Variants in the ATM-BRCA2-CHEK2 axis predispose to chronic lymphocytic leukemia

Matthew F. Rudd, Gabrielle S. Sellick, Emily L. Webb, Daniel Catovsky, and Richard S. Houlston

We conducted a large-scale association study to identify low-penetrance susceptibility alleles for chronic lymphocytic leukemia (CLL), analyzing 992 patients and 2707 healthy controls. To increase the likelihood of identifying disease-causing alleles we genotyped 1467 coding nonsynonymous single nucleotide polymorphisms (nsSNPs) in 865 candidate cancer genes, biasing nsSNP selection toward those predicted to be deleterious. Preeminent associations were identified in SNPs mapping to genes pivotal in the DNA damage-response and cell-cycle pathways, including ATM F858L (odds ratio [OR] = 2.28, P < .0001) and P1054R (OR = 1.68, P = .0006), CHEK2 I157T (OR = 1.45, P = .0032), and BUB1B Q349R (OR = 1.42, P = .0038). Our findings implicate variants in the ATM-BRCA2-CHEK2 DNA damage-response axis with risk of CLL. (Blood. 2006;108:638-644)

Introduction

Chronic lymphocytic leukemia (CLL) is the most common form of leukemia and is 1 of a number of B-cell lymphoproliferative disorders (B-cell LPDs) that include Hodgkin lymphoma (HL) and non-Hodgkin lymphoma (NHL). Inherited predisposition to CLL and other B-cell LPDs is well documented, with epidemiologic studies showing that the risk of CLL in first-degree relatives of patients with CLL is elevated 7-fold.1,2 Furthermore, studies have demonstrated that familial associations exist between different types of B-cell LPDs, with risks of HL and NHL increased 2-fold in relatives of patients with CLL.3 Whereas part of the familial risk could be due to high-penetrance mutations in as-yet-undiﬁed genes, a polygenic model based on low-penetrance alleles provides an alternative explanation. Such a hypothesis is supported by the recent observation that monoclonal B-cell lymphocytosis with an identical phenotype to indolent CLL can be detected in a high proportion of healthy members of CLL families.3,4

Alleles conferring small relative risks are difﬁcult, if not impossible, to identify through classic genome-wide linkage scans.5 The search for low-penetrance disease alleles has therefore centered on association studies based on comparing the frequency of polymorphic genotypes in patients and control subjects. The spectrum of mutations in Mendelian disease genes, coupled with issues of statistical power, provides a compelling rationale for the application of a sequence-based approach targeting nonsynonymous single nucleotide polymorphisms (nsSNPs) rather than reliance on a map of anonymous haplotypes.6 We sought to identify novel low-penetrance susceptibility alleles for CLL by genotyping nsSNPs across 865 genes with relevance to cancer biology, biasing selection of nsSNPs toward those likely to have deleterious consequences. Genotyping 992 patients with CLL and 2707 healthy controls from the United Kingdom population across 1467 nsSNPs provided strong evidence for association between genes in the DNA damage-response and cell-cycle pathways and risk of CLL.

Patients, materials, and methods

Patients and control subjects

Patients with adult CLL (992 total; 688 men, 304 women; mean age at diagnosis, 61 years; SD ± 11.4) ascertained through the Royal Marsden Hospital National Health Service Trust (RMHNHT) Haemato-Oncology Unit were included in the study. The RMHNHT serves as a tertiary referral center and patient selection was not signiﬁcantly biased to any speciﬁc geographic region within the United Kingdom. The diagnosis of B-cell CLL in patients was established using standard clinico-pathologic and immunologic criteria in accordance with current World Health Organization classiﬁcation guidelines.7 A total of 2707 healthy individuals were recruited as part of the National Cancer Research Network Trial (1999-2002), the Royal Marsden Hospital Trust/Institute of Cancer Research Family History and DNA Registry (1999-2004), or the National Study of Colorectal Cancer Genetics Trial (2004), all established within the United Kingdom. Control subjects (836 men, 1871 women; mean age, 59 years; SD ± 10.9) were the spouses of patients with nonhematologic malignancies. None of the controls had a personal history of malignancy. All patients and control subjects were white and British, and there were no obvious differences in the demography of patients and control subjects in terms of place of residence within the United Kingdom. Blood samples were obtained with informed consent and ethics review board approval in accordance with the tenets of the Declaration of Helsinki. DNA was extracted from samples using conventional methodologies and quantiﬁed using PicoGreen (Invitrogen, Carlsbad, CA).

