Research Article

Cancer Epidemiology, Biomarkers & Prevention

Multilocus Association of Genetic Variants in *MLL*, *CREBBP*, *EP300*, and *TOP2A* with Childhood Acute Lymphoblastic Leukemia in Hispanics from Texas

Duangjai Piwkham[1,2,5], Jonathan A.L. Gelfond[3], Budsaba Rerkamnuaychoke[5], Samart Pakakasama[6], Vivienne I. Rebel[1,4], Brad H. Pollock[3], Naomi J. Winick[7], Anderson B. Collier III[7,8], Gail E. Tomlinson[1,2], and Joke Beuten[1,2]

Abstract

Background: Hispanic children have both a higher incidence and a poorer outcome in acute lymphoblastic leukemia (ALL). Moreover, a higher incidence for therapy-related acute myeloid leukemia with 11q23 translocations after treatment with topoisomerase II (topo II) inhibitors has been observed in Hispanic children with ALL. We sought to determine the potential role of genetic variants within the topoisomerase IIα gene (*TOP2A*), within the mixed lineage leukemia gene (*MLL*) and two of its translocation partners, cyclin AMP response element-binding protein gene (*CREBBP*) and E1A binding protein gene (*EP300*) in the increased sensitivity of Hispanic children with ALL to topo II inhibitors.

Methods: Fifty-two tagged single nucleotide polymorphisms (SNP) covering the four genes were genotyped in 241 samples (66 children with ALL and 175 age matched controls) of self-identified Hispanic origin.

Results: Two SNPs within *MLL* (rs525549 and rs6589664) and three SNPs within *EP300* (rs5758222, rs7286979, and rs20551) were significantly associated with ALL ($P = 0.001$–0.04). A significant gene-dosage effect for increasing numbers of potential high-risk genotypes (OR = 16.66; $P = 2 \times 10^{-5}$) and a major haplotype significantly associated with ALL (OR = 5.68; $P = 2 \times 10^{-6}$) were found. Replication in a sample of 137 affected White children and 239 controls showed that only rs6589664 (*MLL*) was significantly associated in this ethnic group.

Conclusions: Our findings indicate that the association between ALL and common genetic variants within *MLL* and *EP300* is population specific.

Impact: Replication of our findings in independent Hispanic populations is warranted to elucidate the role of these variants in ALL susceptibility and define their importance in the ethnic specific differences in ALL risk. *Cancer Epidemiol Biomarkers Prev; 20(6); 1204–12. ©2011 AACR.*

Introduction

Acute lymphoblastic leukemia (ALL) is a heterogeneous disease characterized by the predominance of immature hematopoietic cells, in which malignant cells express phenotypes of either T-cell or B-cell lineages (1). ALL accounts for 25% to 30% of all diagnoses of cancer in children (2, 3). The incidence of lymphoid leukemia has been found to be approximately 15% higher among Hispanic children than among White non-Hispanic children (4). No biological factors have been identified which account for this difference in risks among different ethnic or racially defined populations. Substantial evidence exists indicating that several risk factors (including gender, age, ethnicity, exposure to radiation, genetic syndromes) are involved in increased risk of ALL (5–9). Hispanic children with ALL have been found to have a higher incidence rate for secondary or therapy-related acute myeloid leukemia (t-AML) after treatment with topoisomerase II (topo II) inhibitors (10). Topoisomerase IIα (*TOP2A*) is a key enzyme in DNA replication and a molecular target for many anticancer drugs (11). In breast cancer, *TOP2A* gene amplification and deletion are associated with increased and decreased sensitivity to topo II inhibitors, respectively (12). A possible role of *TOP2A* in the development of leukemia has been suggested by Guerin and colleagues (13), who found overexpression

Authors' Affiliations: [1]Greehey Children's Cancer Research Institute; Departments of [2]Pediatrics, [3]Epidemiology and Biostatistics, [4]Cellular and Structural Biology, The University of Texas Health Science Center at San Antonio, San Antonio, Texas; Departments of [5]Pathology, [6]Pediatrics, Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; [7]University of Texas Southwestern Medical Center, Dallas, Texas; and [8]Lehigh Valley Health Network, Bethlehem, Pennsylvania

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

Corresponding Author: Joke Beuten, Department of Pediatrics, The University of Texas Health Science Center at San Antonio/Greehey Children's Cancer Research Institute, 8403 Floyd Curl Dr., San Antonio, TX 78229. Phone: 210-562-9155; Fax: 210-562-9135.
E-mail: beuten@uthscsa.edu

doi: 10.1158/1055-9965.EPI-11-0059

of *TOP2A* in ALL cases. Because of the frequent finding of rearrangements of the mixed lineage leukemia gene (*MLL*) in therapy-related leukemias (14, 15), particularly those associated with previous therapies that included a topo II inhibitor (16), it was hypothesized that translocations could be caused by binding of topo II poisons to topo II cleavage sites. Such topo II consensus sites have been identified in the 8.3 kb breakpoint cluster region (BCR) of the gene (15), encompassing exons 8 to 14 [originally designated exons 5–11, where the breakpoints typically occur (17)]. The *MLL* gene encodes a histone H3 lysine 4–specific methyltransferase which is essential for normal hematopoiesis and mammalian development via the regulation of homeobox (*HOX*) genes (18). Previous studies have shown that overexpression of certain *HOX* genes can induce leukemia in mice (19).

