Genetic Polymorphisms in Adaptive Immunity Genes and Childhood Acute Lymphoblastic Leukemia

Jeffrey S. Chang1,4,5, Joseph L. Wiemels1, Anand P. Chokkalingam4, Catherine Metayer4, Lisa F. Barcellos4, Helen M. Hansen2, Melinda C. Aldrich3, Neela Guha4,6, Kevin Y. Urayama4,6, Ghislaine Scélo6, Janet Green4, Suzanne L. May4, Vincent A. Kiley7, John K. Wiencke2, and Patricia A. Buffler4

Abstract

Background: Childhood acute lymphoblastic leukemia (ALL) has been hypothesized to have an infection- and immune-related etiology. The lack of immune priming in early childhood may result in abnormal immune responses to infections later in life and increase ALL risk.

Methods: The current analyses examined the association between childhood ALL and 208 single-nucleotide polymorphisms (SNP) of 29 adaptive immune function genes among 377 ALL cases and 448 healthy controls. Single SNPs were analyzed with a log-additive approach using logistic regression models adjusted for sex, age, Hispanic ethnicity, and race. Sliding window haplotype analyses were done with haplotypes consisting of 2 to 6 SNPs.

Results: Of the 208 SNPs, only rs583911 of IL12A, which encodes a critical modulator of T-cell development, remained significant after accounting for multiple testing (odds ratio for each copy of the variant G allele, 1.52; 95% confidence interval, 1.25-1.85; \( P = 2.9 \times 10^{-5} \)). This increased risk was stronger among firstborn children of all ethnicities and among non-Hispanic children with less day care attendance, consistent with the hypothesis about the role of early immune modulation in the development of childhood ALL. Haplotype analyses identified additional regions of CD28, FCGR2, GATA3, IL2RA, STAT4, and STAT6 associated with childhood ALL.

Conclusion: Polymorphisms of genes on the adaptive immunity pathway are associated with childhood ALL risk.

Impact: Results of this study support an immune-related etiology of childhood ALL. Further confirmation is required to detect functional variants in the significant genomic regions identified in this study, in particular for IL12A. Cancer Epidemiol Biomarkers Prev; 19(9): 2152–63. ©2010 AACR.

Introduction

The etiology of childhood acute lymphoblastic leukemia (ALL) is likely to be affected by history of infections and immune development as suggested by Greaves’ “delayed infection” hypothesis (1) and Kinlen’s “population mixing” hypothesis (2). Although these two hypotheses differ on the existence of specific leukemia-causing agents, both suggest that the lack of immune priming in a child’s early development may result in abnormal immune responses to microbial challenges later in life thereby increasing the risk of childhood ALL. Both hypotheses are similar to the “hygiene hypothesis” proposed by Strachan to explain the increasing prevalence of allergies in the western population (3). Many studies have been conducted to test these two hypotheses using proxy measures of infection (4-11). An increased risk of childhood ALL has been observed with proxy measures of early childhood infections such as low birth order (4, 5) and low day care attendance (6-9, 12), although negative findings have also been reported (10, 11). Studies examining associations between reports of specific early childhood infections and childhood ALL have yielded variable findings ranging from inverse association (8, 13, 14) to no association (7, 15-17) to positive association (18). These inconsistencies may be due to difficulty in obtaining information about asymptomatic infections, recall bias, variable sample sizes, and differences in questionnaire design and data collection. Inverse associations have been reported by the majority of studies.

Authors' Affiliations: Departments of 1Epidemiology and Biostatistics, 2Neurological Surgery, and 3Medicine, University of California, San Francisco, San Francisco, California; 4School of Public Health, University of California, Berkeley, Berkeley, California; 5National Institute of Cancer Research, National Health Research Institutes, Tainan, Taiwan R.O.C.; 6ARC, Lyon, France; and 7Pediatric Oncology, Kaiser Permanente Roseville, Roseville, California

Note: Supplementary data for this article is available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Author: Jeffrey S. Chang, National Health Research Institutes, 1F, No. 367, Sheng-Li Road, Tainan 70456, Taiwan R.O.C.; Phone: 886-6-208-3422 ext. 65260; Fax: 886-6-208-3427. E-mail: jeffreychang@nhri.org.tw

doi: 10.1158/1055-9965.EPI-10-0389

©2010 American Association for Cancer Research.
of allergies and childhood leukemia (13, 19-25), indicating that immune function may play an important role.

Although the majority of studies support an infection- and immune-related etiology of childhood ALL, little is known about the underlying role of genetics. The development of immune function is a complex process that involves the interplay between many cell types including Th1, Th2, T regulatory, and Th17 cells (26). Variations in the genes affecting the development and the function of these cell types may affect a child’s immune responses and thus his/her risk of childhood ALL. The current analyses examine the association between childhood ALL and 208 polymorphisms of 29 adaptive immune function genes involved in the development and the function of Th1, Th2, T regulatory, and Th17 cells. In addition, analysis was done to assess the interaction between immune function genes and two proxy measures of early childhood infections (early day care attendance and birth order) on the risk of childhood ALL.