From the Sections of Cancer Genetics and Haemato-Oncology, Institute of Cancer Research, Sutton, Surrey, United Kingdom.


M.F.R. and G.S.S. contributed equally to this study.

© 2006 by The American Society of Hematology
Selection of candidate genes and SNPs

We have previously established a publicly accessible PICS (Predicted Impact of Coding SNPs) database of potentially functional nsSNPs in genes with relevance to cancer biology. Briefly, candidate genes were identified by interrogating the Gene Ontology Consortium database, Kyoto Encyclopedia of Genes and Genomes database, Iobion’ Interaction Explorer PathwayAssist Program, National Center for Biotechnology Information (NCBI) Entrez Gene database, and the CancerGene database. Both keyword and gene-pathway specific queries were performed using the following categories: catalytic activity; cellular processes, growth, and death; development; enzyme regulator activity; folding, sorting, and degradation; ligand-receptor interaction; nucleotide metabolism; physiologic processes; regulation of biologic processes; replication and repair; signal transduction and signal transducer activity; transcription and transcription regulator activity; translation and translation regulator activity; and transporter activity. A total of 9537 validated nsSNPs with minor allele frequency (MAF) data were identified within 21 506 LocusLink annotated genes in NCBI dbSNP Build 123. Filtering this list and linking it to 7080 candidate cancer genes yielded 3666 validated nsSNPs with MAF of 0.01 or more in white populations. The functional impact of each nsSNP was predicted using the in silico computational tools PolyPhen and SIFT. SNP genotyping and data manipulation

Genotyping of samples was performed using customized Illumina Sentrix Bead Arrays (Illumina, San Diego, CA) according to the manufacturer’s protocols. DNA samples with GenCall (Illumina) scores lower than 0.25 at any locus were considered “no calls.” A DNA sample was deemed to have failed if it generated genotypes at fewer than 95% of loci. A SNP was deemed to have failed if fewer than 95% of DNA samples generated a genotype at the locus. Conversion of genotypic data into formats suitable for processing was performed using in-house Perl scripts (available upon request). Conventional statistical manipulations were undertaken in STATA (version 8; http://www.stata.com), S-Plus (version 7; http://www.insightful.com) or R (version 2.0.0; http://www.r-project.org). Population stratification

Genotypic frequencies in control subjects for each SNP were tested for population stratification or R (version 2.0.0; http://www.r-project.org). Processing was performed using in-house Perl scripts (available upon request). Genotyping errors, and were thus removed from further analyses. To detect and control for possible population stratification, we employed the genomic control approach using all SNPs to estimate the stratification parameter λ and its associated 95% confidence interval (CI). The possibility of sex differences as a source of population structure in our controls was evaluated by χ² test or Fisher exact test.

Risk of CLL associated with nsSNPs

The most efficient test of association depends on the true mode of inheritance of alleles. Since this is not known, we based our analyses on the difference between allelic frequencies in patients and control subjects using the χ² test with 1 degree of freedom or Fisher exact test if an expected cell count was less than 5. SNPs that violate the HWE in the control population can indicate selection bias or genotyping errors, and were thus removed from further analyses. To detect and control for possible population stratification, we employed the genomic control approach using all SNPs to estimate the stratification parameter λ and its associated 95% confidence interval (CI). The possibility of sex differences as a source of population structure in our controls was evaluated by χ² test or Fisher exact test.