Alternatively, chromatin structure and DNA accessibility have been suggested to be the critical determinants of cleavage within *MLL* resulting in translocations (20–22), a hypothesis that is corroborated by the findings of Khobta and colleagues (23), who observed specific histone modifications and acetylase/deacetylase within the BCR of *MLL.* In this context, the cyclin AMP response element binding protein gene (*CREBBP*, also known as CBP) located at 16p13.3 and its paralogue, the E1A-binding protein gene (*EP300*, also known as *p300*), located at 22q13 are of interest (24). Both genes have been characterized as fusion genes of *MLL* translocations in t-AML (25–28) or occasionally in *de novo* AML (29). *CREBBP* and *EP300* are transcriptional coactivators that play an essential role in transcriptional regulation by chromatin remodeling through their own histone acetyltransferase activity and by interacting with transcription factors and other histone acetylators (30). A potential regulatory role for *CREBBP*/*EP300* for protein acetylation in base mismatch repair and in maintaining genomic stability has been previously suggested (31, 32). A study by Shigeno and colleagues further found *CREBBP*/*EP300* mutations in leukemias indicating that genetic alterations in these genes might be involved in leukemogenesis (30). Furthermore, Rubinstein–Taybi syndrome, characterized by mutations in *CREBBP*, is associated with increased risk to developing cancer, including leukemia (33). In mice, it was showed that *EP300* and *CREBBP* function as suppressor of hematologic tumor formation (34).

Because of the higher risk in Hispanics to develop secondary leukemias with translocations involving 11q23 after treatment with topo II inhibitors, we sought to determine the potential role of genetic variants in *TOP2A*, *MLL*, *CREBBP*, and *EP300* in the increased sensitivity of Hispanic children with ALL to topo II inhibitors. We investigated possible association between 52 tagged single nucleotide polymorphisms (SNP) covering the 4 genes in 241 Hispanic children with ($n = 66$) and without ($n = 175$) ALL and conducted a similar analysis in a White case–control sample to compare findings.

With a multicausal etiology of ALL, it is likely that multiple risk variants act simultaneously, synergistically or additively, to influence ALL susceptibility. We therefore determined associations of individual SNPs and higher-order SNP–SNP interactions. Our results suggest that common genetic variants in *MLL* and *EP300*, both individually and jointly, are associated with an increase in risk for ALL in Hispanics.

Materials and Methods

Subjects

Consecutive case DNAs were obtained via blood draw from 66 Hispanic and 137 White non-Hispanic (referred to as "White") unrelated children that were diagnosed with B-lineage ALL between 1990 and 2005 at the Children's Medical Center of Dallas ($n = 196$) and from a pilot study project between 2007 and 2009 at The University of Texas Health Science Center at San Antonio ($n = 7$). Collection of blood samples and clinical information from cases and controls was undertaken with signed informed consent from the parents of each child and ethical review board approval at both universities in accordance with the principles of the Declaration of Helsinki. A total of 414 age, gender, and ethnicity-matched unrelated 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 including 175 Hispanics and 239 Whites. Exclusion criteria for the controls consisted of any history of malignancy as confirmed by review of medical records. Matching of the sample was done on frequency of the whole group, not individually. Ethnicity was self-reported as determined by parental report. The Hispanics were primarily children of Mexican descent, all with Hispanic surnames. To our knowledge, there was only 1 self-declared Hispanic case, which had 1 non-Hispanic parent. All samples underwent standard DNA preparation procedures and DNA was stored at 4°C before genotyping. Characteristics of the study group are summarized in Table 1.

SNP selection and genotyping

SNPs were selected from the dbSNP database Build 129 (35) and SNPper (36). First, SNPs with potential functional effects and a minor allele frequency (MAF) more than 0.05 were defined from programs predicting functional annotation (F-SNP; ref. 37), PupaSuite (38), SNPinfo (39), and SNPs associated with leukemia, and in particular childhood ALL, as derived from literature (40), were determined. After this initial selection, tagging SNPs within each gene were identified by using Haploview (41) with the following criteria: (i) inclusion of preselected SNPs (see above), (ii) MAF more than 0.05 to gain more statistical power, (iii) r^2 threshold of 0.8, (iv) a log of odds threshold for multimarker 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. Because of limited amount of

Table 1. Characteristics of the study group

Subgroup	Cases ($n = 203$)	Controls ($n = 414$)
Ethnic background (*n*)		
White	137	239
Hispanic	66	175
Age (mean ± SD)		
White	6.0 ± 4.8	10.6 ± 5.9
Hispanic	6.8 ± 4.8	9.3 ± 5.2
Male/Female		
White	76/61	126/113
Hispanic	28/38	86/89
Ancestry in Hispanics (%)		
Native American	41.9	48.7
European	53.7	43.5
African	4.4	7.8

genotype data on Hispanics in dbSNP Build 129, the selection of tagged SNPs was based on the information on the European population. Genotyping was done with the GoldenGate assay of the VeraCode technology by using the BeadXpress reader system according to the manufacturer's protocol (Illumina). Duplicate samples were included as quality controls and the reproducibility rate was more than 99.8% for all SNPs. The observed error rate of duplicate samples was due to the fact that in 2 occasions 1 sample of a duplicate had too low a signal to being called by the program. An average 98% of the samples were successfully genotyped per SNP.