Materials and Methods

Study subjects

The study subjects were recruited by the Northern California Childhood Leukemia Study (NCCLS), a case-control study that began in 1995. Major medical centers in 17 counties in the San Francisco Bay Area were included in the study from 1995 to 1999, and 18 additional counties in the California Central Valley were added in 1999. The eligibility criteria for case and control subjects were (a) being a resident of the study area; (b) being less than 15 years old at the time of the leukemia diagnosis (reference date for controls); (c) having at least one English or Spanish speaking parent; and (d) having no previous diagnosis of cancer. Cases were identified from four (1995-1999) and later nine hospitals (1999-2008) in the study area. Comparison of case ascertainment in the 35-county study area to the California Cancer Registry data (1997-2003) showed that the NCCLS ascertained 93% to 96% of children diagnosed with leukemia in the participating hospitals. When considering both participating and nonparticipating hospitals within the 35 study counties, cases ascertainment through the NCCLS protocol represented 76% of all the diagnosed cases, making the study approximately population based. Controls were randomly selected from birth certificates through the California Office of Vital Records and individually matched to cases on birth date, gender, maternal race, and Hispanic ethnicity and were shown to be representative of the source population of the cases (27). The current genetic study began with 928 NCCLS subjects recruited during 1995-2002 (464 cases, 464 controls), with 825 subjects retained (377 childhood ALL cases, 448 healthy controls) for this analysis. Reasons for exclusion were insufficient DNA for genotyping from either buccal cytobrush swabs or archived newborn dried bloodspot specimens (n = 21), ineligibility after genotyping (respondent was not a biological parent; n = 1), or Illumina single-nucleotide polymorphism (SNP) call rates <95% (n = 22). Acute myelogenous leukemia cases (n = 59) were excluded from the analysis because ALL was the primary childhood leukemia subtype that has been hypothesized to have an infection- and immune-related etiology (1). The study included both Hispanic and non-Hispanic subjects. A child was considered Hispanic if either parent self-reported Hispanic ethnicity (156 cases, 179 controls). The non-Hispanic group (221 cases, 269 controls) consisted of 73.5% whites, 11.8% Asians, 5.5% blacks, 0.4% Native Americans, and 8.8% others.

The study was approved by the Committee for Protection of Human Subjects of the University of California, Berkeley and by the Institutional Review Boards of all collaborating institutions.

Biospecimen collection and DNA processing

Buccal cytobrushes were collected at the time of interview by trained interviewers as the primary DNA source for case and control children. DNA from cytobrush samples was extracted by heating (98-100°C) in the presence of NaOH, followed by neutralization with Tris-HCl buffer and whole-genome amplification (WGA) using GenoPlex reagents (Sigma Aldrich). Archived newborn blood (ANB) specimens, which are collected at birth on a paper card for each child born in California and archived at −20°C by the California Department of Public Health, were used as a secondary DNA source when buccal cell DNA was insufficient for genotyping (26.6% of subjects). For each child, the NCCLS receives one spot of ANB containing approximately 60 μL of blood. A small piece of the bloodspot was excised for DNA extraction using the QIAamp DNA mini-extraction kit. Isolated ANB DNA was whole-genome amplified using REPLI-g reagents (Qiagen). WGA products were tested for minimum acceptable amplifiable human DNA content using an ALUq real-time PCR method published elsewhere (28). WGA DNA from both buccal cells and ANB specimens and genomic DNA from peripheral blood produced genotypes that were highly concordant when analyzed using multiplexed GoldenGate genotyping (Illumina; refs. 28, 29).

SNP selection

We focused on 29 adaptive immunity genes (Th1: IL12A, IL12B, IL12RB1, IL12RB2, PHF11, STAT4; Th2/allergy: ADAM33, GATA3, IL4, IL4R, MS4A2, STAT6; T regulatory cells: CD28, CD80, CTLA4, IL2, IL2RA, IL6, IL10, STAT5A, STAT5B, TGFβ1; Th17: STAT3; B-cell: CD40, FCGR2A, MME; others: NFKB1, NFKBIA, NFKBIB) based on assessment of their importance in published literature. SNPs genotyped by the International HapMap Project (30) were selected, including 10-kb regions upstream and downstream to include variants in potential regulatory elements. Haplovie (31) was used to construct haplotype blocks according to the block definition by
Gabriel et al. (32) to select haplotype-tagging SNPs. Additional SNPs specific to the Hispanic population were supplemented from the SNP500Cancer database (33). Additionally, SNPs reported previously in the literature for these genes were also included. In total, 244 SNPs in the 29 adaptive immunity genes were selected for genotyping.

Genotyping
The genotyping of 244 SNPs of the 29 adaptive immunity genes was done on whole-genome amplified DNA using a custom Illumina GoldenGate assay panel. SNPs were excluded from statistical analyses if they had a call rate of <90% (29 SNPs), had minor allele frequencies <5% in both Hispanics and non-Hispanics (6 SNPs), or failed Hardy-Weinberg equilibrium \( (P < 0.01) \) in both Hispanic and non-Hispanic controls (1 SNP), leaving a total of 208 SNPs for analysis. Ninety-five ancestry informative markers (AIM) were included to account for potential population stratification in our study population, but only 80 AIMs passed quality control.

