Title

Genomic ancestry and somatic alterations correlate with age at diagnosis in Hispanic children with B-cell acute lymphoblastic leukemia

Journal

American Journal of Hematology, 89(7)

ISSN

0361-8609

Authors

Walsh, Kyle M
de Smith, Adam J
Welch, Tara C
et al.

Publication Date

2014-07-01

DOI

10.1002/ajh.23727

Peer reviewed

Published in final edited form as:
Am J Hematol. 2014 July ; 89(7): 721–725. doi:10.1002/ajh.23727.

Genomic ancestry and somatic alterations correlate with age at diagnosis in Hispanic children with B-cell ALL

Kyle M. Walsh[1,2], **Adam J. de Smith**[3], **Tara C. Welch**[2], **Ivan Smirnov**[2], **Marc J. Cunningham**[3], **Xiaomei Ma**[4], **Anand P. Chokkalingam**[5], **Gary V. Dahl**[6], **William Roberts**[7], **Lisa F. Barcellos**[5], **Patricia A. Buffler**[5], **Catherine Metayer**[5], and **Joseph L. Wiemels**[3]

[1] Division of Neuroepidemiology, Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA 94143, USA

[2] Program in Cancer Genetics, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA

[3] Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco , CA 94143, USA

[4] Division of Chronic Disease Epidemiology, Yale University School of Public Health, New Haven, CT 06510, USA

[5] School of Public Health, University of California Berkeley, Berkeley, CA 94704, USA

[6] Division of Pediatric Hematology/Oncology, Department of Pediatrics, Stanford School of Medicine, Stanford, CA 94304, USA.

[7] Department of Pediatric Hematology/Oncology, Rady Children's Hospital, University of California, San Diego, San Diego, California

Abstract

Background—Hispanic children have a higher incidence of acute lymphoblastic leukemia (ALL) than non-Hispanic whites, but tend to be diagnosed at older ages. In genome-wide association studies, Native American ancestry and polymorphisms in six genes have been associated with ALL risk.

Methods—In multivariable regression models, we investigated whether genomic ancestry, inherited risk SNPs, or acquired somatic alterations were associated with differences in age at diagnosis in Hispanic children with B-cell ALL. Genome-wide array data were used to estimate each participant's percent membership in the three Hispanic ancestral populations: Native American, African, and European.

Results—Each 20% increase in European ancestry was associated with a six month younger age at diagnosis (95% CI=0.36-11.6 months, P=0.037). Correspondingly, each 20% increase in Native American ancestry was associated with a six month older age at diagnosis (P=0.037). Both the

CORRESPONDENCE TO: Kyle M. Walsh, PhD UCSF Helen Diller Family Cancer Center, Box 0520 1450 3rd Street HD276 San Francisco, CA 94143-0520 Kyle.Walsh@ucsf.edu.

CONFLICT OF INTEREST STATEMENT:
The authors do not have any conflicts of interest to declare, financial or otherwise.

TEL-AML1 translocation and high-hyperdiploidy were associated with younger age at diagnosis (24.4 months, $P=2.0\times10^{-4}$ and 12.4 months, P=0.011, respectively), while *CDKN2A* and *IKZF1* deletions were associated with older age at diagnosis (19.7 months, $P=7.0\times10^{-4}$ and 18.1 months, P=0.012, respectively). No associations with age at diagnosis were observed for *RAS* mutation, *PAX5* deletion or for known heritable risk alleles in *IKZF1, CDKN2A, PIP4K2A, GATA3, ARID5B* or *CEBPE*.

Conclusion—Because younger age at diagnosis is associated with improved treatment outcomes for children with ALL, the effect of European ancestry on ALL survival may be mediated by its effect on age at diagnosis, or by proxy, its association with more treatable molecular subtypes of ALL.

