = 7.8 \times 10^{-5} \text{ and } 0.389, \text{combined } P = 4.4 \times 10^{-5} \); Figure 1). This result suggests a biologically plausible mechanism by which reduction of BAK1 expression alleviates repression of antiapoptotic proteins, thereby inhibiting apoptosis and hence contributing to B-cell neoplasia. BAK1 does not, however, directly interact with BCL2, but its interaction is dependent on the allelic form of BCL2 present in cells; thus, the effects of BAK1 are context dependent.

---

**Table 1. Summary results for rs210134 and rs210142 SNPs associated with CLL risk**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position, bp</th>
<th>Risk allele</th>
<th>RAf</th>
<th>( \text{OR} ) (95% CI) P ( \text{P}_{\text{meta}} )</th>
<th>( \text{I}^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs210134</td>
<td>33,648,186</td>
<td>G</td>
<td>0.68</td>
<td>1.31 (1.13-1.51) 2.95 \times 10^{-4} 0.69</td>
<td>57.06</td>
</tr>
<tr>
<td>rs210142</td>
<td>33,644,815</td>
<td>C</td>
<td>0.70</td>
<td>1.38 (1.17-1.60) 6.82 \times 10^{-5} 0.70</td>
<td>58.10</td>
</tr>
</tbody>
</table>

RAf indicates risk allele frequencies (based on control data).
rs210134 genotype influences overall survival or progression-free trial patients (supplemental Methods), we found no evidence that (supplemental Table 3). Furthermore, using data from the CLL4 evidence of a relationship with age, sex, or (supplemental Methods). rs210142 and rs210134 showed no assessed the relationship between age, sex, MuTHER project17 as of December 15, 2008 (GENEVAR project). The G allele is data on 156 lymphoblastoid cell lines derived from healthy female twins as part of the correlation and publicly available Sentrix Human-6 Expression BeadChips (Illumina) gene expression and the count of the G allele of rs210134 using Spearman rank lymphoblastoid cell lines.

Figure 1. Lower BAK1 expression associated with the G allele of rs210134 in lymphoblastoid cell lines. We assessed the correlation between normalized BAK1 gene expression and the count of the G allele of rs210134 using Spearman rank correlation and publicly available Sentrix Human-6 Expression BeadChips (Illumina) data on 156 lymphoblastoid cell lines derived from healthy female twins as part of the MuTHER project17 as of December 15, 2008 (GENEVAR project). The G allele is associated with the elevated risk of CLL.

CLL shows male predominance and can be classified on the basis of the presence or absence of somatic hypermutations of IGVH genes, with mutated CLL having a better prognosis.19 We assessed the relationship between age, sex, IGVH mutation, and rs210142 and rs210134 genotypes by case-only logistic regression (supplemental Methods). rs210142 and rs210134 showed no evidence of a relationship with age, sex, or IGVH mutation status (supplemental Table 3). Furthermore, using data from the CLL4 trial patients (supplemental Methods), we found no evidence that rs210134 genotype influences overall survival or progression-free survival (hazard ratio = 1.11, P = .41 and hazard ratio = 1.12, P = .23; respectively).

We investigated the combined effect of the 6p21.33 variation and the previously identified risk variants on CLL risk. No evidence of interaction between any of the loci (P > .05) was observed, compatible with each locus having an independent role in defining risk. Whereas the risks of CLL associated with the 6p21.33 and other variants are individually modest, the carrier frequencies of the risk alleles are high in those persons of European ancestry; therefore, the loci contribute significantly to the development of CLL.

Collectively, the currently identified susceptibility loci account for approximately 16% of the familial CLL risk. Previous genetic linkage studies have failed to provide evidence that rare, high-penetrance genes contribute significantly to the familial risk. Our study had moderate power (~50%) to detect variants, such as 6p21.33, indicating that additional common variants with similar or smaller effects might be identified with additional GWAS data.