Additional quality control of genotyping was done by comparing duplicate samples: (a) 59 samples were run in duplicate after processing with the same WGA method and genotyped on the same plate; these showed a 99.1% concordance of genotype; (b) DNA specimens extracted from both buccal cell and archived newborn dried bloodspots were genotyped for 9 subjects; these showed a 98.9% concordance of genotype; and (c) the Mendelian errors (the child does not have the expected genotype compared with the parents) were estimated in 10 trios from the HapMap Centre d'Etude du Polymorphisme Humain and only 28 pedigree errors in 25 markers were found (overall Mendelian error rate, 0.20%). For subjects that had samples run in duplicate for quality control purpose, one sample from each pair of duplicates was chosen on the basis of SNP call rates. If one of the replicates had a higher call rate than the other (higher number of SNPs successfully genotyped), then its data were retained in the analysis, while the data of the other replicate were discarded.

Proxy measures of early childhood exposure to infections
Information on day care attendance and birth order was collected through an in-person interview with the biological parent (usually the mother) of the child. Detailed information on collection and calculation of day care attendance was presented previously (7). Briefly, data on day care and preschool attendance before the date of case

<table>
<thead>
<tr>
<th>Table 1. Demographic characteristics, genetic ancestry, birth order, and day care attendance of Hispanic and non-Hispanic subjects, NCCLS, 1995-2002</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
</tr>
<tr>
<td><strong>Case, n (%)</strong></td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td><strong>Age</strong></td>
</tr>
<tr>
<td>Mean age, y (SE)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black</td>
</tr>
<tr>
<td>Native Americans</td>
</tr>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>Others</td>
</tr>
<tr>
<td><strong>Genetic ancestry</strong></td>
</tr>
<tr>
<td>European, % (SE)</td>
</tr>
<tr>
<td>African, % (SE)</td>
</tr>
<tr>
<td>Amerindian, % (SE)</td>
</tr>
<tr>
<td><strong>Birth order</strong></td>
</tr>
<tr>
<td>Firstborn</td>
</tr>
<tr>
<td>Having at least one older sibling</td>
</tr>
<tr>
<td><strong>Day care attendance</strong></td>
</tr>
<tr>
<td>Thousand child hours before 6 mo of age (SE)</td>
</tr>
<tr>
<td>Thousand child hours before 1 y of age (SE)</td>
</tr>
</tbody>
</table>

*P values were generated using \( \chi^2 \) tests for categorical variables (sex and race) and \( t \) tests for continuous variables (age and genetic ancestry).
The 208 adaptive immunity SNPs.

Tests of heterogeneity by Hispanic ethnicity; otherwise, all subjects were combined to assess SNP-disease association.

Multifactor dimensionality reduction (MDR) analysis (34) was done to assess SNP-SNP interactions between the 6 SNPs that were significantly (P < 0.05) associated with ALL among all subjects. For the current MDR analysis, we allowed for combinations of 1 to 4 SNPs. The 10-fold cross-validation was repeated 10 times using 10 different random seeds to reduce the probability of spurious findings due to chance division of the data. P values were calculated by permutation testing with 1,000 permutations. The best combination of SNPs was determined based on cross-validation consistency and testing accuracy.

To assess the influence of adaptive immunity genes on the association between early life exposure to infections and childhood ALL, gene-environment interaction analysis was done with the one SNP (IL12A rs583911) that passed the multiple testing adjustment. The risk associated with the minor variant G allele was evaluated by stratifying on the two proxy measures of early exposure to common infections (total child-hours of day care attendance and birth order). There were 365 ALL cases and 429 controls retained in the gene-environment interaction analyses after excluding subjects under the age of 1 year to allow for sufficient exposure to infectious factors before the development of childhood leukemia. Product terms for interaction between infectious exposure variables and rs583911 were included in the statistical model, and statistical significance was assessed by the log-likelihood ratio test comparing the full model containing the product terms to the sub-model without the product terms. Interaction between rs583911 and day care attendance was evaluated separately for Hispanics and non-Hispanics because the association between day care attendance and childhood ALL differed significantly by Hispanic ethnicity (P < 0.05). Hispanics and non-Hispanics were combined for assessing interaction between rs583911 and birth order because the association between ALL and birth order did not differ by Hispanic ethnicity (P values ranged from 0.35 to 0.89).

Haplotype analyses were done using the haplo.stats R package (35) for each gene separately to capture the information potentially missed by the single SNP analyses by increasing the statistical power to tag causal variants and by accounting for cis-interactions between two or more SNPs (36). All subjects were combined for haplotype analysis if none of the SNPs in a gene had evidence of heterogeneity by Hispanic ethnicity; otherwise, separate analyses by Hispanic ethnicity were done. Haplotype analyses were done using the sliding window approach (haplotype windows of 2-6 SNPs) confined to the region of a single gene. Global P values were calculated for each haplotype window to evaluate whether...
the distribution of haplotypes was significantly different between cases and controls. Graphical representations of the sliding window results were constructed using GrASP (37). The most significant $P$ value (minimal $P$ value) for each SNP across all haplotype windows was used to determine the most significant genomic region for each gene. Analyses to examine specific haplotypes in significant genomic regions were done with haplotype trend regression to calculate the OR associated with each copy of a specific haplotype using the most frequent haplotype as the reference group.

Eighty AIMs were used to estimate genetic ancestry (percent of European, Amerindian, and African ancestry) using the methods described by Chakraborty and Weiss (38) and Hanis et al. (39). Genetic ancestry was included in statistical models to assess the effect of potential population stratification on the association between adaptive immunity SNPs and childhood ALL.

**Results**

Cases and controls were comparable in the distribution of sex, age, race, and genetic ancestry (Table 1). The quantile-quantile plot (Fig. 1) compares the distribution of the observed versus the expected $-\log_{10} P$ values (log-additive model not adjusted for genetic ancestry using data of combined race/ethnicity) of the 208 SNPs. All of the observed $-\log_{10} P$ values except the most significant one follow closely the 45-degree angle line expected under the null hypothesis of no association, indicating minimal evidence of an inflated test statistic associated with population stratification or some other systematic bias. In addition, sensitivity analyses including genetic ancestry in the statistical models did not change the ORs by more than 10%, suggesting minimal effect of population substructure, and therefore genetic ancestry was not included in the final analytic models.

### Table 2. Single SNP analysis of adaptive immunity genes with $P < 0.05$, NCCLS, 1995-2002

<table>
<thead>
<tr>
<th>Rs no.</th>
<th>Gene</th>
<th>Alleles*</th>
<th>Chromosomal location</th>
<th>Location</th>
<th>Tag or literature SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Th1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs583911</td>
<td>IL12A</td>
<td>A/G</td>
<td>3q25.33-q26</td>
<td>Intron 2</td>
<td>Tag</td>
</tr>
<tr>
<td>rs2243154</td>
<td>IL12A</td>
<td>A/G</td>
<td>2q32.2-2q32.3</td>
<td>Intron 3</td>
<td>Tag</td>
</tr>
<tr>
<td>rs17769459</td>
<td>STAT4</td>
<td>A/G</td>
<td>13q14.3</td>
<td>Intron 2</td>
<td>Tag</td>
</tr>
<tr>
<td>rs16833215</td>
<td>STAT4</td>
<td>A/G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2247119</td>
<td>PHF11</td>
<td>A/G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1925742</td>
<td>PHF11</td>
<td>A/G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3181224</td>
<td>IL12B</td>
<td>A/G</td>
<td>5q31.1-q33.1</td>
<td>3′ near gene</td>
<td>Tag</td>
</tr>
<tr>
<td>rs11574790</td>
<td>IL12B</td>
<td>A/G</td>
<td></td>
<td>Intron 6</td>
<td>Tag</td>
</tr>
<tr>
<td>rs1368439</td>
<td>IL12B</td>
<td>A/C</td>
<td></td>
<td>3′ UTR</td>
<td>Tag</td>
</tr>
<tr>
<td>Th2/Allergies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3781093</td>
<td>GATA3</td>
<td>A/G</td>
<td>10p15</td>
<td>Intron 3</td>
<td>Tag</td>
</tr>
<tr>
<td>rs1059513</td>
<td>STAT6</td>
<td>A/G</td>
<td>12q13</td>
<td>3′ UTR</td>
<td>Literature</td>
</tr>
<tr>
<td>Treg</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1554286</td>
<td>IL10</td>
<td>A/G</td>
<td>1q31-q32</td>
<td>Intron 3</td>
<td>Tag</td>
</tr>
<tr>
<td>rs3024505</td>
<td>IL10</td>
<td>A/G</td>
<td></td>
<td>3′ near gene</td>
<td>Tag</td>
</tr>
<tr>
<td>rs3024490</td>
<td>IL10</td>
<td>A/C</td>
<td></td>
<td>Intron 1</td>
<td>Tag</td>
</tr>
<tr>
<td>rs926169</td>
<td>CTLA4</td>
<td>A/G</td>
<td>2q33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs733618</td>
<td>CTLA4</td>
<td>A/G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs706778</td>
<td>IL2RA</td>
<td>A/G</td>
<td>10p15-p14</td>
<td>Intron 1</td>
<td>Tag</td>
</tr>
<tr>
<td>rs942201</td>
<td>IL2RA</td>
<td>A/C</td>
<td></td>
<td>Intron 1</td>
<td>Tag</td>
</tr>
<tr>
<td>B-cell</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1385477</td>
<td>MME</td>
<td>A/C</td>
<td>3q25.1-q25.2</td>
<td></td>
<td>Tag</td>
</tr>
</tbody>
</table>

(Continued on the following page)
accounting for multiple testing using the Bonferroni correction (significance threshold = 0.05/235 = 2.1 × 10^{-4}). No P values from the dominant model reached statistical significance after correcting for multiple testing (Supplementary Table S1). Supplementary Table S2 compares the results of the 19 statistically significant SNPs (P < 0.05) with and without adjustment for genetic ancestry, and the results were very similar. Single SNP analyses were further performed among three ALL subgroups with a sufficient subject number (n > 50): (a) ALL with TEL-AML translocation (n = 62); (b) ALL with high hyperdiploidy (n = 110); and (c) c-ALL (n = 189). Although statistical power decreased with smaller sample size in the subgroup analyses, rs583911 of IL12A remained statistically significant (P < 0.05) in all subgroups.

