Candidate gene association analysis of acute lymphoblastic leukemia identifies new susceptibility locus at 11p15 (LMO1)

Joke Beuten1,2,*, Jonathan A.L.Gelfond3, Duangjai Piwksam1,4, Brad H.Pollock3, Naomi J.Winick5, Anderson B.Collier III5,6 and Gail E.Tomlinson1,2

1Greehey Children’s Cancer Research Institute, 2Department of Pediatrics, and 3Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA, 4Department of Pathology, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand, 5Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX, USA and 6Department of Pediatrics, Lehigh Valley Health Network, Bethlehem, PA, USA

To whom correspondence should be addressed. Tel: +210 562 9155; Fax: +210 562 9135; Email: beuten@uthscsa.edu

Introduction

Acute lymphoblastic leukemia (ALL) is the most common childhood cancer with a peak prevalence between the ages of 2 and 5 years (1). Acute leukemias may be of B- or T-lymphoid lineage and are characterized by recurring chromosomal aberrations and gene mutations often within members of key cellular pathways including lymphoid development, cell cycle regulation and tumor suppression, regulation of apoptosis, cell signaling and drug responsiveness (2,3). Although considered as important initiating events in leukemogenesis, these alterations alone usually do not cause leukemia (4,5), implicating that there are other etiologic factors. A weak increased familial risk of ALL has been shown suggesting that inherited genetic factors influence the risk of developing ALL (6). In addition, three recent genome-wide association (GWA) studies have revealed that genetic variation likely influences risk of ALL (7–9). Furthermore, genetic associations with ALL risk and outcome have been identified with ALL during the study period between 1990 and 2005 at the Children’s Medical Center of Dallas (n = 161) and from a pilot study project between 2007 and 2009 at the University of Texas Health Science Center at San Antonio (n = 2). A total of 251 age-matched unrelated Caucasian control samples were obtained from surplus deidentified blood samples from children after emergency room treatment for diagnoses other than cancer at the laboratory services of Children’s Medical Center. Exclusion criteria for the controls consisted of any history of malignancy as confirmed by review of medical records. Approval from Institutional Review Board at both universities was attained and the parents of each child signed consent to participate in genetic studies. Ethnicity was self-reported as determined by parental report. The average age of cases was 6.5 years and for controls 10.5 years, with an age range between 1 month and 21 years in all samples except for three controls that were older. The case samples consisted of 92 males and 71 females, and controls included 131 males and 120 females. All samples underwent standard DNA preparation procedures and DNA was stored at 4°C prior to genotyping. Of the 163 cases, 138 were diagnosed with precursor B-cell leukemia, 16 were T-cell leukemia and 9 were other B-cell subtypes.

Gene selection

To select and prioritize highly potential candidate genes, we performed an extended literature search and used annotated pathway databases and available programs including Endoavour (http://homes.esat.kuleuven.be/~bioiuser/endoeavour/index.php), Pathway Studio (Ariadne, Rockville, MD), Cytoscape (http://www.cytoscape.org/) and SNPs3D (http://www.snp3d.org/).

SNP selection, genotyping and quality control

A total of 672 SNPs (supplementary Table 1 is available at Carcinogenesis Online) within the 29 candidate genes were selected from the dbSNP database Build 129 (http://www.ncbi.nlm.nih.gov/projects/SNP/) and SNPper (http://snpper.chip.org/) using the following criteria: (i) within each gene, SNPs with a minor allele frequency > 0.05 with potential functional effects retrieved from programs predicting functional annotation and (ii) SNPs associated with leukemia, and in particular childhood ALL, as derived from literature were chosen. After this initial selection, tagging SNPs within each gene were identified using Haplovie (http://www.broad.mit.edu/mpg/haplovie/) with the following criteria: (i) inclusion of presellected SNPs (see above), (ii) minor allele frequency > 0.05 in order to gain more statistical power, (iii) r² threshold of 0.8, (iv) a log of odds (LOD) threshold for multi-marker testing of 3.0, (v) a minimum distance between tags of 60 bp and (vi) the 2- and 3-marker haplotype tagging option. For each gene, the search for tagging SNPs extended to a 10 kb region surrounding the gene. Genotyping was performed with the Golden Gate Assay of the VeraCode technology using the BeadXpress reader System according to the manufacturer’s protocol (Illumina, San Diego, CA). Duplicate samples were included as quality control.

Abbreviations: ALL, acute lymphoblastic leukemia; GWA, genome-wide association; OR, odds ratio; SNP, single-nucleotide polymorphisms; TFBS, transcription factor binding sites.

© The Author 2011. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Advance Access publication May 21, 2011

Carcinogenesis vol.32 no.9 pp.1349–1353, 2011

doI:10.1093/carcin/bgr091

Downloaded from https://academic.oup.com/carcin/article-abstract/32/9/1349/2391449 on 16 February 2018
controls and >99.8% concordant results were obtained for all SNPs. An average 96% of the samples were successfully genotyped per SNP.

Statistical analysis
We used Haplovlew 4.2 to measure linkage disequilibrium (LD) between the SNPs. Allele frequencies among the cases and controls were computed using the chi-square test. The odds ratio (OR) and its 95% confidence interval were estimated by unconditional logistic regression using R statistical software version 2.9.1 (http://www.r-project.org/). All the tests were two sided and the unadjusted \( \alpha \) was considered 0.05. To test for the independent association of a significant SNP while adjusting for other SNPs, we used a generalized linear model function from the R statistical package for which each SNP selected is entered into a single multivariable logistic regression model. SNPs in the model were taken to have additive effects. To calculate ORs of the haplotypes, we used the method implemented in the haplo.ccs package in R (15). Only major haplotypes, with an estimated frequency of >5%, are considered in this report. The OR of each major haplotype was computed relative to a reference group consisting of all other haplotypes including rare haplotypes.

To correct for multiple testing in the single SNP analysis, we used the single-nucleotide spectral decomposition method that accounts for LD among SNPs [http://gump.qimr.edu.au/general/dale/N/5PspDj/ (16)]. This method showed that 390 SNPs were independent, resulting in an experiment-wide significance threshold for \( \alpha = 0.05 \text{ and } P < 0.0001 \). The Bonferroni correction was used for multiple comparisons in haplotype analysis, where the adjusted \( P \) value is given by \( P \) at the 0.05 significance level divided by the number of major (frequency > 5%) haplotypes.

Results
Genes and allele frequencies
Polymorphisms of 29 genes with known or postulated roles in normal lymphoid development and/or lymphoid leukemogenesis and/or that commonly dysregulate hematopoietic transcription factors and tyrosine kinases were investigated. Genes studied in this report are genes shown previously to affect leukemogenesis, including fusion genes of translocations (BCR, ABL1, ET1V, FBX6, LMO1, LLI, EP300, CREBBP, MLL, JAK2, RUNX1, TCF3). DNA repair genes (MSH2, CHEK2, ATM, CCND1, MLH1, MSH2, TOP2A, CDKN1B), genes encoding transcription factor binding sites (TFBS) in DNA sequences (http://www.cbi.unc.edu/cgi-bin/tess/tess). Two potential interesting TFBS, for the PU.1 and for the PU.1 class 2 homeobox 2 (POU2F2) transcription factors, respectively, are located <1.5 kb downstream of rs442264 and show a high log likelihood score of 20 (Figure 2).

Quality control analysis
Quantile–quantile plots of the negative logarithm of all \( P \) values did not show significant differences between unadjusted and European ancestry-adjusted outcomes indicating that hidden population structure from differences between genotype calling of cases and controls is unlikely (supplementary data are available at Carcinogen-esis Online). Deviation from the null distribution was only observed at the upper tail for SNPs within the LIM domain only 1 (LMO1) gene (supplementary Figure 1 is available at Carcinogen-esis Online).

Haplotypes
Multiple SNPs in LMO1 were significantly associated with ALL risk at the \( P < 0.05 \) level (Table I and Figure 2) and we next determined whether joint effects of these SNPs had significant risk associations in our sample. Although perfect LD (\( D' = 1 \)) is observed between several pairs of SNPs, the \( r^2 \) is low (<0.4), and we therefore used all 14 SNPs with significant risk associations within LMO1 in haplotype analysis (rs11041815 is downstream from the gene and was not included). A major haplotype (31%) A-G-G-A-G-A-C-A-C-C-A-C for the combination of the 14 SNPs with significant risk associations within the gene showed a significant increase in risk for ALL (OR = 1.79, 95% confidence interval = 1.29–2.49, \( P = 0.0006 \), Table II), indicating that the boundaries of the susceptibility region could be delineated to a 20 kb region at the 5′ end of the LMO1 gene. Except for rs2071458, located in exon 1 and 215 bp upstream from the transcription start site, all other significant SNPs from the haplotype are within intronic regions of the gene.

Single SNP analysis
Logistic regression analysis was performed for all SNPs in the context of recessive, dominant and additive models. Since the underlying genetic model for ALL is unknown, we chose to present the findings considering the additive mode of inheritance in this report. Fifty SNPs in 15 genes were significantly associated with ALL risk at the two-sided \( P < 0.05 \) level (supplementary Table 1 is available at Carcinogen-esis Online, Figure 1). There was a moderate increase (~2.03) or decrease (~0.41) in ORs associated with these significant results. We used single nucleotide spectral decomposition (16) to determine a threshold for significance at this more stringent level (\( P < 0.0001 \)) that accounts for multiple tests and rs442264 (LMO1) remained significant (OR = 1.90, \( P = 3 \times 10^{-2} \)). A stratified analysis by subtype further showed that risk associations of LMO1 variants were significant in children with B-cell leukemia and rs442264 remained significant after correction for multiple testing in this subgroup (OR = 1.98, \( P = 3 \times 10^{-3} \), Table I), suggesting a potential role of this gene in the development of precursor B-lineage leukemias.

The effects of rs11041815, rs3799012, rs11041830 and rs11041831 on ALL remained when each is adjusted for the other significant SNPs by multiple logistic regression (\( P = 0.012–0.004 \), supplementary Table 4 is available at Carcinogen-esis Online). Except for rs3799012 and rs11041830 or rs11041831 that showed high LD (\( D' = 0.936, r^2 = 0.5 \) and \( D' = 1, r^2 = 0.36 \) in our control sample), the correlation between rs11041815 and any of the other three markers or between rs11041830 and rs11041831 is poor (\( D' < 0.5, r^2 < 0.04 \) in our control sample, supplementary Table 5 is available at Carcinogen-esis Online).

We further explored the region around rs442264 using the TESS web tool for predicting transcription factor binding sites (TFBS) in DNA sequences (http://www.cbi.unc.edu/cgi-bin/tess/tess). Two potentially interesting TFBS, for the PU.1 and for the PU.1 class 2 homeobox 2 (POU2F2) transcription factors, respectively, are located <1.5 kb downstream of rs442264 and show a high log likelihood score of 20 (Figure 2).

Fig. I. Association results from 612 SNPs using the additive model. For each SNP at the position (kb) shown on the x-axis, \(-\log_{10} P \) from the allelic association test is indicated on the y-axis. Dashed horizontal line represents significance threshold after adjustment for multiple comparisons (\( P < 0.0001 \)).
Table I. Significant risk associations of single SNPs in \textit{LMO1}

<table>
<thead>
<tr>
<th>Marker</th>
<th>Controls, no. of risk alleles</th>
<th>All cases Cases, no. of risk alleles</th>
<th>OR (95% CI)</th>
<th>(P)</th>
<th>Pre-B Cases, no. of risk alleles</th>
<th>OR (95% CI)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11041815</td>
<td>47</td>
<td>52</td>
<td>1.96 (1.28–3.01)</td>
<td>0.002</td>
<td>45</td>
<td>2.02 (1.49–3.16)</td>
<td>0.002</td>
</tr>
<tr>
<td>rs442264</td>
<td>212</td>
<td>181</td>
<td>1.90 (1.41–2.56)</td>
<td>0.00003*</td>
<td>156</td>
<td>1.98 (1.44–2.73)</td>
<td>0.00003*</td>
</tr>
<tr>
<td>rs7394015</td>
<td>201</td>
<td>98</td>
<td>0.66 (0.48–0.89)</td>
<td>0.008</td>
<td>87</td>
<td>0.70 (0.51–0.97)</td>
<td>0.033</td>
</tr>
<tr>
<td>rs379769</td>
<td>203</td>
<td>100</td>
<td>0.66 (0.49–0.89)</td>
<td>0.007</td>
<td>83</td>
<td>0.63 (0.46–0.86)</td>
<td>0.004</td>
</tr>
<tr>
<td>rs765813</td>
<td>25</td>
<td>6</td>
<td>0.41 (0.17–0.97)</td>
<td>0.042</td>
<td>4</td>
<td>0.33 (0.12–0.92)</td>
<td>0.033</td>
</tr>
<tr>
<td>rs127710</td>
<td>158</td>
<td>77</td>
<td>0.70 (0.51–0.97)</td>
<td>0.030</td>
<td>66</td>
<td>0.72 (0.51–1.01)</td>
<td>0.054</td>
</tr>
<tr>
<td>rs3794012</td>
<td>185</td>
<td>152</td>
<td>1.62 (1.20–2.19)</td>
<td>0.001</td>
<td>131</td>
<td>1.70 (1.24–2.34)</td>
<td>0.001</td>
</tr>
<tr>
<td>rs7094134</td>
<td>55</td>
<td>20</td>
<td>0.56 (0.33–0.94)</td>
<td>0.029</td>
<td>17</td>
<td>0.56 (0.32–0.98)</td>
<td>0.041</td>
</tr>
<tr>
<td>rs4237770</td>
<td>217</td>
<td>157</td>
<td>1.41 (1.06–1.87)</td>
<td>0.020</td>
<td>138</td>
<td>1.50 (1.11–2.03)</td>
<td>0.009</td>
</tr>
<tr>
<td>rs11041830</td>
<td>247</td>
<td>124</td>
<td>0.69 (0.52–0.92)</td>
<td>0.010</td>
<td>101</td>
<td>0.65 (0.48–0.88)</td>
<td>0.005</td>
</tr>
<tr>
<td>rs204938</td>
<td>238</td>
<td>120</td>
<td>0.64 (0.47–0.86)</td>
<td>0.003</td>
<td>105</td>
<td>0.67 (0.48–0.92)</td>
<td>0.012</td>
</tr>
<tr>
<td>rs204957</td>
<td>230</td>
<td>108</td>
<td>0.66 (0.48–0.89)</td>
<td>0.007</td>
<td>93</td>
<td>0.67 (0.49–0.93)</td>
<td>0.018</td>
</tr>
<tr>
<td>rs11041831</td>
<td>189</td>
<td>84</td>
<td>0.58 (0.42–0.80)</td>
<td>0.001</td>
<td>69</td>
<td>0.55 (0.39–0.78)</td>
<td>0.001</td>
</tr>
<tr>
<td>rs11041833</td>
<td>182</td>
<td>141</td>
<td>1.52 (1.12–2.05)</td>
<td>0.007</td>
<td>125</td>
<td>1.65 (1.20–2.27)</td>
<td>0.002</td>
</tr>
<tr>
<td>rs2071458</td>
<td>139</td>
<td>65</td>
<td>0.68 (0.49–0.95)</td>
<td>0.023</td>
<td>50</td>
<td>0.58 (0.40–0.84)</td>
<td>0.004</td>
</tr>
</tbody>
</table>

SNPs in italics were used for haplotype analysis. CI, confidence interval; OR, odds ratio.

*Significant results after multiple correction are in bold (\(P < 1 \times 10^{-4}\)).

**Discussion**

Large-scale candidate gene approaches allow for a specific and focused search for biomarkers that play a role in the risk for childhood ALL. To estimate ALL risk conferred by individual SNPs and/or SNP-SNP interactions, we studied 29 candidate genes known to be involved in the biology of the disease. We report on the association between ALL susceptibility and 612 tagged SNPs, fully covering each of the genes and enriched with functional and potential deleterious SNPs. A total of 414 samples, 163 cases and 251 healthy controls of Caucasian descent were genotyped.

The major finding was obtained for variants within the \textit{LMO1} gene. \textit{LMO1} encodes a cysteine-rich two LIM domain transcriptional regulator and is located on chromosome 11 in an area of consistent chromosomal translocation in T-cell leukemia (18). The gene is abundantly expressed in several tissues, including bone marrow [TIGER database, http://bioinfo.wilmer.jhu.edu/tiger/, (19)], where it plays a role in protein–protein interaction (20). Deregulation of LIM protein expression has been found to contribute to T-cell oncogenesis (18,21) and 45% of T-ALL cases, with or without chromosomal translocations, show aberrant T-cell-specific expression of \textit{LMO1} (22). Of interest is that rs379951 that showed an association in the Korean GWA study in childhood ALL (7) is tagged by rs442264 (\(D^2 = 0.9, r^2 = 0.503\) in Europeans), and our results therefore support the findings in Korean children with ALL. On the other hand, \textit{LMO1} has not been reported previously in GWA studies of ALL susceptibility in Caucasians and to our knowledge, our findings are the first to show evidence that genetic variation of this gene is involved in ALL in Caucasian children. Of note is that rs442264 is not present on the chip panels used in previous GWA studies of leukemia in European children (8,9).

Two potentially interesting TFBS, located >1.5 kb downstream of rs442264, were predicted by the web tool TESS (Figure 2). The transcription factor PU.1, a hematopoietic-specific member of the erythroid differentiation family of transcription factors, is expressed in macrophages and B cells (23) and plays an essential role in granulocytic, monocytic and lymphoid development (24) and as such has overlapping functions with other DNA-binding proteins like Ikaros (25). Mice with germ line mutations in the gene encoding PU.1 (\textit{Sfpi1}) have severely impaired myeloid and lymphoid differentiation and die shortly after birth (26,27). The second predicted TFBS close to rs442264 is for the POU2F2 transcription factor, a lymphoid-specific transcription factor, which has been shown to concomitantly act with PU.1 to activate immunoglobulin promoters and regulate immunoglobulin expression (28,29). One could hypothesize that a disruption of one or both transcription binding site activities, impairing the gene’s normal function, could be (indirectly) implicated in aberrant B-cell development, which might result in leukemogenesis. Further studies are necessary to identify the functional effect of these allelic variations and their contribution to leukemogenesis.

Our study further showed that upon stratification on subtype, risk associations of \textit{LMO1} variants were significant in children with precursor B-cell leukemia. Moreover, the effect size becomes stronger when examining the precursor B-ALL subset, indicating a potential role in the development of B-cell leukemia separate from the previously reported association with T-cell ALL, which could not be confirmed due to the small sample size (18,21,22). Although the

---

**Fig. 2.** Outline of the \textit{LMO1} gene. SNPs analyzed and LD structure. (a) Exons are shown as black boxes, dashed lines indicate predicted promoter. Vertical arrows denote potential interesting transcription factor binding sites. TSS, translation start site. (b) SNPs analyzed, their physical position based on dbSNP database Build 129 of the human genome and \(P\) values of single SNP analysis. Asterisks indicate significant SNPs from the study by Wang et al. (17). SNPs present on Illumina Infinium HD Human370 Duo BeadChips (8), Illumina 550K (17) and Affymetrix 550K (9) are in bold, underlined and in italics, respectively (c) Major haplotypes for SNPs with risk associations, with the risk haplotype denoted in black. (d) Pair-wise LD between SNPs in our sample presented as \(D^2\) values. The color intensity represents the strength of LD according to the standard Haplovio scheme for \(D^2\) (black indicating values > 0.90, shades of gray 0.90 > \(D^2\) > 0.00, white indicate a value of 0.0).
susceptibility loci could be in LD with nearby SNPs/gene and thus merely markers for the true causal factor, the finding of an associated haplotype may suggest that the true causal variant(s) is (are) likely to reside on the haplotype.