Multiple testing

Correction for multiple testing in association studies using simple adjustment approaches such as the Bonferroni correction are known to be conservative due to the assumption of independence between tests, which can lead to type II errors. To control the type II error rate, we adopted an empirical Monte Carlo simulation approach based on 10 000 permutations, which takes into account the fact that tests may be correlated due to the presence of linkage disequilibrium (LD) throughout the genome. At each iteration, patient and control subject labels are permuted at random and maximum test statistics TAmax, TDmax, and TRmax are determined. For each of these statistics (allelic, dominant, or recessive models), significance levels of the observed statistics from the original data are then estimated by the proportion of permutation samples with TAmax, TDmax, and TRmax larger than that in the observed data. Although this approach adjusts for multiple testing for each of the 3 statistics separately, the consequent increase in false-positive rate is expected to be small due to the strong dependence between tests.

Assessment of linkage disequilibrium between SNPs

To assess the level of LD between SNPs, we calculated the pairwise LD measure D’ between consecutive pairs of markers throughout the genome using the expectation-maximization algorithm to estimate 2-locus haplotype frequencies. We chose to use the measure D’ as it is less sensitive to small minor allele frequencies than other measures such as r². This information was used to investigate the relationship between haplotypes and disease status. Specifically, haplotypes were reconstructed using a Markov chain Monte Carlo method, and their frequencies in patient and control samples compared by permutation testing, using the PHASE program (http://www.stat.washington.edu/stephens/software.html).

Covariates and interactions

Information on a number of covariates was available for the patients, including sex, family history of CLL, and age at diagnosis. The test statistic TA was computed for all subgroups, together with ORs and their associated 95% CIs. Under certain conditions, a 2-stage process incorporating estimates of pairwise interactions between significant SNPs can yield greater power to detect association. To investigate epistatic interactions, each pair of SNPs that displayed a significant allelic association at the 5% level was evaluated by fitting a saturated logistic regression model and the log likelihood ratio statistic for comparison with the main effects model computed. This was compared against a χ² distribution with 1 degree of freedom (d.f.). Statistics were then adjusted for multiple testing using a Bonferroni correction.

Results

Data quality and genotyping success

Of the 3699 DNA samples submitted for genotyping, a total of 3657 samples were successfully processed, generating in excess of 4 million genotypes. Genotypes were obtained for 962 (97.0%) of 992 patients and 2695 (99.6%) of 2707 control subjects. The likelihood of a DNA sample failing to genotype correlated with sample DNA concentration. SNP call rates per sample for each of the 3657 DNA samples were greater than 99.6% in patients and control subjects. Of the 1467 SNPs submitted for analysis, 1218 SNPs were satisfactorily genotyped (83%), with mean individual sample call rates of 99.7% and 99.8% in patients and control subjects. Of the 3657 samples were successfully processed, generating in excess of 4 million genotypes. Genotypes were obtained for 962 (97.0%) of 992 patients and 2695 (99.6%) of 2707 control subjects. The likelihood of a DNA sample failing to genotype correlated with sample DNA concentration. SNP call rates per sample for each of the 3657 DNA samples were greater than 99.6% in patients and control subjects. Of the 1467 SNPs submitted for analysis, 1218 SNPs were satisfactorily genotyped (83%), with mean individual sample call rates of 99.7% and 99.8% in patients and control subjects.
subjects, respectively. Of the 1218 SNP loci satisfactorily genotyped, 188 were fixed in all samples, leaving 1030 SNPs for which genotype data were informative.

Population stratification

Of the 1030 polymorphic nsSNPs, 55 were found to violate HWE in controls at the 5% significance level (expected number of failures, 52). After Bonferroni correction, 6 SNPs still violated HWE and were removed, leaving a total of 1024 for further analysis. Each of the 6 SNPs removed had low genotyping reliability scores. Table S1 details all MAF data in 2695 controls for each of the 1024 nsSNPs. Of the remaining SNPs that violated HWE at the nominal 5% level, none was associated ($P_{.05}$) with risk of CLL. Implementing the genomic control method indicated no evidence of population stratification in our data as a cause of false-positive results, as the 95% confidence interval for the stratification parameter $\lambda$ (0.92-1.31) encompassed unity. Furthermore, no evidence was found for differences in allele frequencies of SNPs between male and female control subjects as a source of potential confounding in subsequent analyses.