**Multifactor dimensionality reduction**

MDR analysis showed that rs583911 is the best predictor of case-control status with the highest testing accuracy (P = 0.05; Supplementary Table S3). No additional SNPs were able to improve the prediction accuracy, suggesting no evidence of SNP-SNP interactions.

**Interaction between the variant G allele of IL12A rs583911 and proxies for infectious exposures**

Cases and controls were comparable in the distribution of birth order (Table 1). For Hispanics, cases had more total child-hours of day care attendance before 6 months and 1 year of age (Table 1). In contrast, cases had less total child-hours of day care attendance before 6 months and 1 year of age for non-Hispanics. Among controls, non-Hispanics had more total child-hours of day care attendance before 6 months and 1 year of age compared with Hispanics.

The increased ALL risk associated with each copy of the variant G allele of IL12A rs583911 seemed stronger among firstborn children (OR, 2.14; 95% CI, 1.52-3.01) compared with children with older siblings (OR, 1.30; 95% CI, 1.00-1.69; Table 3).

Among non-Hispanics, those children who had less than 2,000 child-hours in day care before the age of...
6 months had an increased risk of ALL associated with each copy of the G allele of *IL12A* rs583911 (OR, 1.68; 95% CI, 1.27-2.22). This was not apparent among children with 2,000 child-hours or more of day care attendance. Similar results were seen for day care attendance before the age of 1 year using 5,000 child-hours as the cutoff. Among Hispanic children, the risk of ALL associated with each copy of the G allele of *IL12A* rs583911 did not differ significantly by child-hours of day care attendance.

### Haplotype analyses with all subjects combined

Because no evidence of heterogeneity was observed among all SNPs of each gene by Hispanic status, haplotype analyses using a sliding window approach were performed with all subjects combined for 16 of the 29 genes (Supplementary Fig. S1). Among these 16 genes, four (CD28, CTLA4, FCGR2A, and *IL12A*) showed significant haplotype associations (global *P* < 0.05). However, for CTLA4 and *IL12A*, haplotype analyses did not contribute additional information because the strength of association either did not improve or weakened with increasing haplotype window size compared with the results of single SNP analyses.

Haplotype analyses of CD28 and FCGR2A identified significant regions in the genes not observed with the single SNP analyses. For CD28, a region tagged by rs1879877, rs3181096, rs1181389, and rs3769683 showed the strongest significance (global *P* = 0.02), and the haplotype CGAA was associated with an increased risk of ALL (OR for each copy of the haplotype, 1.55; 95% CI, 1.03-2.34; *P* = 0.04; Table 4). For FCGR2A, a region tagged by rs10800309 and rs4656308 showed the strongest significance (global *P* = 0.02), and the haplotype AA was associated with an increased risk of ALL (OR for each copy of the haplotype, 1.46; 95% CI, 1.13-1.90; *P* = 0.004; Table 4).

### Haplotype analyses stratified by Hispanic ethnicity

Due to the presence of significant heterogeneity for at least one SNP in the gene by Hispanic ethnicity, haplotype analyses using a sliding window approach were done separately for Hispanic and non-Hispanic subjects for 13 of the 29 genes (Supplementary Fig. S2). Eleven of the 13 genes had significant haplotype regions among either the Hispanics or the non-Hispanics; however, only four (GATA3, IL2RA, STAT4, and STAT6) of the 11 genes had common regions shared by Hispanics and non-Hispanics in their association with childhood ALL.

For GATA3, a region tagged by rs4143094, rs3781093, and rs3802604 was significantly associated with childhood ALL (global *P* = 0.03 for non-Hispanics and 0.04 for Hispanics).

### Table 3. Interaction between birth order or day care attendance and *IL12A* rs583911 on the risk of childhood ALL, NCCLS, 1995-2002

<table>
<thead>
<tr>
<th>Birth order</th>
<th>n cases, AA/AG/GG</th>
<th>n controls, AA/AG/GG</th>
<th>OR for every one copy of G allele of rs583911*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firstborn</td>
<td>29/66/52</td>
<td>57/80/25</td>
<td>2.14 (1.52-3.01)</td>
</tr>
<tr>
<td>Having at least one older sibling</td>
<td>58/88/63</td>
<td>81/124/53</td>
<td>1.30 (1.00-1.69)</td>
</tr>
</tbody>
</table>

Interaction *P* = 0.01

<table>
<thead>
<tr>
<th>Day care attendance before 6 mo of age (thousand child hours)†</th>
<th>n cases, AA/AG/GG</th>
<th>n controls, AA/AG/GG</th>
<th>OR for every one copy of G allele of rs583911*</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;2</td>
<td>49/86/61</td>
<td>83/102/41</td>
<td>1.68 (1.27-2.22)</td>
</tr>
<tr>
<td>≥2</td>
<td>10/4/2</td>
<td>9/19/3</td>
<td>0.44 (0.15-1.31)</td>
</tr>
</tbody>
</table>