Recently, *LMO1* was identified as neuroblastoma oncogene (17). Four SNPs reached genome-wide significance (rs4758051, rs10840002, rs10420, rs204938), three of which we also genotyped in our study (Figure 2). Aside from rs204938, significant in both our and their study, the remaining three are located outside the boundaries of the risk haplotype we found. Although neuroblastoma and ALL are clinically different entities, it remains to be determined whether *LMO1* triggers tumorigenesis through a common mechanism in both pediatric malignancies.

It is noteworthy that we observed a trend toward significance for seven variants within the Ikaros family zinc finger protein 1 (*IKZF1*) gene (supplementary Table 3 is available at Carcinogenesis Online), previously shown to play a role in ALL from GWA analysis (8,9). However, we were not able to replicate the significant findings for rs4132601 but rather found other SNPs to be marginally significant in this region that did not reach genome-wide significance in the reported GWA studies. The most significant finding in our study was for rs6959427 (OR = 2.03, P = 0.003), located just downstream from *IKZF1* and upstream from the dopa decarboxylase aromatic l-amino acid (DDC) gene. This SNP is not present on the SNP chip panels used in the GWA studies reported previously. However, statistical significance did not remain after correction for multiple testing implicating that a role of this gene in leukemogenesis is not corroborated by our findings.

We selected 29 candidate genes known to be involved in the biology of ALL based on biochemical and physiological studies. Because the exact mechanism by which these genes might act in ALL development and progression is largely unknown, the genes investigated may represent only a fraction of all relevant genes. Furthermore, the SNPs selected in the study are not an exhaustive list of all existing common SNPs and may not fully represent the genetic variability of these genes. However, a careful selection of genes and SNPs was made based on published data of functional and/or association analyses related to ALL, and thus, high-priority genes and SNPs were included in the analysis. We only evaluated ALL as a single trait without considering interactions with environmental and other non-genetic factors that may be involved in leukemogenesis, which is a limitation of studies examining only genetic variants. In addition, no gender-specific associations were tested for due to our limited sample size and overestimation of the data which would likely result in uncertainty of the risk estimates. The minimal detectable OR, corresponding to 80% power, is 1.55 considering a median minor allele frequency of 24.6% for our Caucasian sample of 163 cases and 251 controls. Replication of our findings in independent Caucasian samples is warranted to elucidate the role of these variants in ALL susceptibility.

In summary, our results indicate that genetic alterations play a significant role in the risk of ALL. Our approach of evaluating a large number of genetic variants in preselected candidate genes identified a new susceptibility locus at 11p15 for ALL. Further studies are warranted to explore whether our findings can be replicated not only in other European samples but also in other populations of different ethnic background.

### Supplementary material

Supplementary Tables 1–5 and Figure 1 can be found at http://carcin.oxfordjournals.org/

### Funding

Hyundai Hope on Wheels [44096 to J.B. and G.E.T.]; the Texas Higher Education Coordinating Board [010019-0105-2001 to G.E.T.]; the Children’s Cancer Fund Dallas to G.E.T.; the Greerhey Family Foundation to G.E.T. and the Commission on Higher Education Scholarship, Thailand, to D.P.

### Acknowledgements

We thank the family members for their participation and clinical staff for their skilled assistance. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Conflict of Interest Statement

None declared.

### References


### Table II. Association of common haplotypes from significant SNPs within LMO1 with ALL risk

<table>
<thead>
<tr>
<th>SNP combination</th>
<th>Frequency (%)</th>
<th>No. of haplotypes</th>
<th>OR</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-A-G-G-C-A-A-A-G-G-C-G-C</td>
<td>17</td>
<td>41</td>
<td>86</td>
<td>0.83</td>
<td>0.55–1.24</td>
</tr>
</tbody>
</table>

*Significant results after Bonferroni correction are in bold (P < 0.01). Only common haplotypes (>5%) are shown.*


Received February 25, 2011; revised April 26, 2011; accepted May 6, 2011