Statistical analysis

Allele frequencies among the cases and controls in each ethnic group were compared by using the χ^2 test. The OR and its 95% CI was estimated by logistic regression by using *R* statistical software version 2.9.1. All statistical tests were 2-sided and significance was set at $P < 0.05$. We used additive, dominant, and recessive models in the test analysis and chose the model with strongest association for presentation (i.e., model with smallest *P* value with minimum 5 individuals). The cumulative effect of combined genotypes was estimated by counting the number of genotypes associated with ALL on the basis of the best-fitting genetic inheritance from single SNP analysis. ORs and their 95% CIs were calculated for carriers of any combination of 1, or any combination of 2 or more alleles associated with ALL as compared with carriers of none of the risk genotypes by using the unconditional logistic regression analysis. The total number of risk alleles for each subject was treated as an additive risk factor, and P_{trend} is the *P* value for the association between the number of risk alleles and disease outcome. Logistic regression was used to calculate ORs of the haplotypes by using the method implemented in the haplo.ccs package in *R* (42). Only major haplotypes, with an estimated frequency of more than 5%, are considered in this report. The model was fit for each major haplotype so that the OR of each major haplotype was computed relative to a reference group consisting of all other haplotypes including rare haplotypes. All analyses were stratified by ethnicity and Native American ancestry was used as covariate in the analysis of the Hispanic sample. Haploview was used to measure linkage disequilibrium and to define the block structure between the tagging SNPs for each ethnicity. To correct for multiple testing in the SNP analysis, we used the method of Storey and Tibshirani on the basis of the concept of false discovery rate (FDR; ref. 43). The FDR estimation showed that for $P < 0.05$, the probability that the association is a true positive is more than 60% and for $P < 0.006$ the probability is more than 90%. The Bonferroni correction was used for multiple comparisons in haplotype analysis in which the adjusted *P* value is given by *P* at the 0.05 significance level divided by the number of major (frequency >5%) haplotypes.

Estimates of genetic ancestry of each individual in Hispanics were derived from 96 ancestral informative markers by using a maximum likelihood method implemented in the program Maximum Likelihood Individual Admixture Estimation, as described previously (Beuten and colleagues, in press; ref. 44). The ancestral proportions were estimated at 46% European, 47% Native American, and 7% West-African. To determine population stratification genetic ancestry was used as a covariate in the SNP association analysis and compared with the unadjusted outcome.

Results

Allele frequency distribution

A total of 52 tagged SNPs were initially selected and genotyped in each sample. Ten SNPs were not polymorphic or failed genotyping and were left out from the study. All of the remaining 42 markers analyzed were in Hardy–Weinberg equilibrium in control samples. Table 2 shows the MAF distribution in the cases and the controls in each ethnic group. A significant difference in allele frequency between cases and controls was found for 5 SNPs in Hispanics; 2 (rs525549 and rs6589664) in the *MLL* gene and 3 (rs5758222, rs7286979, and rs20551) in the *EP300*. In Whites, none of the SNPs showed a significantly different allele frequency between cases and controls. Several markers showed a significant difference in allele frequencies across ethnic groups. In particular for variants within EP300 highly significant allele frequency differences were found between the control groups of Hispanics and Whites (Table 2). High pairwise linkage disequilibrium (LD) was found between SNPs within each of the 4 genes in both ethnicities (Supplementary Table S1). Several pairwise LD measures within each gene were significantly different between Hispanics and Whites. For example, perfect LD between rs9611502 and rs20551 or rs20552 in *EP300* was found in Whites, whereas both combinations showed low LD in Hispanics.

Table 2. Tagged SNPs, their location and minor allele frequencies by race/ethnicity and case–control status

SNP	Chr	Position	Function SNP	Minor allele	Hispanics			Whites			Hispanics vs. Whites	
					MAF case	MAF control	*P*	MAF case	MAF control	*P*	*P* cases	*P* controls
MLL												
rs4456287	11	117803462	5′ flank	G	0.19	0.18	0.91	0.24	0.21	0.31	0.20	0.33
rs629470	11	117825763	Intron	A	0.09	0.14	0.18	0.19	0.15	0.15	**0.02**	0.84
rs9332801	11	117860852	Coding exon	C	0.03	0.05	0.40	0.07	0.04	0.06	0.12	0.45
rs525549	11	117863278	Intron	A	0.39	0.26	**0.005**	0.38	0.41	0.39	0.83	***P* < 0.0001**
rs562780	11	117867624	Intron	A	0.05	0.06	0.64	0.15	0.11	0.13	**0.003**	**0.01**
rs2071702	11	117879071	Coding exon	A	0.04	0.06	0.35	0.06	0.04	0.14	0.29	0.19
rs573971	11	117898606	3′-UTR	A	0.11	0.15	0.39	0.23	0.17	0.05	**0.01**	0.34
rs6589664	11	117910014	3′ flank	A	0.42	0.54	**0.02**	0.28	0.33	0.16	**0.004**	***P* < 0.0001**
rs3741324	11	117911045	3′ flank	A	0.05	0.09	0.28	0.16	0.13	0.20	**0.003**	0.06
CREBBP												
rs2191416	16	3710730	3′ flank	G	0.47	0.42	0.29	0.33	0.36	0.51	**0.01**	0.09
rs39733	16	3713165	3′ flank	A	0.20	0.17	0.38	0.25	0.27	0.48	0.31	**0.0003**
rs3751845	16	3718338	Coding exon	A	0.00	0.02	0.13	0.01	0.00	0.25	0.31	**0.02**
rs129974	16	3735293	Coding exon	A	0.00	0.01	0.22	0.01	0.00	0.26	0.32	0.09
rs129963	16	3736148	Intron	A	0.47	0.53	0.99	0.44	0.45	0.77	0.10	**0.03**
rs17136507	16	3740546	Intron	G	0.19	0.22	0.80	0.11	0.13	0.48	**0.02**	**0.003**
rs886528	16	3751557	Intron	G	0.33	0.37	0.45	0.43	0.43	0.93	0.07	0.09
rs3025694	16	3760668	Coding exon	A	0.00	0.01	0.22	0.00	0.00	0.69	0.49	0.09
rs130005	16	3768349	Intron	G	0.07	0.06	0.71	0.10	0.11	0.76	0.27	**0.01**
rs130021	16	3772472	Intron	G	0.28	0.25	0.49	0.32	0.36	0.24	0.40	**0.001**
rs130023	16	3773307	Intron	G	0.01	0.04	0.14	0.01	0.04	0.06	0.54	0.52
rs130024	16	3774317	Intron	G	0.04	0.03	0.60	0.08	0.10	0.33	0.20	**0.0002**
rs11076787	16	3792777	Intron	A	0.14	0.17	0.47	0.20	0.18	0.50	0.11	0.48
rs1296720	16	3813643	Intron	C	0.09	0.11	0.49	0.18	0.17	0.69	**0.02**	**0.03**
rs2530890	16	3844547	Intron	G	0.06	0.10	0.20	0.18	0.16	0.48	**0.002**	**0.01**
rs2239316	16	3853996	Intron	G	0.29	0.28	0.78	0.26	0.29	0.26	0.47	0.69
rs2526687	16	3872726	5′ flank	A	0.02	0.03	0.50	0.02	0.03	0.29	0.83	0.62
TOP2A												
rs16965741	17	35794735	3′ flank	A	0.18	0.12	0.11	0.11	0.10	0.74	0.05	0.32
rs2715555	17	35798150	3′ flank	A	0.07	0.07	0.90	0.13	0.14	0.70	0.08	**0.001**
rs13695	17	35798719	3′ UTR	A	0.12	0.15	0.42	0.26	0.26	0.80	**0.002**	**0.0004**
rs11540720	17	35799350	Coding exon	A	0.01	0.02	0.44	0.02	0.02	0.97	0.29	0.55
rs525812	17	35828818	5′ region	G	0.37	0.37	0.99	0.20	0.19	0.70	**0.0004**	***P* < 0.0001**
rs11650680	17	35832762	5′ flank	A	0.20	0.18	0.70	0.18	0.22	0.29	0.78	0.18
rs1530363	17	35836302	5′ flank	G	0.39	0.41	0.79	0.22	0.21	0.78	**0.0003**	***P* < 0.0001**
EP300												
rs5758222	22	39816139	5′ flank	G	0.28	0.15	**0.002**	0.33	0.34	0.77	0.30	***P* < 0.0001**
rs4822002	22	39817290	5′ flank	G	0.21	0.22	0.77	0.33	0.37	0.37	**0.01**	***P* < 0.0001**
rs5758223	22	39819866	Intron	G	0.13	0.13	0.97	0.24	0.25	0.71	**0.01**	***P* < 0.0001**
rs7286979	22	39828573	Intron	A	0.28	0.16	**0.003**	0.34	0.35	0.68	0.24	***P* < 0.0001**
rs9611502	22	39861182	Intron	A	0.02	0.02	0.62	0.04	0.03	0.57	0.24	0.55
rs20551	22	39877954	Coding exon	G	0.50	0.62	**0.02**	0.30	0.29	0.74	**0.0001**	***P* < 0.0001**
rs20552	22	39880985	Coding exon	A	0.19	0.20	0.80	0.37	0.36	0.77	**0.0004**	***P* < 0.0001**
rs2294976	22	39894654	Intron	A	0.05	0.05	0.97	0.09	0.10	0.50	0.15	**0.003**
rs1046088	22	39904329	Coding exon	C	0.02	0.02	0.99	0.02	0.03	0.24	0.77	0.33