Interaction *P* = 0.02

<table>
<thead>
<tr>
<th>Day care attendance before 1 y of age (thousand child hours)†</th>
<th>n cases, AA/AG/GG</th>
<th>n controls, AA/AG/GG</th>
<th>OR for every one copy of G allele of rs583911*</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;5</td>
<td>48/83/58</td>
<td>80/97/39</td>
<td>1.64 (1.24-2.18)</td>
</tr>
<tr>
<td>≥5</td>
<td>11/7/5</td>
<td>12/24/5</td>
<td>0.84 (0.37-1.90)</td>
</tr>
</tbody>
</table>

Interaction *P* = 0.11

*ORs were adjusted for age, sex, Hispanic status (for birth order analysis), race, and annual household income using unconditional logistic regression.

†Cutoff was selected based on the median thousand hours of day care attendance among controls who had day care attendance.
for Hispanics) in both Hispanics and non-Hispanics (Table 5). The CAG haplotype was associated with a reduced childhood ALL risk compared with the most common haplotype among both non-Hispanics and Hispanics (although only borderline significant among non-Hispanics). However, the CGG haplotype was positively associated with childhood ALL only among non-Hispanics.

For IL2RA, a region tagged by rs6602398, rs942201, rs791587, and rs706778 was significantly associated with childhood ALL (global P = 0.02 for non-Hispanics and 0.004 for Hispanics) in both Hispanics and non-Hispanics (Table 5). The ACAG haplotype was associated with an increased risk of childhood ALL compared with the most common haplotype among both non-Hispanics and Hispanics, but Hispanics had two additional common (>5%) haplotypes (CAGA and CCGA) that were associated with an increased risk of childhood ALL compared with the most common haplotype.

For STAT4, even though a region tagged by rs17769459, rs4853546, and rs1031509 was found to be associated with childhood ALL among both Hispanics and non-Hispanics, these associations were mainly driven by the rare (<5%) haplotypes (Table 5).

For STAT6, a region tagged by rs4559, rs1059513, and rs324015 was associated with childhood ALL among both non-Hispanics and Hispanics. Particularly, the GAG haplotype was associated with an increased risk of childhood ALL, although more significantly for Hispanics than for non-Hispanics (Table 5). In addition, the AGG haplotype was associated with an increased risk of childhood ALL among non-Hispanics but not among Hispanics.

Discussion

Of the 208 SNPs analyzed in the study, 19 SNPs belonging to 10 genes (IL12A, STAT4, IL12B, GATA3, PHF11, STAT6, IL10, CTLA4, IL2RA, and MME) showed a significant (P < 0.05) association with childhood ALL. However, only rs583911 of IL12A remained significant after accounting for multiple testing (OR for each copy of variant allele, 1.52; 95% CI, 1.25-1.85; P = 2.9 × 10⁻⁵). The increased risk associated with IL12A rs583911 G allele was stronger among firstborn children of all ethnicities and among children with less child-hours at day care in non-Hispanics. In addition to single SNP analyses, haplotype analyses further identified regions of CD28, FCGR2A, GATA3, IL2RA, STAT4, and STAT6 that may be associated with childhood ALL risk.

Previous studies have shown that newborns have Th2-skewed immune profiles (40-42). Furthermore, during the normal course of immune development, a shift from Th2-dominant to Th1-dominant immune profiles occurs with increasing age (43). It is thought that the major driving force for this immune shift is the production of interleukin-12 (IL12) by innate immune cells (e.g., dendritic cells) after exposure to microbial challenges (44). The IL12 protein is a heterodimer that consists of two subunits, IL12A (p35) and IL12B (p40; ref. 45). In the current study, the most significant result was observed with rs583911 of IL12A, a finding that is strikingly robust against multiple testing. The functional effect of the IL12A SNPs has yet to be characterized. A recent study by Pistiner et al. showed that rs2243123 of IL12A, a SNP in intron 2 that is 739 bp away from rs583911, is associated with immune sensitization to cockroach

Table 4. Haplotype analyses of CD28 and FCGR2A for combined race/ethnicity, NCCLS, 1995-2002

<table>
<thead>
<tr>
<th>Genes/haplotypes</th>
<th>Control (%)</th>
<th>Case (%)</th>
<th>OR (95% CI)*</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CD28</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1879877, rs3181096, rs1181389, and rs3769683</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAAG</td>
<td>36</td>
<td>37</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>CGAG</td>
<td>30</td>
<td>27</td>
<td>0.89 (0.70-1.13)</td>
<td>0.33</td>
</tr>
<tr>
<td>AGGG</td>
<td>15</td>
<td>16</td>
<td>1.09 (0.81-1.46)</td>
<td>0.58</td>
</tr>
<tr>
<td>AGAA</td>
<td>9</td>
<td>9</td>
<td>0.98 (0.67-1.44)</td>
<td>0.92</td>
</tr>
<tr>
<td>CGAA</td>
<td>5</td>
<td>9</td>
<td>1.55 (1.03-2.34)</td>
<td>0.04</td>
</tr>
<tr>
<td>Rare haplotypes†</td>
<td>4</td>
<td>2</td>
<td>0.47 (0.25-0.88)</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Global P = 0.02</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FCGR2A</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10800309 and rs4656308</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>61</td>
<td>57</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>AG</td>
<td>23</td>
<td>21</td>
<td>0.98 (0.78-1.25)</td>
<td>0.90</td>
</tr>
<tr>
<td>AA</td>
<td>16</td>
<td>21</td>
<td>1.46 (1.13-1.90)</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>Global P = 0.02</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*ORs were adjusted for age, sex, Hispanic status, and race using unconditional logistic regression.
†Haplotypes with frequencies <5% were grouped together for analysis.
antigen (46); this lends further support that the region around rs583911 may be important in either the function or the expression of IL12A and may be a promising candidate region to perform fine mapping and functional studies to determine causal variants. The current study also observed 3 SNPs in IL12B (rs3181224, rs1368439, and rs11574790) that are associated with childhood ALL risk only among non-Hispanics, although not statistically significant after multiple testing adjustment. Rs3181224 is located near the 3′ end of IL12B and rs1368439 is located in the 3′ untranslated region (UTR) and may tag polymorphisms in regulatory elements located in this region. Another SNP in the 3′ UTR of IL12B, rs3212227, has been associated with psoriasis, a chronic T-cell–mediated inflammatory disease of the skin (47); however, we observed a null association between this SNP and childhood ALL.

### Table 5. Haplotype analyses of GATA3, IL2RA, STAT4, and STAT6 by Hispanic status, NCCLS, 1995–2002

<table>
<thead>
<tr>
<th>Genes/haplotypes</th>
<th>Non-Hispanics</th>
<th>Hispanics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (%)</td>
<td>Case (%)</td>
</tr>
<tr>
<td><strong>GATA3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4143094, rs3781093, and rs3802604</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA</td>
<td>52</td>
<td>51</td>
</tr>
<tr>
<td>AAG</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>CGG</td>
<td>14</td>
<td>21</td>
</tr>
<tr>
<td>AAA</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>CAG</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Rare haplotypes‡</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>IL2RA</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6602398, rs942201, rs791587, and rs706778</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCGG</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>ACAG</td>
<td>20</td>
<td>26</td>
</tr>
<tr>
<td>CCAA</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>CAGA</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>CCAG</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>CCGA</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Rare haplotypes‡</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td><strong>STAT4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs17769459, rs4853546, and rs1031509</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GGC</td>
<td>59</td>
<td>62</td>
</tr>
<tr>
<td>GAA</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>GAC</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Rare haplotypes‡</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td><strong>STAT6</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4559, rs1059513, and rs324015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAG</td>
<td>56</td>
<td>47</td>
</tr>
<tr>
<td>GAG</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>AGG</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Rare haplotypes‡</td>
<td>8</td>
<td>13</td>
</tr>
</tbody>
</table>

*ORs were adjusted for age, sex, Hispanic status, and race using unconditional logistic regression.
†Included in the rare haplotypes group.
‡Haplotypes with frequencies <5% were grouped together for analysis.
Although our data showed an overall increase in the risk of ALL associated with the variant G allele of IL12A rs583911, this association was stronger among firstborn children who likely have less exposures to infections early in life compared with those with older siblings. These observations further support that childhood ALL may result from inadequate immune modulation early in life due to decreased microbial exposures.

A similar gene-environment interaction was observed between IL12A rs583911 and early day care attendance on the risk of childhood ALL among non-Hispanics. An increased ALL risk associated with the variant G allele of IL12A rs583911 was observed among those children who had fewer child-hours of day care attendance. However, a similar modifying effect of day care attendance was not observed in Hispanics, consistent with published data from the NCCLS reporting ethnic differences in the association between day care attendance and childhood ALL (7). In their report, Ma et al. suggested that day care attendance may not be the primary source of early life exposure to infections for Hispanic children because it was observed that fewer Hispanic children started day care before the age of 1 year and Hispanic children tended to live with more other children in the same household compared with non-Hispanic white children (7). Among the subjects included in the current analysis, fewer Hispanic control children than non-Hispanic control children attended day care before the age of 1 year (12.8% versus 31.6%), and among those who did attend day care, the mean child-hours of day care was significantly lower among Hispanics compared with non-Hispanics (3,900 versus 7,300 child-hours, P = 0.002).

As stated previously, IL12 is a key cytokine for the normal switching, triggered by exposure to microbial challenges, of a Th2-dominant to a Th1-dominant immune profile in early childhood (43, 44). Although the functional effect of rs583911 or the actual causal SNP linked to rs583911 has yet to be elucidated, it seems that the increased ALL risk associated with genetic factors, possibly the effect of rs583911 or the actual causal SNP linked to IL12A, may result from inadequate immune modulation early in life compared with those with older siblings. These findings are in agreement with previous studies that used nongenetic measures and surrogates of infection such as day care attendance, birth order, vaccination history, allergies, and parental reports of infections for investigating the infection- and immune-related etiology of childhood leukemia. In addition, unlike interview-based studies, genetic studies do not suffer from recall errors or recall biases.