By pooling GWAS data and conducting replication analyses, we have identified a novel CLL susceptibility locus. Although additional analyses are required to determine the functional consequences of 6p21.33 variation, the findings further highlight the importance of genetic variation in B-cell developmental pathway as a biologic basis to CLL pathogenesis. The frequency of the rs210142 risk allele is substantially higher in Europeans than in other ancestry groups consistent with adaptive selection. Hence, it will also be intriguing to explore how our findings translate to non-European populations, some of which are typified by a significantly lower prevalence of CLL.20

Acknowledgments

The authors thank the study participants and the study coordinators for work in recruitment. This study made use of genotyping data from the 1958 Birth Cohort and NBS samples, kindly made available by the Wellcome Trust Case Control Consortium 2. A full list of the investigators who contributed to the generation of the data is available at http://www.wtccc.org.uk/. NHS funding for the Royal Marsden Biomedical Research Center is acknowledged.


In the GEC Consortium, the work was supported in part by the National Institutes of Health (NIH; grants CA118444 and CA148690, S.L.S.; and grant CA92153, J.R.C.); the Intramural Research Program of the NIH, National Cancer Institute (NCI); and the Veterans Affairs Research Service. 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 the National Cancer Institute Surveillance Epidemiology and End Results program (contract HHSN261201000026C) with additional support from the Utah State Department of Health and the University of Utah. Sample collection at Duke University was supported by a Leukemia & Lymphoma Society Career Development Award (M.C.L.), the Bernstein Family Fund for Leukemia and Lymphoma Research, and the NIH (1K08CA134919, M.C.L.). The SF-GWAS was supported by the NIH (grants CA122663, CA154643-01A1, and CA104682, to C.F.S.) and the NCI NIH (grants CA45614 and CA89745, to P.M.B.). E.H. is a faculty fellow of the Edmond J. Safra Bioinformatics program at Tel-Aviv University. In the United Kingdom, Leukemia Lymphoma Research Fund provided principal funding for the study (LRF05001 and 06002). Additional funding was provided by Cancer Research UK (C1298/A8362 supported by the Bobby Moore Fund), and the Arbib Fund. M.C.D.B. was supported by the NIH (CA148690).
Authorship

Contribution: S.L.S., C.F.S., and R.S.H. designed the study, obtained financial support, and drafted the manuscript; M.C.D.B., L.C., E.H., R.W., D.J.S., and S.K.M. conducted statistical and bioinformatic analyses; P.B. performed laboratory management and oversaw genotyping of United Kingdom cases and controls; N.C., A.H., and S.H. performed genotyping of United Kingdom and oversaw recruitment of persons with CLL and sample acquisition within the United Kingdom; P.M.B. developed protocols for recruitment of persons from Duke University; S.L.S., C.D., and E.M. performed ascertainment and oversampling samples; R.S.H. and D.C. developed protocols for recruitment of persons from the MD Anderson Cancer Center; N.J.C. oversaw recruitment of persons from Utah; L.G.S. and V.A.M. oversaw recruitment of persons from University of Minnesota/Minneapolis Veterans Administration Medical Center; and all authors contributed to the final manuscript.

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

Correspondence: Susan L. Slager, Mayo Clinic College of Medicine, 200 1st St SW Rochester, MN 55905; e-mail: slager@mayo.edu; Christine F. Skibola, Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, B84 Hildebrand Hall MC 7356, Berkeley, CA; e-mail: chrisfs@berkeley.edu; and Richard S. Houlston, Molecular and Population Genetics, Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, SM2 5NG, United Kingdom; e-mail: richard.houlston@icr.ac.uk.

References

11. Leu JJ, George DL. Hepatic IGFBP1 is a prosurvival factor that binds to BAK, protects the liver from apoptosis, and antagonizes the proapoptotic actions of p53 at mitochondria. Genes Dev. 2007;21(23):3095-3109.
Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia