Human Genome Epidemiology (HuGE) Review

Synopsis and Synthesis of Candidate-Gene Association Studies in Chronic Lymphocytic Leukemia: The CUMAGAS-CLL Information System

Elias Zintzaras and George D. Kitsios

Initially submitted December 10, 2008; accepted for publication June 9, 2009.

A comprehensive and systematic assessment of the current status of candidate-gene association studies for chronic lymphocytic leukemia (CLL) was conducted. Data from 989 candidate-gene association studies (1992–2009) involving 905 distinct genetic variants were analyzed and cataloged in CUMAGAS-CLL, a Web-based information system which allows the retrieval and synthesis of data from candidate-gene association studies on CLL (http://biomath.med.uth.gr). Nine genetic variants (BAX (rs4645878), GSTM1 (null/present), GSTT1 (null/present), IL10 (rs1800896), LTA (rs909253), MTHFR (rs1801131), MTHFR (rs1801133), P2RX7 (rs3751143), and TNF (rs1800629)) were investigated in 4 or more studies, and their results were meta-analyzed. In individual studies, 147 variants showed a significant association with CLL risk under any genetic model. For 53 variants, the association was significant at \( P < 0.01 \) with an increased risk greater than 40%. Only 0.3% of studies had statistical power greater than 80%. In meta-analyses, none of the variants showed significant results, and heterogeneity ranged from none to high. Large and rigorous genetic studies (candidate-gene association studies and genome-wide association studies) designed to investigate epistatic and gene-environment interactions may produce more conclusive evidence about the genetic etiology of CLL. CUMAGAS-CLL would be a useful tool for current genomic epidemiology research in the field of CLL.

database; epidemiology; genes; genome, human; information systems; leukemia, lymphocytic, chronic, B-cell; meta-analysis; polymorphism, genetic

Abbreviations: BAX, BCL2-associated X protein; CI, confidence interval; CLL, chronic lymphocytic leukemia; CUMAGAS, Cumulative Meta-Analysis of Genetic Association Studies; GSTM1, glutathione S-transferase M1; GSTT1, glutathione S-transferase T1; IL10, interleukin-10; LTA, lymphotoxin alpha; MTHFR, methylenetetrahydrofolate reductase; OR, odds ratio; P2RX7, purinergic receptor P2X, ligand-gated ion channel, 7; rs, reference SNP; SLL, small lymphocytic lymphoma; SNP, single nucleotide polymorphism; TNF, tumor necrosis factor.

Editor’s note: This article also appears on the Web site of the Human Genome Epidemiology Network (http://www.cdc.gov/genomics/hugenet/default.htm).

Candidate-gene association studies represent a popular approach for investigating the genetic etiology of complex diseases, such as chronic lymphocytic leukemia (CLL) (1, 2). These studies examine the relation between gene variants and a disease based on evidence regarding the function or location of postulated causal genes, and they are conducted using a sample of unrelated cases and controls (1).

CLL represents the most frequent type of adult leukemia in the Western world, and substantial evidence supports the existence of a genetic component in disease susceptibility (2, 3). CLL is currently classified within the category of malignant B-cell lymphomas. In addition, CLL and small lymphocytic lymphoma (SLL) are classified as the same disease at different stages, because they have the same immunophenotype and similar clinical courses (4). The number of published candidate-gene association studies on CLL...
has increased rapidly, and this trend is expected to accelerate because of the availability of mapped single nucleotide polymorphisms (SNPs) and advances in genotyping technology (2, 5). Considering the exponential accumulation of genetic association data, keeping track of the rapidly growing amount of evidence is a major challenge (5). The available evidence published on CLL to date is weak, owing to inconsistent and inconclusive results from individual studies, mainly attributed to small samples and the heterogeneity of studied populations. Aiming to overcome the obstacle of limited sample sizes and to harmonize ongoing research in the genetics of lymphoma, a movement towards consortial efforts has been made (http://epi.grants.cancer.gov/InterLymph/); this permits data pooling and the conduct of genome-wide association studies, in addition to candidate-gene association studies.

In order to explore the involvement of gene variants in CLL susceptibility and to make available to researchers the genetic risk effect of each variant implicated in CLL, we systematically searched the literature for all available candidate-gene association studies on CLL and created CUMAGAS-CLL [Cumulative Meta-Analysis of Genetic Association Studies–CLL], a Web-based information system (http://biomath.med.uth.gr) which allows the retrieval and synthesis of data from candidate-gene association studies on CLL with the capacity for continuous updating. In this project, we cataloged all retrieved articles and estimated the risk effects (for the allele-contrast, dominant, recessive, and/or additive models) of all individually investigated variants. Finally, we synthesized the available data using meta-analytic techniques in order to increase statistical power for detecting significant results and to decrease the uncertainty of the estimated genetic risk effects (1).

MATERIALS AND METHODS

Selection of studies

All studies published before February 2009 were identified through extended computer-based searches of the PubMed (National Library of Medicine (http://www.ncbi.nlm.nih.gov/PubMed)) and HuGE Literature Finder (Human Genome Epidemiology Network (http://hugenavigator.net/HuGENavigator/startPagePubLit.do)) databases. The search criterion included combinations of the following terms: “chronic lymphocytic leukemia,” “CLL,” “chronic lymphocytic leukaemia,” “chronic lymphoid leukemia,” “chronic lymphoid leukaemia,” “small lymphocytic lymphoma,” “SLL,” “lymphocytic lymphoma,” “non-Hodgkin’s lymphoma,” “NHL,” “gene,” “polymorphism,” “SNP,” “association,” “risk,” and “susceptibility.” Study eligibility criteria included peer-reviewed reports of candidate gene studies that provided the germ-line genotypic distribution of the genetic variant in CLL cases and cancer-free controls. Only studies carried out in human subjects that had used DNA-based analysis methods for genotyping were considered. Family-based association studies were not considered because of different design considerations. Case reports, editorials, and review articles were excluded. Non-English articles were also excluded. Abstracts of retrieved studies were independently read by each of the investigators (G. D. K., E. Z.) to assess their appropriateness for this study. The results were compared, and disagreements were resolved by consensus. Full-text articles of the studies were evaluated according to the inclusion criteria. We also reviewed all references cited in the studies to identify additional published work not indexed by the searched databases.

Data abstraction

From each study, the following information was extracted: name of first author, year of publication, race/ethnicity of the study population, demographic data, definitions of cases and controls, and numbers of cases and controls for each genotype. The frequencies of the alleles and the genotypic distributions were extracted or calculated for both cases and controls. The reference SNP (rs) numbers and the nucleotide base changes for all genetic variants were identified through extended searches of bioinformatics databases (6–9).

Data analysis and synthesis

Prior to meta-analysis, the risk effects of gene variants for the allele-contrast (mutant-type vs. wild-type allele), recessive, dominant, and additive models were evaluated for each study separately. All associations are expressed as odds ratios with corresponding 95% confidence intervals.

When the same variant had been investigated in 4 or more studies, a meta-analysis of the published results was conducted. In the meta-analysis, heterogeneity between studies was tested using the Q statistic and quantified with the $I^2$ metric (1). The pooled odds ratio was estimated using a random-effects (DerSimonian and Laird) model (10). Random-effects modeling assumes a genuine diversity in the results of various studies, and it incorporates between-study variance into the calculations. When there is lack of heterogeneity, the random-effects model coincides with the fixed-effects model (1). The differential magnitude of effect in large studies versus small studies (of variants included in the meta-analysis) was checked for the allele contrast, using the test proposed by Harbord et al. (11). Each meta-analysis consisted of the main (overall) analysis, which included all available data; subgroup analyses carried out by race/ethnicity; and sensitivity analysis, which examined the effect of excluding specific studies (1).

The distribution of each variant in the control group was tested for Hardy-Weinberg equilibrium. Since lack of Hardy-Weinberg equilibrium indicates possible genotyping errors and/or population stratification, a sensitivity analysis was carried out for these studies (1). Hardy-Weinberg equilibrium was tested for the whole study population in each study. For studies with mixed populations, testing for Hardy-Weinberg equilibrium by race/ethnicity was performed when relevant data were available. The statistical power of each study to assess the allele contrast was calculated, assuming an increase in effect size of 20% (a modest effect), a significance level of 0.05, and a minor allele frequency equal to that of the study population.
Analyses were performed using CUMAGAS-CLL (http://biomath.med.uth.gr) and Compaq Visual Fortran 90 (Hewlett-Packard Company, Palo Alto, California) with the International Mathematics and Statistics Library (Visual Numerics, Inc., Houston, Texas). Power was calculated using the CaTS Power Calculator for Genetic Studies (Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan).

CUMAGAS-CLL

CUMAGAS-CLL performs meta-analysis for all genetic models (allele-contrast, dominant, recessive, and additive) and provides information on study design and gene polymorphism characteristics, including nucleotide-base changes and dbSNP (National Center for Biotechnology Information) rs numbers. CUMAGAS-CLL has the capacity for continuous updating (we currently aim to update the system annually), and authors of published studies can contribute new data, correct existing data, or notify us about missed studies by contacting us via e-mail (cumagas@med.uth.gr).

RESULTS

Eligible articles

The literature review identified 719 titles that met the search criteria. After abstract selection, 217 articles remained (see the Web Appendix, which is posted on the Journal’s Web site (http://aje.oxfordjournals.org/)). Eighty-one articles in which investigators had studied associations between genetic variants and CLL fulfilled the inclusion criteria. Figure 1 shows a flowchart of the studies retrieved and the studies excluded, with specification of reasons. Overall, 260 candidate genes and 905 distinct variants of these genes investigated in 989 gene-disease association studies were identified in the field of CLL. The studies were published between 1992 and 2009.

Studies’ characteristics and association results

The characteristics of each study and the association results for variants are shown in Web Table 1, which is posted on the Journal’s Web site (http://aje.oxfordjournals.org/).

Studies were conducted in various populations of different racial/ethnic/ancestry backgrounds: 268 studies involved solely whites, 1 study involved East Asians, and 720 studies were conducted in ethnically mixed populations. The distribution of genotypes in the control group departed from Hardy-Weinberg equilibrium in 126 studies. In 133 studies, there was not enough information to test for Hardy-Weinberg equilibrium. In 5.6% of the studies, statistical power was greater than 50%, and in 2.6% of studies, the power ranged from 25% to 50%. Only 0.3% of studies had power greater than 80%.

In total, 147 variants produced a significant association with CLL risk under any genetic model. The association was significant at $P < 0.01$ for any genetic model with an increased risk (or prevention) of more than 40% for 53 variants: ADCY4 (rs3212254), APAF1 (rs17028658), B3GNT3 (rs36686), BCL2 (rs2849377), BCL6 (rs1056932), BCL6 (rs3172469), CASP8 (rs1045485), CCNH (rs2230641), CCR7 (rs3136687), CD28 (rs3116496), CD38 (rs6449182), CD5 (rs1800561), CD5 (19CA repeats), CRY2 (rs7123390), CRY2 (rs1401417), CTLA4 (rs5742909), CYP19A1 (rs2899472), CYP19A1 (rs2008691), CYP19A1 (rs9944225), CYP19A1 (rs7181886), CYP19A1 (rs936306), CYP19A1 (rs10046), CYP19A1 (rs1870049), ESR1 (rs2813545), ETS1 (SstI C1/C2), FAS (rs12765241), FAS (rs9658761), FGG (rs1800792), GSTM1 (null/present), IL13 (rs20541), IL1B (rs16944), LSP1 (rs2089910), LTA (rs909253), MCL1 (rs3738484), MMP3 (rs564018), MYC (rs16902359), P2RX7 (rs3751143), PON1 (rs854560), RIPK3 (rs3212254), SCNN1A (rs11064145), SLC23A2 (rs6133175), SLC23A2 (rs1776948), SLC23A2 (rs1715364), TNF (rs1800629), TNFRSF10A (rs20575), TNFRSF10A (rs20576), TNFSF17 (rs11570151), TNFSF10 (rs3136609), TNFSF10 (rs9879554), TNFSF10 (rs17601879), TNFSF13B (rs17499386), TNFSF13B (rs1224163), and TRAF1 (rs2269059).

Meta-analysis results

In total, 9 variants were investigated in 4 or more studies, and their results were meta-analyzed. None of the variants showed a significant association for any genetic model. Table 1 shows the meta-analysis results for the associations between the different variants and the risk of developing CLL. Below, we present briefly the findings of the meta-analyses.

BAX (rs4645878). Four studies investigated the association between the BCL2-associated X protein (BAX) rs4645878 SNP and CLL in white populations (12–15). The meta-analysis of the allele contrast showed significant heterogeneity between studies ($p_{Q} = 0.03, I^{2} = 64%$), and the association was nonsignificant (odds ratio (OR) = 1.42, 95% confidence interval (CI): 0.82, 2.44). The dominant, recessive, and additive models also produced nonsignificant results.

GSTM1 (null/present) and GSTT1 (null/present). The glutathione S-transferase M1 (GSTM1) gene was investigated in 5 studies (16–20), and in 4 of them investigators also examined the glutathione S-transferase T1 (GSTT1) gene (16–18, 20). The overall meta-analysis for the association between the GSTM1 null genotype and CLL risk, relative to the normal genotype, revealed significant heterogeneity between studies ($p_{Q} = 0.01, I^{2} = 67%$). The association for the allele contrast was nonsignificant (OR = 1.41, 95% CI: 0.82, 2.42). In subgroup analysis carried out according to race/ethnicity, the odds ratio was nonsignificant in whites (OR = 1.49, 95% CI: 0.76, 2.92). For the GSTT1 gene, the association was also nonsignificant (OR = 1.31, 95% CI: 0.87, 1.97; $p_{Q} = 0.24, I^{2} = 27%$). In subgroup analysis for whites, similar results were obtained.

IL10 (-1082 A/G, rs1800896). The meta-analysis of the rs1800896 variant of the interleukin-10 (IL10) gene in relation to CLL development involved 4 studies (21–24) and produced nonsignificant results for the allele contrast (OR = 1.07, 95% CI: 0.88, 1.30; $p_{Q} = 0.34, I^{2} = 10%$). The dominant, recessive, and additive models also showed
nonsignificant associations. The subgroup analysis by race/ethnicity for whites produced the same pattern of results.

**LTA (rs909253)**. Five studies investigated the association between lymphotoxin alpha (LTA) rs909253 and CLL in white and mixed populations (22, 24–27). The meta-analysis of findings from these studies derived a nonsignificant association for all genetic models. The odds ratio for the allele contrast was 1.15 (95% CI: 0.80, 1.67), and the heterogeneity was large ($I^2 = 64\%$).

**MTHFR (rs1801131)** and (rs1801133). Four studies examined both the methylene-tetrahydrofolate reductase (MTHFR) rs1801133 variant and the MTHFR rs1801131 variant in relation to CLL (16, 28–30). The meta-analysis of the rs1801133 variant derived significant heterogeneity between studies for the allele contrast ($p_Q = 0.02, I^2 = 64\%$), with the association being nonsignificant (OR = 0.97, 95% CI: 0.86, 1.09). The dominant, recessive, and additive models also showed nonsignificant associations. The subgroup analysis by race/ethnicity for whites showed the same pattern of results. The meta-analysis of the rs1801131 variant also produced a nonsignificant association for the allele contrast (OR = 1.00, 95% CI: 0.88, 1.11; $p_Q = 0.74, I^2 = 0\%$). The dominant and recessive models produced no significant associations overall or in whites.

**P2RX7 (rs3751143)**. Eight studies in whites investigated the association between purinergic receptor P2X, ligand-gated ion channel, 7 (P2RX7) rs3751143 and CLL (31–38). The heterogeneity between studies for the allele contrast was significant ($p_Q = 0.05, I^2 = 49\%$), but the association was not significant (OR = 0.98, 95% CI: 0.78, 1.22). The dominant, recessive, and additive models also produced nonsignificant results.

**TNF (rs1800629)**. Eight studies examined the association between the tumor necrosis factor (TNF) rs1800629 variant and CLL (22, 24–27, 39–41). The meta-analysis for the allele contrast showed that the association was nonsignificant (OR = 0.92, 95% CI: 0.67, 0.97).
<table>
<thead>
<tr>
<th>Gene, Variant (Wild Type/Mutant), and rs No.</th>
<th>Population(s)</th>
<th>No. of Cases</th>
<th>No. of Controls</th>
<th>Allele-Contrast (Mutant vs. Wild Type) Model</th>
<th>Dominant Model</th>
<th>Recessive Model</th>
<th>Additive Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAX (-248 G/A), rs4645878</td>
<td>Whites</td>
<td>812</td>
<td>462</td>
<td>4</td>
<td>1.42</td>
<td>0.82, 2.44</td>
<td>0.03</td>
</tr>
<tr>
<td>GSTM1 (null/present)</td>
<td>Whites and ethnically mixed populations</td>
<td>288</td>
<td>1,187</td>
<td>5</td>
<td>1.41</td>
<td>0.82, 2.42</td>
<td>0.01</td>
</tr>
<tr>
<td>GSTT1 (null/present)</td>
<td>Whites and ethnically mixed populations</td>
<td>275</td>
<td>1,067</td>
<td>4</td>
<td>1.31</td>
<td>0.87, 1.97</td>
<td>0.24</td>
</tr>
<tr>
<td>IL10 (-1082 A/G), rs1800896</td>
<td>Whites and ethnically mixed populations</td>
<td>304</td>
<td>1,735</td>
<td>4</td>
<td>1.15</td>
<td>0.87, 1.52</td>
<td>0.35</td>
</tr>
<tr>
<td>In HWE</td>
<td>204</td>
<td>1,075</td>
<td>3</td>
<td>1.20</td>
<td>0.86, 1.66</td>
<td>0.03</td>
<td>30</td>
</tr>
<tr>
<td>LTA (-252 A/G), rs909253</td>
<td>Whites and ethnically mixed populations</td>
<td>316</td>
<td>1,276</td>
<td>5</td>
<td>1.15</td>
<td>0.80, 1.67</td>
<td>0.02</td>
</tr>
<tr>
<td>MTHFR (1298 A/C), rs1801131</td>
<td>Whites and ethnically mixed populations</td>
<td>1,170</td>
<td>2,090</td>
<td>4</td>
<td>1.00</td>
<td>0.88, 1.11</td>
<td>0.74</td>
</tr>
<tr>
<td>MTHFR (677 C/T), rs1801133</td>
<td>Whites and ethnically mixed populations</td>
<td>1,168</td>
<td>2,080</td>
<td>4</td>
<td>0.97</td>
<td>0.86, 1.09</td>
<td>0.76</td>
</tr>
<tr>
<td>P2RX7 (-308 A/G), rs3751143</td>
<td>Whites, and rs3751143</td>
<td>1,142</td>
<td>1,725</td>
<td>8</td>
<td>0.98</td>
<td>0.78, 1.22</td>
<td>0.05</td>
</tr>
<tr>
<td>In HWE</td>
<td>972</td>
<td>1,525</td>
<td>7</td>
<td>1.04</td>
<td>0.84, 1.28</td>
<td>0.15</td>
<td>36</td>
</tr>
<tr>
<td>TNF (-308 G/A), rs1800629</td>
<td>Whites, ethnically mixed populations, and East Asians</td>
<td>591</td>
<td>3,278</td>
<td>8</td>
<td>0.92</td>
<td>0.67, 1.26</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Abbreviations: BAX, BCL2-associated X protein; CI, confidence interval; GSTM1, glutathione S-transferase M1; GSTT1, glutathione S-transferase T1; HWE, Hardy-Weinberg equilibrium; IL10, interleukin-10; LTA, lymphotoxin alpha; MTHFR, methylenetetrahydrofolate reductase; OR, odds ratio; P2RX7, purinergic receptor P2X, ligand-gated ion channel; 7; rs, reference SNP; SNP, single nucleotide polymorphism; TNF, tumor necrosis factor.