AIMs were included in the genotyping to assess potential population stratification especially among the Hispanic population. It was reassuring that the results of the analysis with AIMs indicated that population stratification is minimal in our study most likely because of the matched design. Another strength of the current analysis is the haplotype approach to provide a more comprehensive assessment of variations within candidate genes than has been possible using single SNP analyses. These analyses indicate regions within several genes that would be suitable candidates for replication in other study populations, as well as further studies to identify potential causal variants.

In summary, the current analysis identified associations between polymorphisms of several adaptive immunity genes and childhood ALL. Although only one single-SNP association (rs583911 of IL12A) was statistically robust, these findings provide important support for a role of the adaptive immunity pathway in childhood ALL through immune modulation in early childhood development. Additional support was provided by the results showing that the risk associated with IL12A rs583911 G allele was stronger among children with fewer opportunities for infectious exposures (among firstborn children of all ethnicities and non-Hispanic children with less day care attendance). Further confirmation is needed to determine functional variants in the significant genomic regions identified by this study, in particular for IL12A, which encodes a critical modulator of T-cell development.

In addition, the statistical power to detect an OR of <1.5 may be insufficient in our study especially when the minor allele frequency is 0.10 or less. For example, based on 377 ALL cases and 448 controls, type I error (α) = 0.05, and a minor allele frequency of 0.10, the statistical power to detect an OR of 1.4 is 0.59. Nevertheless, this makes the finding associated with rs583911 of IL12A all the more impressive. Besides an insufficient statistical power for detecting a main effect OR of <1.5 when the minor allele frequency is low, the current study may have an even more limited statistical power for haplotype and interaction analyses (both gene-environment and gene-gene via MDR analysis).

The interpretation of results from complex genetic analyses when multiple comparisons are inherent is always an issue of concern and this constraint exists with the current analysis. However, our most significant result associated with rs583911 of IL12A remained highly significant even after a stringent multiple testing correction using the Bonferroni method.

Results of this genetic study complement those of studies that used nongenetic measures and surrogates of infection such as day care attendance, birth order, vaccination history, allergies, and parental reports of infections for investigating the infection- and immune-related etiology of childhood leukemia. In addition, unlike interview-based studies, genetic studies do not suffer from recall errors or recall biases.

The results of these analyses need to be interpreted in the context of several limitations. We attempted to be inclusive in gene selection, but we were not able to include all of the known adaptive immune function genes. In addition, the statistical power to detect an OR of <1.5 may be insufficient in our study especially when the minor allele frequency is 0.10 or less. For example, based on 377 ALL cases and 448 controls, type I error (α) = 0.05, and a minor allele frequency of 0.10, the statistical power to detect an OR of 1.4 is 0.59. Nevertheless, this makes the finding associated with rs583911 of IL12A all the more impressive. Besides an insufficient statistical power for detecting a main effect OR of <1.5 when the minor allele frequency is low, the current study may have an even more limited statistical power for haplotype and interaction analyses (both gene-environment and gene-gene via MDR analysis).

The interpretation of results from complex genetic analyses when multiple comparisons are inherent is always an issue of concern and this constraint exists with the current analysis. However, our most significant result associated with rs583911 of IL12A remained highly significant even after a stringent multiple testing correction using the Bonferroni method.
No potential conflicts of interest were disclosed.

Acknowledgments

We thank our clinical collaborators and participating hospitals without whose strong support this research could not have been conducted: University of California Davis Medical Center (Dr. Jonathan Ducore), University of California San Francisco (Drs. Mignon Loh and Katherine Matthay), Children’s Hospital of Central California (Dr. Vonda Crouse), Lucile Packard Children’s Hospital (Dr. Gary Dahl), Children’s Hospital Oakland (Dr. James Freasure), Kaiser Permanente Roseville (Drs. Kent Jolly and Vincent Kiley), Kaiser Permanente Santa Clara (Drs. Alan Wong and Carolyn Russo), Kaiser Permanente San Francisco (Dr. Kenneth Leung), and Kaiser Permanente Oakland (Drs. Daniel Kronish and Stacy Month).

References

34. Ritchie MD, Hahn LW, Roedl N, et al. Multifactor-dimensionality...
44. Romagnani S. The increased prevalence of allergy and the hygiene hypothesis: missing immune deviation, reduced immune suppression, or both? Immunology 2004;112:352–63.
Genetic Polymorphisms in Adaptive Immunity Genes and Childhood Acute Lymphoblastic Leukemia

Jeffrey S. Chang, Joseph L. Wiemels, Anand P. Chokkalingam, et al.


Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-10-0389

Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2010/08/16/1055-9965.EPI-10-0389.DC1

This article cites 46 articles, 8 of which you can access for free at:
http://cebp.aacrjournals.org/content/19/9/2152.full#ref-list-1

This article has been cited by 1 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/19/9/2152.full#related-urls

Sign up to receive free email-alerts related to this article or journal.

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.