NOTE: Significant *P* values are in bold.
Chr: Chromosome.

Association between individual SNPs and ALL

Five SNPs showed a significant association with ALL in Hispanics, of which rs525549 (*MLL*) and rs20551 (*EP300*) were significant at $\alpha = 0.05$ (FDR <40%) and rs65896642 (*MLL*), rs5758222, and rs7286979 (*EP300*) at $\alpha = 0.006$ (FDR < 10%; Table 3). The strongest effect was seen for which homozygous AA carriers at rs525549 had a 3.35-fold increase in ALL risk (OR = 3.35; 95% CI = 1.37–8.18; $P = 0.008$). Results shown in Table 3 are not adjusted; however, inclusion of age or gender as covariate showed very similar results (data not shown). We also tested whether population stratification was confounding the outcome. Including individual Native American ancestral estimates as covariate in the regression analysis, indicated that significance remained for all 5 SNPs, although was slightly reduced for the 2 MLL variants (Table 3). A similar observation was found after correction for European ancestral proportions, whereas the outcome after adjusting for African ancestry did not differ from unadjusted results (data not shown).

Replication of the association findings in a White sample group showed that only rs65896642 within *MLL* was significantly associated with ALL ($P = 0.029$, Table 3). Of note is that for rs6589664 the A allele is the minor allele in White controls, whereas it is the major allele in Hispanic controls (the allele frequency of the controls is used as reference). However, *G* is the risk allele in both ethnic groups; in Whites, homozygous AA carriers have a 2.50-fold decrease in ALL risk whereas homozygotes for the G allele show a 2.32-fold increase in ALL risk in Hispanics. Results from logistic analysis for all SNPs by inheritance models and by ethnicities are shown in Supplementary Table S2.

Cumulative effects of significant SNPs

Multivariate logistic regression on combinations of multiple risk alleles defined in the SNP analysis as compared with the combination with no risk allele as reference, showed cumulative risk effects. Because rs5758222 and rs7286979 in *EP300* are in complete LD, we only included one of both SNPs in the model. An interaction was found for rs525549 and rs6589664 in *MLL* and rs5758222 in *EP300*, with a more than 16-fold increase in ALL risk (OR = 16.66; 95% CI = 4.48–61.9; $P_{trend} < 0.0001$; Table 4). It was noted that including rs7286979 instead of rs5758222 in the model, showed a similar outcome (data not shown). Moreover, adding rs20551 in the model also resulted in a significant cumulative effect. However, the 95% CI of the OR was much wider because of the absence of individuals with zero-risk genotypes (used as reference).

Interaction between haplotypes

To examine the joint effect of risk alleles, the 5 SNPs that were found to be significantly associated with ALL from SNP analysis were included in haplotype analysis. A major haplotype A-G-G-A-A (6%), from the interaction between haplotypes of *MLL* and *EP300*, is significantly associated with ALL with an adjusted OR of 5.68 (95% CI = 2.82–11.44; $P < 0.0001$) and carries all 5 alleles associated with increased risk in SNP analysis (Table 5). This association remained significant after Bonferroni correction (the adjusted *P* value at the 0.05 significance level for 6 major haplotypes is 0.0083).