* Odds ratios for chronic lymphocytic leukemia, corresponding 95% confidence intervals, and heterogeneity metrics (q and I²) for the allele-contrast, dominant, recessive, and additive models. Results of the analysis for the differential magnitude of effect in large studies versus small studies (pH) are described in the text.
1.26; \( p_Q = 0.08, I^2 = 44\% \)). The remaining genetic models also produced nonsignificant results. The subgroup analysis in whites produced nonsignificant results as well. **Potential bias.** The sensitivity analyses excluding studies not conforming to Hardy-Weinberg equilibrium (1 study on **IL10** (rs1800896), 1 study on **P2RX7** (rs3751143), and 1 study on **TNF** (rs1800629)) did not alter the pattern of results for any variant in any contrast examined. There was no differential magnitude of effect in large studies versus small studies for any variant \((p_H \geq 0.10)\), except for **BAX** (rs4645878) \((p_H = 0.04)\).

**DISCUSSION**

This project represents, to our knowledge, the first comprehensive and systematic assessment of the current status of genetic epidemiology research for **CLL**. Our primary aim in the present study was to help guide the direction of future **CLL** research by identifying all potential candidate variants. This goal was achieved through the establishment of CUMAGAS-CLL, a database information system for performing cumulative meta-analysis on each variant implicated in **CLL** susceptibility. Our secondary aim was to stress the necessity of moving towards consortial efforts for performing large candidate-gene association studies and even genome-wide association studies, via shared resources and harmonized methods.

Data from 989 candidate-gene association studies described in 81 published articles were cataloged in CUMAGAS-CLL. With the implementation of CUMAGAS-CLL, summary effect estimates were calculated in the context of 9 meta-analyses for genetic variants and risk of **CLL**. The resulting evidence provided insights regarding the roles of numerous candidate genes in **CLL** susceptibility.

In general, individual candidate-gene association studies have small sample sizes, and this results in insufficient power to detect the minor contributing role of common alleles. Indeed, the most realistic genetic association between a polymorphic locus and a disease has been claimed to yield an odds ratio between 1.1 and 1.5. Therefore, to achieve satisfactory statistical power (>80%) to identify a modest genetic effect \((\text{OR} = 1.2)\) of a polymorphism present in 10% of individuals, a sample size of 10,000 subjects or more would be needed for a candidate-gene association study (1, 5). The studies on **CLL** have had modest sample sizes and have consequently been underpowered for detection of weak associations. Overall, no compelling associations between genetic variants and risk of **CLL** emerged from our analysis. Only 5.9% of investigated variants showed significant results in individual studies, whereas no meta-analysis detected any significant association. This lack of positive signals with strong credibility should be viewed in the light of the inherent weaknesses of genetic epidemiology investigations.

The rarity of **CLL** is probably a major limiting factor in recruiting large study samples (42). Since the requirements for sample sizes will be far beyond what is currently available and no single institution alone will be able to provide a reasonable number of patients, the creation of large databases and consortia to facilitate the sharing of resources among investigators would be a straightforward step.

The need for data-sharing has been highlighted by the formation of the Genetic Association Information Network, which aims to facilitate the subsequent and joint analysis of data from genome-wide association studies (43). In an effort to pool data across studies and to undertake collaborative research in lymphoma, the International Lymphoma Epidemiology Consortium (http://epi.grants.cancer.gov/InterLymph/) was established in 2001. Apart from the need for larger samples, selecting cases that are genetically loaded (i.e., cases with a strong familial history) may also improve power (44).

Meta-analysis clearly has a role in offering an analysis with the potential for higher statistical power, although there is no formal method for calculating the power of a meta-analysis. Meta-analysis is a retrospective all-inclusive synthesis of results from published studies, and power analysis is not applicable. Nevertheless, type II errors are expected to be less common in a meta-analysis than in single studies (1). In the present meta-analyses, the results were based on a small number of studies and subjects; therefore, interpretation should be cautious.