SNPs and risk of CLL

Statistically significant associations were identified for 49 of 1024 SNPs at the 5% level by means of the $T_A$ statistic, 3 of which were significant at the 0.1% level (Table 1). The test statistics $T_D$ and $T_R$ and ORs under dominant and recessive models were computed for 1024 and 866 SNPs with sufficient MAF, respectively.

Of the 49 SNPs showing significant association ($P_{.05}$), 2 SNPs have previously been documented to be functional: I157T in CHK2 checkpoint yeast homolog (CHEK2 [MIM 604373]), a cell-cycle checkpoint regulator, and P1054R in ataxia telangiectasia mutated (ATM [MIM 607585]), a cell-cycle checkpoint kinase required for cellular response to DNA damage. In addition, 1 SNP encodes a termination codon; S474X in lipoprotein lipase (LPL [MIM 238600]), and a further 31 SNPs are predicted by at least 1 in silico algorithm to be deleterious (Table 2).

ATM SNPs F858L (rs1800056) and P1054R (rs1800057), which are in strong LD, showed the most significant allelic association with CLL, with strongest association under a dominant model ($OR_D = 2.28; 95\% CI, 1.53-3.40; P_D < .0001; OR_D = 1.68; 95\% CI, 1.25-2.28; P_D = .0006$), respectively. After permutation analysis to adjust for multiple testing, $ATM$ F858L (rs1800056) was found to still be significantly associated with CLL risk, with adjusted $P = .03$ at the genome-wide level. Additionally, the haplotype formed by the minor alleles of $ATM$ F858L and P1054R was significantly overrepresented in patients compared with control subjects ($OR_{.05} = 2.32; 95\% CI, 1.56-3.45; P_D < .0001; P = .01$ after permutation testing).

Nine additional SNPs located within the DNA damage-response axis also showed significant association (Figure 1; Table 1). Pre-eminent SNPs on the basis of biologic relevance were $I157T$ (rs17879961) in CHEK2 (OR$_D = 14.83; 95\% CI, 1.85-8.83; P_D = .0008$), N372H (rs144848) in breast cancer 2 early onset (BRCA2 [MIM 600185]), a tumor suppressor involved in DNA double-strand break repair (OR$_{.05} = 1.45; 95\% CI, 1.13-1.86; P_{.05} = .0032$), and Q349R (rs1801376) in BUB1 budding uninhibited by benzimidazoles yeast homolog 1 (BUB1B [MIM 602860]), encoding a kinase involved in spindle checkpoint function (OR$_{.05} = 1.42; 95\% CI, 1.12-1.81; P_{.05} = .0038$).

Stratification of patients by sex, family history of the disease, and age at diagnosis ($\leq 60$ years, $>60$ years) did not significantly affect study findings. We examined for interactive effects between the 49 SNPs significantly associated with risk of CLL ($P_{.05}$) by fitting full logistic regression models for each pair, generating 1176 models, and comparing these with the main effects model. The strongest interaction was between $BRCA2$ N372H and $EPH$ receptor A7 (EPHA7 [MIM 602190]) P278S ($P = .0007$), albeit nonsignificant after correction for multiple testing.

Discussion

We evaluated nsSNPs on the basis that each has the capacity to directly affect the function of expressed proteins, implying a higher probability of being directly causally related to susceptibility. Allelic loss in cells used in genetic analyses is a potential source of bias, because an apparent increase in homozygosity may be due to loss of heterozygosity in tumor leukocytes. There was no evidence of such confounding in our study as a source of spurious results, since the number of SNPs showing deviation from Hardy-Weinberg equilibrium followed the expected distribution, and associations were primarily based upon an overrepresentation of heterozygotes.

For 2 of the nsSNPs identified, CHEK2 I157T and ATM P1054R, there is evidence they are likely to directly affect the risk of malignancy. Furthermore, for an additional 32 of the SNPs significantly associated with CLL risk, the substitution either resulted in a termination codon or was predicted to be functionally deleterious using the in silico algorithms PolyPhen and/or SIFT. Although predictions about the functional consequences of amino acid changes are not definitive, these algorithms have been demonstrated in benchmarking studies to successfully categorize 80% of amino-acid substitutions.

Through interrogation of the Pathway Assist program (Stratagene, La Jolla, CA), 11 of the 49 associated SNPs were found within genes encoding pivotal components of the ATM-BRCA2-CHEK2 DNA damage-response and cell-signaling pathways.

The 3 SNPs in ATM associated with increased risk of CLL, F582L, F858L, and P1054R, are each predicted to be deleterious. Heterozygosity for P1054R has been reported to be associated with decreased ATM expression in CLL; furthermore, cell lines from breast cancer patients harboring the linked heterozygous F858L and P1054R variants exhibited increased radiosensitivity. ATM P1054R has previously been associated with an increased risk of breast and prostate cancer. While the functional significance of F582L is unknown, this SNP has previously been reported to confer an elevated risk of acute lymphocytic leukemia.

We have recently conducted a genome-wide linkage search of 115 families segregating CLL and other related B-cell LPDs but did not demonstrate significant linkage to ATM ($P = .08$). This observation is not contradictory to our current findings of an overrepresentation of the minor alleles of ATM F858L and P1045L in patients with CLL as the impact of ATM on the familial risk of CLL generated by both variants is approximately 1.03, insufficient to generate a significant departure in expected allele-sharing probabilities between affected individuals in the 115 families.

ATM is critical for regulation of cell-cycle checkpoints, and activation of ATM by DNA damage leads to ATM-dependent phosphorylation of CHEK2. CHEK2 I157T is localized in a functionally important domain of CHEK2, and the variant protein has been shown to be defective in its ability to bind TP53 and BRCA1. Previously, CHEK2 I157T has been associated with...
increased risk of breast, colon, kidney, and prostate cancers.34 Furthermore, possession of 157T has been shown to confer a 2-fold increase in risk of NHL,34 supporting the role of inherited dysregulation of CHEK2 in the development of B-cell LPDs.

**BRCA2** is involved in the monitoring and repair of DNA double-strand breaks.35 The minor allele of N372H has been documented to confer an elevated risk of breast36 and ovarian cancers.37 N372H is located between residues 290 and 453 of BRCA2, a region shown to interact with the transcriptional coactivator P/CAF,38 and hence has the potential to directly modify BRCA2-mediated regulation of transcription.

An additional 6 nsSNPs were identified in genes that interact either directly or indirectly with the *ATM-BRCA2-CHEK2* DNA damage-response axis. These include SNPs D784V in *EGF*, I253M in insulin-like growth factor–binding protein 1 (*IGFBP1* [MIM 146730]), and R574P in matrix metalloproteinase 9 (*MMP9* [MIM 120361]), which are involved in Sp1-mediated down-regulation of *ATM* transcription by EGF.39 Despite the low minor-allele frequency of

### Table 1. SNPs showing significant allelic association with chronic lymphocytic leukemia

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene*</th>
<th>Substitution</th>
<th>MAF†</th>
<th>OR (95% CI)</th>
<th>P*</th>
<th>ORsim (95% CI)†</th>
<th>Psim‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs180056</td>
<td>ATM</td>
<td>F858L</td>
<td>0.014</td>
<td>2.14 (1.45, 3.16)</td>
<td>.001</td>
<td>2.28 (1.53, 3.40)</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>SNP</td>
<td>Substitution</td>
<td>Predicted functionality*</td>
<td>Gene†</td>
<td>Gene description†</td>
<td>Gene ontology‡</td>
<td>MIM§</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>--------------</td>
<td>--------------------------</td>
<td>-------</td>
<td>------------------</td>
<td>----------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>rs1800056</td>
<td>F858L</td>
<td>Possibly damaging/potentially intolerant</td>
<td>ATM</td>
<td>Ataxia telangiectasia mutated</td>
<td>DNA repair, transcription regulation, cell-cycle control</td>
<td>607585</td>
<td></td>
</tr>
<tr>
<td>rs1800057</td>
<td>P1054R</td>
<td>Probably damaging/Intolerant</td>
<td>ATM</td>
<td>Ataxia telangiectasia mutated</td>
<td>DNA repair, transcription regulation, cell-cycle control</td>
<td>607585</td>
<td></td>
</tr>
<tr>
<td>rs1801376</td>
<td>Q349R</td>
<td>Potentially damaging</td>
<td>BUB1B</td>
<td>BUB1 budding uninhibited by benznidazoles 1 homolog beta</td>
<td>Cell cycling and proliferation</td>
<td>602860</td>
<td></td>
</tr>
<tr>
<td>rs17879961</td>
<td>I157T</td>
<td>—</td>
<td>CHEK2</td>
<td>CHK2 checkpoint homolog</td>
<td>DNA damage checkpoint</td>
<td>604373</td>
<td></td>
</tr>
<tr>
<td>rs10927851</td>
<td>F191S</td>
<td>—</td>
<td>FBLIM1 (FBLP-1)</td>
<td>Filamin-binding LIM protein-1</td>
<td>Cell adhesion</td>
<td>607747</td>
<td></td>
</tr>
<tr>
<td>rs9975588</td>
<td>S102L</td>
<td>Potentially damaging</td>
<td>MCM3AP</td>
<td>MCM3 minichromosome maintenance—deficient 3 associated protein</td>
<td>DNA replication</td>
<td>603294</td>
<td></td>
</tr>
<tr>
<td>rs1048201</td>
<td>R209Q</td>
<td>Potentially damaging/potentially intolerant</td>
<td>NUDT6</td>
<td>Nudix (nucleoside diphosphate linked moiety X) motif 6</td>
<td>Growth factor</td>
<td>606261</td>
<td></td>
</tr>
<tr>
<td>rs2250889</td>
<td>R574P</td>
<td>Possibly damaging</td>
<td>MMP9</td>
<td>Matrix metalloproteinase 9</td>
<td>Collagen catabolism</td>
<td>120361</td>
<td></td>
</tr>
<tr>
<td>rs1801265</td>
<td>C29R</td>
<td>—</td>
<td>DPDYD</td>
<td>Dihydropyrimidine dehydrogenase</td>
<td>Thymidine/uracil catabolism</td>
<td>274270</td>
<td></td>
</tr>
<tr>
<td>rs2293925</td>
<td>R525W</td>
<td>Possibly damaging/Intolerant</td>
<td>TOP1MT</td>
<td>DNA topoisomerase 1</td>
<td>DNA unwinding</td>
<td>606387</td>
<td></td>
</tr>
<tr>
<td>rs4874147</td>
<td>G452S</td>
<td>Potentially damaging</td>
<td>ZC3H3 (ZC3HDC 3)</td>
<td>Zinc finger CCCH-type domain containing 3</td>
<td>Nucleic acid binding</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>rs7288201</td>
<td>I475V</td>
<td>Intolerant</td>
<td>GGTLA1</td>
<td>Gamma-glutamyltransferase-like activity 1</td>
<td>Amino acid metabolism</td>
<td>137168</td>
<td></td>
</tr>
<tr>
<td>rs10067</td>
<td>H604Q</td>
<td>Probably damaging/Intolerant</td>
<td>OVG1</td>
<td>Oviductal glycoprotein 1</td>
<td>Chitin catabolism</td>
<td>603578</td>
<td></td>
</tr>
<tr>
<td>rs17409304</td>
<td>Q306E</td>
<td>Intolerant</td>
<td>CLCA2</td>
<td>Chloride channel, calcium activated, member 2</td>
<td>Chloride transport</td>
<td>604003</td>
<td></td>
</tr>
<tr>
<td>rs11569017</td>
<td>D784V</td>