Discussion

Because of our previously observed predilection for Hispanic children to develop secondary leukemias after treatment with topo II inhibitors, often showing 11q23 translocations, we focused on examining the association of variants within *TOP2A*, *MLL*, and 2 of its fusion partners, *CREBBP* and *EP300*, hypothesizing that variation in these genes could influence leukemia risk in general in this population, and estimated risks conferred by both individual SNPs and SNP–SNP joint effects. No association studies of genetic variants within these genes with ALL susceptibility have been reported to date. Only very few association studies of candidate gene in Hispanics have been reported so far (45–48). Being the fastest-growing cultural group and second-largest racial/ethnic group of the population in the United States, genetic studies in Hispanics, such as our report, are important steps in understanding the unique cancer risks in this susceptible population.

Genotyping of 52 haplotype tagged SNPs covering the 4 genes showed that 2 SNPs within *MLL* and 3 within *EP300* were significantly associated with ALL ($P = 0.001$–0.04). The most significant association was found for rs525549 (OR = 3.52; $P = 0.004$), located in intron 11 of the *MLL* gene. The breakpoints for most *MLL* rearrangements, in particular in t-AML patients with previous exposure to drugs targeting topo II, cluster in the BCR encompassing exons 8 through 14 (15). The association finding for rs525549 in Hispanics is therefore of particular interest because this ethnic group has been shown to be at increased risk to develop secondary leukemias related to therapy with topo II inhibitors (10). The BCR is characterized by multiple topo II consensus sites and rs525549 is 2 base pairs upstream from a 90% matching consensus binding site for topo II. Moreover, rs525549 is less than 1.5 kb upstream and in high LD with the region encompassing exon 12 [or exon 9 according to an older numbering system (17)] containing a strong *in vivo* topo II consensus binding site (49) and less than 1.3 kb from a translocation hotspot (22) as well as a gene-internal promoter having a binding site for RNA polymerase II likely implicated in genetic (in)stability of the *MLL* gene (50). The association finding of rs525549 in Hispanics, therefore suggests that genetic variation in *MLL* could be of biological significance and be implicated in the increased risk of t-AML in Hispanics.

Replication in Whites revealed that only SNP rs6589664 (*MLL*) was found to be significant and the A allele is the protective allele for ALL in both ethnic groups. Rs6589664 is located downstream of the *MLL* gene and maps within

Table 3. Significant results from individual SNP effects on ALL

Gene	SNP	Genotype	Hispanics					Whites			
			Control/case	OR (95% CI)[a]	P_{unadj}	OR (95% CI)[b]	P_{adj}		Control/case	OR (95% CI)[a]	
MLL	rs525549	TT	98/27	Ref	-				83/49	Ref	-
		AA	13/12	3.35 (1.37–8.18)	**0.008**	**2.84 (1.07–7.55)**	**0.036**		41/17	0.70 (0.36–1.37)	0.299
		AT	64/26	1.47 (0.79–2.75)	0.222	1.24 (0.62–2.46)	0.540		115/67	0.99 (0.62–1.57)	0.955
		AA vs. AT/TT	13/12	2.82 (1.21–6.56)	**0.016**	**2.56 (1.02–6.41)**	**0.044**		41/17	0.71 (0.38–1.30)	0.267
MLL	rs6589664	AA	46/16	Ref				GG	112/66	Ref	-
		GG	31/25	2.32 (1.07–5.03)	**0.034**	1.85 (0.76–4.47)	0.174	AA	30/7	**0.40 (0.16–0.95)**	**0.038**
		AG	98/24	0.70 (0.34–1.45)	0.342	0.65 (0.29–1.44)	0.286	AG	97/60	1.05 (0.67–1.63)	0.830
		GG vs. AG/AA	31/25	2.90 (1.54–5.47)	**0.001**	**2.49 (1.23–5.04)**	**0.011**	AA vs. AG/GG	30/7	**0.39 (0.17–0.91)**	**0.029**
EP300	rs5758222	AA	126/32	Ref	-				99/56	Ref	-
		GG	5/4	3.15 (0.80–12.4)	0.101	3.11 (0.74–13.15)	0.123		25/13	0.92 (0.44–1.94)	0.825
		AG	44/28	2.51 (1.36–4.62)	**0.003**	**2.75 (1.41–5.35)**	**0.003**		113/60	0.94 (0.60–1.48)	0.784
		GG/AG vs. AA	49/32	2.57 (1.42–4.64)	**0.002**	**2.79 (1.46–5.32)**	**0.002**		138/73	0.94 (0.61–1.44)	0.762
EP300	rs7286979	GG	126/33	Ref	-				97/56	Ref	-
		AA	6/4	2.55 (0.68–9.55)	0.166	2.58 (0.65–10.31)	0.180		26/12	0.80 (0.37–1.71)	0.563
		AG	43/28	2.49 (1.35–4.58)	**0.003**	**2.68 (1.38–5.21)**	**0.004**		115/66	0.99 (0.64–1.55)	0.979
		AA/AG vs. GG	49/32	2.49 (1.39–4.49)	**0.002**	**2.67 (1.40–5.07)**	**0.003**		141/78	0.96 (0.62–1.47)	0.846
EP300	rs20551	GG	69/14	Ref				AA	120/63	Ref	-
		AA	28/14	2.46 (1.04–5.83)	**0.040**	**2.86 (1.06–7.73)**	**0.039**	GG	21/12	1.09 (0.50–2.36)	0.830
		AG	78/36	2.27 (1.13–4.57)	**0.021**	**2.77 (1.22–6.29)**	**0.015**	AG	97/52	1.02 (0.65–1.61)	0.928
		AA/AG vs. GG	106/50	2.32 (1.19–4.52)	**0.013**	**2.79 (1.26–6.18)**	**0.011**	GG/AG vs. AA	118/64	1.03 (0.67–1.59)	0.882

NOTE: All significant values are in bold.
[a]Unadjusted and [b]Adjusted Native American ancestry.

Table 4. Cumulative effect of risk variants

Markers	Risk genotypes	Controls	Cases	OR[a]	95% CI	*P*
MLL: rs525549-rs6589664 + *EP300*: rs5758222						
	0	97	19	Ref	-	-
	1	64	24	1.78	0.83–3.81	0.87
	2	13	13	5.55	2.01–15.4	0.03
	3	1	6	35.12	3.75–329	0.0004
	Trend			**16.66**	**4.48–61.9**	**0.00002**

NOTE: Significant values are given in bold.
[a]Adjusted for Native American ancestry.

exon 7 of the mRNA-transmembrane protein 25 gene (*TMEM25*), adjacent to *MLL*. *TMEM25* is a member of the immunoglobulin superfamily, implicated in immune responses, growth factor signaling, and cell adhesion (51). No data have been reported on a potential role of *TMEM25* in leukemogenesis and further studies are warranted to identify its contribution to ALL. Alternatively, high LD exists between rs6589664 and other SNPs within *MLL*, and rs6589664 therefore might be merely a marker for the true causal factor.

We also found 3 variants within *EP300* (rs5758222, rs7286979, and rs20551) that were significantly associated with ALL in Hispanics. None of these variants have been previously reported in association studies with childhood ALL. The population-specific allele frequencies, showing significant differences in distribution between Hispanics and Whites, might explain the lack of statistical significance in Whites for genetic variants located within *EP300*. Of note is that perfect LD between rs9611502 and rs20551 or rs20552 was found in Whites, whereas both combinations showed low LD in Hispanics, suggestive for higher recombination rate in this region of the gene for the latter group.

Our study further showed a significant dosage effect with increasing numbers of potential high-risk genotypes. Combining 3 high-risk genotypes of *MLL* and *EP300* in Hispanics showed synergistic effects; the risks conferred by multiple polymorphisms were elevated as compared with any of them alone (OR = 16.66; $P < 0.0001$). From these findings, we conclude that an interaction effect between different loci is operative and likely confers ALL risk. Such interaction was further confirmed by the interaction between haplotypes of both genes. Caution should be used when interpreting the results as to the biological meaning of these multilocus interactions; statistical significance does not necessarily correspond to true biological interactions between the specific genes studied (52). However, our findings are supportive for cooperative roles of *EP300* and *MLL* in leukemogenesis and show for the first time a genetic interaction between *MLL* and *EP300* in Hispanic children with ALL.

Significant heterogeneity exists among Hispanics in the United States and worldwide, however, the individuals designated self-described as Hispanic in this study virtually all derive from Mexico across the U.S. border in South Texas. Therefore, any possible race/ethnicity misclassification would be equally likely in cases and controls of each ethnic group and should therefore not have a substantial impact on our conclusion. In addition, after including the proportion of Native American ancestry as covariate, determined by information retrieved from ancestral informative markers, the associations did not

Table 5. Interaction between haplotypes of *MLL* and *EP300*

SNP combination	Frequency (%)	Controls	Cases	OR[a]	95% CI	*P*
MLL: rs525549-rs6589664 + *EP300*: rs5758222-rs7286979-rs20551[b]						
TA-AGG	29	95	20	0.50	0.28–0.90	0.022
AG-AGG	16	53	15	0.86	0.47–1.58	0.628
TG-AGG	12	45	9	0.72	0.36–1.44	0.346
TA-AGA	11	38	9	0.72	0.35–1.45	0.356
TA-GAA	9	31	10	1.3	0.62–2.63	0.502
AG-GAA	6	11	15	**5.68**	**2.82–11.44**	**0.000002**

NOTE: Significant results after Bonferroni correction are in bold.
Only common haplotypes (>5%) are shown.
[a]Adjusted for Native American ancestry.
[b]Dominant model.

change for variants within *EP300* and were slightly reduced for variants in *MLL*. Therefore, these associations are true positives and the confounding effects because of admixture are insignificant.

Recently, genome-wide association studies in Europeans and Koreans have revealed that genetic variants play a role in childhood ALL (53–55). However, Hispanics have not been analyzed extensively at a genome-wide level. This may be explained by the fact that genotype information on Hispanics has only recently become more abundantly available in HapMap. For that reason, our selection of the 52 tagged SNPs was based on HapMap data of the European population. Because of the ethnic-specific LD patterns, it is possible that some tagged SNPs in Hispanics were omitted and the SNPs selected for our study may therefore not fully represent all tagged variants in this ethnic group. However, our selection of SNPs was enriched with potentially functionally important SNPs and thus high-priority SNPs were included in the analysis.

A limitation of the study is that we only evaluated pre-B ALL and did not consider other phenotypic subtypes neither did we examine interactions with nongenetic factors that may be involved in disease risk. In addition, the statistical power of this study is limited by the sample size, in particular for the analysis of the joint effects of risk alleles. The minimal detectable OR, corresponding to 80% power, is 2.05 considering a median MAF of 15% for our Hispanic sample of 66 cases and 175 controls. Therefore, replication of our findings in independent Hispanic samples is warranted to elucidate the role of these variants in ALL susceptibility and define their importance in the ethnic specific differences in ALL risk.

In summary, our results support the hypothesis that ALL develops through a variety of interrelated genes (polygenic model) and suggest that common genetic variants within *MLL* and *EP300* genes are associated with ALL in Hispanics from Texas. The susceptibility defined by these genetic variants is population specific. To our knowledge, this is the first report to explore the effect of higher-ordered interactions between these genes and ALL risk. Our findings might help to improving prediction risk of an individual for ALL, or in explaining leukemia risks at the population level.

Disclosure of Potential Conflicts of Interest

The authors have declared that no competing interests exist. The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Acknowledgment

We thank family members for their participation and clinical staff for their skilled assistance.

Grant Support

This work was supported by the Texas Higher Education Coordinating Board Grant #010019–0105-2001, the Children's Cancer Fund Dallas, Texas and the Greehey Family Foundation (to G.E. Tomlinson), and the Commission on Higher Education Scholarship, Thailand (to D. Piwkham).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 19, 2011; revised March 23, 2011; accepted April 3, 2011; published OnlineFirst April 14, 2011.

References

1. Krajinovic M, Labuda D, Richer C, Karimi S, Sinnett D. Susceptibility to childhood acute lymphoblastic leukemia: influence of CYP1A1, CYP2D6, GSTM1, and GSTT1 genetic polymorphisms. Blood 1999;93:1496–501.
2. Karathanasis NV, Choumerianou DM, Kalmanti M. Gene polymorphisms in childhood ALL. Pediatr Blood Cancer 2009;52:318–23.
3. Chen CL, Liu Q, Pui CH, Rivera GK, Sandlund JT, Ribeiro R, et al. Higher frequency of glutathione S-transferase deletions in black children with acute lymphoblastic leukemia. Blood 1997;89:1701–7.
4. Glazer ER, Perkins CI, Young JL, Schlag RD, Campleman SL, Wright WE, et al. Cancer among Hispanic children in California, 1988–1994: comparison with non-Hispanic white children. Cancer 1999;86:1070–9.
5. Emerenciano M, Koifman S, Pombo-de-Oliveira MS. Acute leukemia in early childhood. Braz J Med Biol Res 2007;40:749–60.
6. Buffler PA, Kwan ML, Reynolds P, Urayama KY. Environmental and genetic risk factors for childhood leukemia: appraising the evidence. Cancer Invest 2005;23:60–75.
7. Soldin OP, Nsouli-Maktabi H, Nsouly-Maktabi H, Genkinger JM, Loffredo CA, Ortega-Garcia JA, et al. Pediatric acute lymphoblastic leukemia and exposure to pesticides. Ther Drug Monit 2009;31:495–501.
8. Wakeford R.Childhood leukaemia following medical diagnostic exposure to ionizing radiation in utero or after birth. Radiat Prot Dosimetry 2008;132:166–74.
9. Taub JW, Ge Y. The prenatal origin of childhood acute lymphoblastic leukemia. Leuk Lymphoma 2004;45:19–25.
10. Winick N, Buchanan GR, Kamen BA. Secondary acute myeloid leukemia in Hispanic children. J Clin Oncol 1993;11:1433.
11. Wang JC. DNA topoisomerases. New break for archaeal enzyme. Nature 1997;386:329,331.
12. Järvinen TA, Tanner M, Rantanen V, Bärlund M, Borg A, Grénman S, et al. Amplification and deletion of topoisomerase II alpha associate with ErbB-2 amplification and affect sensitivity to topoisomerase II inhibitor doxorubicin in breast cancer. Am J Pathol 2000;156:839–47.
13. Guérin E, Entz-Werlé N, Eyer D, Pencreac'h E, Schneider A, Falkenrodt A, et al. Modification of topoisomerase genes copy number in newly diagnosed childhood acute lymphoblastic leukemia. Leukemia 2003;17:532–40.
14. Aplan PD. Chromosomal translocatios involving the MLL gene: molecular mechanisms. DNA Repair (Amst) 2006;5:1265–72.
15. Broeker PL, Super HG, Thirman MJ, Pomykala H, Yonebayashi Y, Tanabe S, et al. Distribution of 11q23 breakpoints within the MLL breakpoint cluster region in *de novo* acute leukemia and in treatment-related acute myeloid leukemia: correlation with scaffold attachment regions and topoisomerase II consensus binding sites. Blood 1996;87:1912–22.
16. Pedersen-Bjergaard J, Andersen MK, Johansson B. Balanced chromosome aberrations in leukemias following chemotherapy with DNA-topoisomerase II inhibitors. J Clin Oncol 1998;16:1897–8.