Combining patients with **CLL** and SLL could potentially introduce phenotypic heterogeneity in the definition of cases for **CLL** studies, resulting in diminished statistical power. However, the effects of genetic variants for **CLL** and SLL are hypothesized to be similar, since these 2 subtypes share the same molecular immunophenotype (2). Conversely, a significant amount of data was untraceable, because in many candidate-gene association studies on the genetic etiology of non-Hodgkin’s lymphoma, the investigators did not present genotypic data on disease subtypes (including **CLL**/SLL) and thus the studies were excluded from our analysis. Potential misclassification of **CLL**/SLL with other cancers, such as monocytoid B-cell lymphoma or mantle cell lymphoma (45, 46), cannot be ruled out, particularly for early studies that did not include immunophenotypic criteria for diagnosis.

As with other complex diseases, the development of **CLL** is probably determined by multiple epistatic and gene-environment interactions. The search for susceptibility loci has probably been complicated by the increased number of contributing loci and susceptibility alleles. Because most genetic risk factors are believed to contribute only small influences to disease susceptibility, analysis of gene variants in disease-associated molecular pathways may prove to be more useful in finding a panel of genes that operate together (47). In addition, despite difficulties in study design and assessment of environmental exposures, such parameters should be incorporated in future studies (48).

Our analysis used the available study-level allele and genotype distributions, precluding adjusted analysis for potential gene–gene and gene–environment interactions, for which raw genotype data would be required. Failure to account for interactions may have reduced the power of our analysis, but it is unlikely to have inflated the number of positive results.

The lack of consistently replicated susceptibility genes for **CLL** brings into question the reliability of the candidate gene approach for elucidating the complex genetic
architecture of the disease. The biologic plausibility of investigated candidate genes might have been exaggerated, and the selection of genetic variants was possibly unsuccessful. The first genome-wide association studies of CLL provided “hypothesis-free” evidence for low-risk variants predisposing to a hematologic malignancy (49). Interestingly, none of the 6 loci that emerged from this study (IRF4, PRKD2, SP140, GRAMD1B, ACOXL, and a nongenic region on chromosome 15) had been previously investigated in candidate-gene association studies, thus suggesting novel pathogenetic mechanisms. Nevertheless, positive “hits” from the massive genomic scans produced by genome-wide association studies warrant careful replication in independent cohorts.

CUMAGAS-CLL represents an evidence-based approach combined with an electronic information system to systematically search, review, and synthesize the rapidly emerging body of candidate-gene association studies in CLL, with the capacity for continuous updating. Available evidence is cataloged and, where appropriate, synthesized with meta-analytic techniques, highlighting the strengths as well as the gaps of research in the field. The CUMAGAS database will be expanded to additional complex phenotypes; it will incorporate the findings from emerging genome-wide association studies and will be updated on an annual basis as evidence accumulates.

In summary, there was no evidence of strong association between genetic variants and the risk of developing CLL in these individual studies and meta-analyses. Collaborative researchers in lymphoma may help in identifying the contributing role of variants by performing candidate-gene association studies and genome-wide association studies with adequate power. Furthermore, the design of rigorous studies for investigating epistatic and gene-environment interactions and the utilization of data generated by genomic studies may help in deriving more conclusive claims about the genetics of CLL. The CUMAGAS-CLL information system would be a useful resource for reviewing and interpreting the findings of accumulating genomic epidemiology research in CLL.

ACKNOWLEDGMENTS

Author affiliations: Department of Biomathematics, School of Medicine, University of Thessaly, Larissa, Greece (Elias Zintzaras, George D. Kitsios); and Center for Clinical Evidence Synthesis, Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts Medical Center, Tufts University School of Medicine, Boston, Massachusetts (Elias Zintzaras, George D. Kitsios).

This project was partially funded by the Bodosakis Foundation, the University of Thessaly Research Committee, and the ALPHA Bank.

Conflict of interest: none declared.

REFERENCES


Am J Epidemiol 2009;170:671–678