<td>Possibly damaging/Intolerant</td>
<td>EGF</td>
<td>Epidermal growth factor</td>
<td>DNA replication, receptor signaling, regulation of cell proliferation</td>
<td>131530</td>
<td></td>
</tr>
<tr>
<td>rs3206824</td>
<td>Q335R</td>
<td>Intolerant</td>
<td>DKK3</td>
<td>Dickkopf homolog 3</td>
<td>Receptor signaling</td>
<td>605416</td>
<td></td>
</tr>
<tr>
<td>rs144848</td>
<td>N372H</td>
<td>—</td>
<td>BRCA2</td>
<td>Breast cancer 2, early onset</td>
<td>DNA repair, regulation of transcription</td>
<td>600185</td>
<td></td>
</tr>
<tr>
<td>rs4619</td>
<td>I253M</td>
<td>—</td>
<td>IGFBP1</td>
<td>Insulin-like growth factor—binding protein 1</td>
<td>Regulation of cell growth, signal transduction</td>
<td>146730</td>
<td></td>
</tr>
<tr>
<td>rs16978999</td>
<td>S247P</td>
<td>Possibly damaging/Intolerant</td>
<td>ZNF233</td>
<td>Zinc finger protein 233</td>
<td>Regulation of transcription</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>rs712665</td>
<td>S155G</td>
<td>Possibly damaging</td>
<td>U2AF1L1</td>
<td>U2 small nuclear RNA auxiliary factor 1-like 1</td>
<td>Nucleotide binding</td>
<td>601079</td>
<td></td>
</tr>
<tr>
<td>rs5167</td>
<td>L96R</td>
<td>Probably damaging</td>
<td>APOC4</td>
<td>Apolipoprotein C-IV</td>
<td>Lipid metabolism</td>
<td>600745</td>
<td></td>
</tr>
<tr>
<td>rs2235006</td>
<td>F582L</td>
<td>Possibly damaging</td>
<td>ATM</td>
<td>Ataxia telangiectasia mutated</td>
<td>DNA repair, transcription regulation, cell cycle control</td>
<td>607585</td>
<td></td>
</tr>
<tr>
<td>rs328</td>
<td>S474X</td>
<td>Stop codon</td>
<td>LPL</td>
<td>Lipoprotein lipase</td>
<td>Fatty acid metabolism</td>
<td>238600</td>
<td></td>
</tr>
<tr>
<td>rs4150521</td>
<td>S704L</td>
<td>Possibly damaging</td>
<td>ERCC3</td>
<td>Excision repair cross-complementing rodent repair deficiency, complementation group 3</td>
<td>Induction of apoptosis, regulation of transcription</td>
<td>133510</td>
<td></td>
</tr>
<tr>
<td>rs10494745</td>
<td>G306E</td>
<td>Possibly damaging</td>
<td>CFHL4</td>
<td>Complement factor H-related 4</td>
<td>Lipid transporter activity</td>
<td>605337</td>
<td></td>
</tr>
<tr>
<td>rs2269429</td>
<td>G255S</td>
<td>Possibly damaging</td>
<td>TNXB</td>
<td>Tenascin XB</td>
<td>Cell-matrix adhesion</td>
<td>600985</td>
<td></td>
</tr>
<tr>
<td>rs17050550</td>
<td>A85S</td>
<td>—</td>
<td>OGG1</td>
<td>8-Oxoguanine DNA glycosylase</td>
<td>Base-excision repair</td>
<td>601982</td>
<td></td>
</tr>
<tr>
<td>rs17293607</td>
<td>G65R</td>
<td>Intolerant</td>
<td>MMP10</td>
<td>Matrix metalloproteinase 10</td>
<td>Collagen catabolism</td>
<td>185260</td>
<td></td>
</tr>
<tr>
<td>rs619203</td>
<td>S229C</td>
<td>Probably damaging</td>
<td>ROS1</td>
<td>v-ros UR2 sarcoma virus oncogene homolog 1</td>
<td>Signal transduction</td>
<td>165020</td>
<td></td>
</tr>
<tr>
<td>rs459552</td>
<td>V1822D</td>
<td>—</td>
<td>APC</td>
<td>Adenomatosis polyposis coli</td>
<td>Receptor signaling, cell-cycle control</td>
<td>175100</td>
<td></td>
</tr>
<tr>
<td>rs508405</td>
<td>T280A</td>
<td>Potentially intolerant</td>
<td>CAPN13</td>
<td>Calpain 13</td>
<td>Proteolysis and peptidolysis</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>rs3136797</td>
<td>P242R</td>
<td>Possibly damaging/Intolerant</td>
<td>POLB</td>
<td>Polymerase (DNA directed), beta</td>
<td>DNA repair, DNA replication</td>
<td>174760</td>
<td></td>
</tr>
</tbody>
</table>
the SNPs individually associated with risk of CLL in our study, there was some evidence for an interaction between BRCA2 N372H and EPHA7 P278S (P = .0007), albeit nonsignificant after correction for multiple testing.