17. Gu Y, Alder H, Nakamura T, Schichman SA, Prasad R, Canaani O, et al. Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved in acute leukemia. Cancer Res 1994;54:2327–30.
18. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, et al. MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol Cell 2002;10:1107–17.
19. Abramovich C, Humphries RK. Hox regulation of normal and leukemic hematopoietic stem cells. Curr Opin Hematol 2005;12:210–6.
20. Udvardy A, Schedl P. Chromatin structure, not DNA sequence specificity, is the primary determinant of topoisomerase II sites of action *in vivo*. Mol Cell Biol 1991;11:4973–84.
21. Kas E, Laemmli UK. *In vivo* topoisomerase II cleavage of the Drosophila histone and satellite III repeats: DNA sequence and structural characteristics. EMBO J 1992;11:705–16.
22. Mirault ME, Boucher P, Tremblay A. Nucleotide-resolution mapping of topoisomerase-mediated and apoptotic DNA strand scissions at or near an MLL translocation hotspot. Am J Hum Genet 2006; 79:779–91.
23. Khobta A, Carlo-Stella C, Capranico G. Specific histone patterns and acetylase/deacetylase activity at the breakpoint-cluster region of the human MLL gene. Cancer Res 2004;64:2656–62.
24. Hayashi Y. The molecular genetics of recurring chromosome abnormalities in acute myeloid leukemia. Semin Hematol 2000;37:368–80.
25. Rowley JD, Reshmi S, Sobulo O, Musvee T, Anastasi J, Raimondi S, et al. All patients with the T(11;16)(q23;p13.3) that involves MLL and CBP have treatment-related hematologic disorders. Blood 1997;90: 535–41.
26. Zhang Y, Zeleznik-Le N, Emmanuel N, Jayathilaka N, Chen J, Strissel P, et al. Characterization of genomic breakpoints in MLL and CBP in leukemia patients with t(11;16). Genes Chromosomes Cancer 2004; 41:257–65.
27. Sobulo OM, Borrow J, Tomek R, Reshmi S, Harden A, Schlegelberger B, et al. MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). Proc Natl Acad Sci U S A 1997;94:8732–7.
28. Ohnishi H, Taki T, Yoshino H, Takita J, Ida K, Ishii M, et al. A complex t (1;22;11)(q44;q13;q23) translocation causing MLL-p300 fusion gene in therapy-related acute myeloid leukemia. Eur J Haematol 2008; 81:475–80.
29. Ida K, Kitabayashi I, Taki T, Taniwaki M, Noro K, Yamamoto M, et al. Adenoviral E1A-associated protein p300 is involved in acute myeloid leukemia with t(11;22)(q23;q13). Blood 1997;90:4699–704.
30. Shigeno K, Yoshida H, Pan L, Luo JM, Fujisawa S, Naito K, et al. Disease-related potential of mutations in transcriptional cofactors CREB-binding protein and p300 in leukemias. Cancer Lett 2004; 213:11–20.
31. Tini M, Benecke A, Um SJ, Torchia J, Evans RM, Chambon P, et al. Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. Mol Cell 2002;9:265–77.
32. Datta A, Bagchi S, Nag A, Shiyanov P, Adami GR, Yoon T, et al. The p48 subunit of the damaged-DNA binding protein DDB associates with the CBP/p300 family of histone acetyltransferase. Mutat Res 2001;486:89–97.
33. Miller RW, Rubinstein JH. Tumors in Rubinstein-Taybi syndrome. Am J Med Genet 1995;56:112–5.
34. Rebel VI, Kung AL, Tanner EA, Yang H, Bronson RT, Livingston DM, et al. Distinct roles for CREB-binding protein and p300 in hematopoietic stem cell self-renewal. Proc Natl Acad Sci U S A 2002;99: 14789–94.
35. NCBI dbSNP database. Available from: http://www.ncbi.nlm.nih.gov/SNP/
36. CHIP bioinformatics tool. Available from: http://snpper.chip.org/
37. F-SNP. Available from: http://compbio.cs.queensu.ca/F-SNP/
38. Pupasuite. Available from: http://pupasuite.bioinfo.cipf.es/
39. SNPinfo Web Server. Available from: http://manticore.niehs.nih.gov/
40. Monzo M, Brunet S, Urbano-Ispizua A, Navarro A, Perea G, Esteve J, et al. Genomic polymorphisms provide prognostic information in intermediate-risk acute myeloblastic leukemia. Blood 2006;107: 4871–9.
41. Haploview. Available from: http://www.broad.mit.edu/mpg/haploview/
42. French B, Lumley T, Monks SA, Rice KM, Hindorff LA, Reiner AP, et al. Simple estimates of haplotype relative risks in case-control data. Genet Epidemiol 2006;30:485–94.
43. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003;100:9440–5.
44. Beuten J, Halder I, Fowler S, Göring H, Duggirala R, Arya R, et al. (2011). Wide disparity in genetic admixture among Mexican Americans from San Antonio, TX. Ann Hum Genet (in press).
45. Gallegos-Arreola MP, González-García JR, Figuera LE, Puebla-Pérez AM, Delgado-Lamas JL, Zúñiga-González GM, et al. Distribution of CYP1A1*2A polymorphism in adult patients with acute lymphoblastic leukemia in a Mexican population. Blood Cells Mol Dis 2008;41:91–4.
46. Harvey RC, Mullighan CG, Chen IM, Wharton W, Mikhail FM, Carroll AJ, et al. Rearrangement of CRLF2 is associated with mutation of JAK kinases, alteration of IKZF1, Hispanic/Latino ethnicity, and a poor outcome in pediatric B-progenitor acute lymphoblastic leukemia. Blood 2010;115:5312–21.
47. Swinney RM, Hsu SC, Hirschman BA, Chen TT, Tomlinson GE. MDM2 promoter variation and age of diagnosis of acute lymphoblastic leukemia. Leukemia 2005;19:1996–8.
48. Taja-Chayeb L, Vidal-Millán S, Gutiérrez O, Ostrosky-Wegman P, Dueñas-González A, Candelaria M, et al. Thiopurine S-methyltransferase gene (TMPT) polymorphisms in a Mexican population of healthy individuals and leukemic patients. Med Oncol 2008;25:56–62.
49. Le H, Singh S, Shih SJ, Du N, Schnyder S, Loredo GA, et al. Rearrangements of the MLL gene are influenced by DNA secondary structure, potentially mediated by topoisomerase II binding. Genes Chromosomes Cancer 2009;48:806–15.
50. Scharf S, Zech J, Bursen A, Schraets D, Oliver PL, Kliem S, et al. Transcription linked to recombination: a gene-internal promoter coincides with the recombination hot spot II of the human MLL gene. Oncogene 2007;26:1361–71.
51. Hunkapiller T, Hood L. The growing immunoglobulin gene superfamily. Nature 1986;323:15–6.
52. Siemiatycki J, Thomas DC. Biological models and statistical interactions: an example from multistage carcinogenesis. Int J Epidemiol 1981;10:383–7.
53. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. Nat Genet 2009;41:1006–10.
54. Treviño LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nat Genet 2009;41:1001–5.
55. Yang JJ, Bhojwani D, Yang W, Cai X, Stocco G, Crews K, et al. Genome-wide copy number profiling reveals molecular evolution from diagnosis to relapse in childhood acute lymphoblastic leukemia. Blood 2008;112:4178–83.