The prior probability of identifying a significant association with CLL risk for a series of SNPs mapping to a single gene pathway is intuitively small. Genotyping a total of 81 SNPs across 50 genes (including ATM, BRCA2, and CHEK2) implicated in the DNA damage-response axis for SNPs associated with risk of CLL. Relationships between genes were established using Pathway Assist software; supporting publications are indicated with their corresponding NCBI Entrez PubMed ID number in square brackets. (1) Regulation [15020226]; (2) Expression [8625897]; (3) Regulation [12511596]; (4) Expression [9096655]; (5) Expression [11751435]; (6) Binding [10973490]; (7) Binding [11438675]; (8) Binding [10866324]; (9) Binding [9774970]; (10) Binding [11034101]; (11) Binding [11034101]; and (12) Binding [12815053].

![Figure 1. Interrelationship between genes involved in the DNA damage-response axis for SNPs associated with risk of CLL. Relationships between genes were established using Pathway Assist software; supporting publications are indicated with their corresponding NCBI Entrez PubMed ID number in square brackets.](image-url)
cell-cycle pathway via Gene Ontology Consortium annotations identified 8 SNPs displaying statistical association with risk of CLL, a significantly greater number than expected (P < .05). By contrast, no significant associations were observed for SNPs mapping to genes encoding components of the cell-cell signaling (21 SNPs, 18 genes) and cell differentiation (20 SNPs, 17 genes) pathways.

Several lines of evidence support a role for inherited dysfunction in the ATM-CHEK2-BRCA2 axis as a cause of predisposition to CLL. Recessive ataxia telangiectasia (A-T), caused by mutations in ATM, is well established to confer a substantive increase in risk of LPD,\(^{40}\) and an overrepresentation of LPD has been documented in relatives of patients with A-T.\(^{41}\) Mutations in ATM, CHEK2, and BRCA2 are documented to confer an increased risk of breast cancer. This fact, coupled with the elevated risk of LPD reported in relatives of patients with breast cancer,\(^{4}\) suggests that a subset of breast cancers and LPDs have a common biology.

Our study provides evidence that inherited predisposition to CLL is in part mediated through low-penetrance alleles, specifically variants in the ATM-BRCA2-CHEK2 DNA damage-response axis. Clearly it is, however, desirable to validate our study findings through analysis of additional large datasets.

Acknowledgments

We gratefully acknowledge the participation of all patients with CLL and control individuals. The authors are indebted to Ruth Allison, Richard Coleman, Christina Fleischmann, Nicholas Hearle, Athena Matakidiou, Mobshra Qureshi, Hayley Spendlove, and Remben Talaban for sample ascertainment.

References

24. XI T, Jones IM, Mohrenweiser HW. Many amino acid substitution variants identified in DNA repair genes during human population screenings are predicted to impact protein function. Genomics. 2004;83:970-979.
Variants in the ATM-BRCA2-CHEK2 axis predispose to chronic lymphocytic leukemia

Matthew F. Rudd, Gabrielle S. Sellick, Emily L. Webb, Daniel Catovsky and Richard S. Houlston