Keywords

Childhood acute lymphoblastic leukemia; Hispanics; ancestry; genetic association; admixture

INTRODUCTION

Hispanic children have a 10-30% greater incidence rate of acute lymphoblastic leukemia (ALL) than non-Hispanic whites (1) and also display a trend toward older ages at diagnosis (2). Studies indicate that age at diagnosis is a strong predictor of survival, with children diagnosed before 10 years of age having better prognosis than those diagnosed later in life (3). Hispanic children often require more intensive ALL treatment regimens than non-Hispanic whites, possibly due to older ages at diagnosis, poor cytogenetic factors, or as-yet unidentified genetic factors (4).

Using data from 10,272 children diagnosed with ALL and recorded by the U.S. SEER program (1988-2008), Goggins *et al.* showed that just 26.1% of non-Hispanic white children with ALL were diagnosed after their tenth birthday. During this same time period, 33.3% of children of Puerto Rican descent were diagnosed after their tenth birthday. An even higher proportion of U.S. children of Mexican and Central-American descent were in this high-risk age group (42.5% and 44.7%, respectively) (5). Similar to previous reports, an elevated proportion (32.6%) of African-American children were also in this highest risk age group (6, 7). These epidemiologic observations suggest that European ancestry may correlate with younger ages at diagnosis. An alternative explanation is that age at diagnosis is, at least in part, a function of socioeconomic status and access to healthcare. While access to healthcare undoubtedly influences survivorship, its relationship to age at diagnosis for an acute malignancy such as ALL is less clear.

In addition to differences in age at diagnosis, the frequency of specific cytogenetic abnormalities differs by ethnicity. Hispanic ALL patients from California have been shown to have a lower frequency of *TEL-AML1 (ETV6-RUNX1*) translocations and a higher frequency of high-hyperdiploid ALL compared to non-Hispanic whites (8). These patterns have also been observed in Hispanic populations in Central America (2). Because the *TEL-AML1* translocation is most common in children under 10 years of age, it is possible that cytogenetic profiles, age at diagnosis, and genetic ancestry are interrelated in ALL etiology (9).

Genome-wide association studies (GWAS) have identified B-cell ALL (B-ALL) risk alleles in six genes (10, 11), all of which have been replicated in Hispanic populations (12-14). The number of risk alleles at five of these genes (*CDKN2A*, *PIP4K2A*, *GATA3*, *ARID5B* and *CEBPE*) is positively and significantly correlated with increased Native American Ancestry (12, 14, 15). As a corollary to these observations, increased Native American ancestry has been associated with both B-ALL risk and relapse in Hispanic populations (4, 12). These observations support the hypothesis that genomic ancestry underlies mechanisms of leukemogenesis.

To identify genetic correlates of age at diagnosis in children with B-cell ALL, we conducted a case-only analysis of Hispanic participants from the California Childhood Leukemia Study (CCLS) who previously underwent genome-wide SNP genotyping as part of a GWAS of ALL risk (12). We investigated whether genomewide ancestry is associated with differences in age at diagnosis among Hispanic children with B-cell ALL. We further investigated whether ALL risk loci detected in previous GWAS, or the combination of these risk alleles, was associated with age at diagnosis in our study population. Finally, we tested for differences in age at diagnosis among cases with common acquired molecular and cytogenetic abnormalities, including: hyperdiploidy, *RAS* gene mutations, *TEL-AML1* translocation, *CDKN2A* deletion, *IKZF1* deletion, and *PAX5* deletion.

MATERIALS and METHODS

Ethics Statement

This study was approved and reviewed by all collaborating institutions, including the institutional review committees at the California Department of Public Health (CDPH) and the University of California, Berkeley. All parent respondents provided written informed consent.

Study Population

Hispanic participants from the CCLS are included in this study. Enrollment and recruitment processes have been described previously (16). This population-based study comprises subjects recruited between 1995 and 2008 from 35 Central and Northern California counties. Newly diagnosed subjects with leukemia were recruited within ~72 hours of diagnosis at nine area hospitals. The CCLS captures ~95% of children diagnosed with leukemia at these hospitals. In the 35-county study region, a comparison of case ascertainment with the California Cancer Registry indicated that this represented ~76% of all cases diagnosed during this time period (1997- 2003).

Eligibility criteria included: 1) no prior diagnosis of cancer 2) younger than 15 years of age at time of diagnosis 3) at least one parent reporting Hispanic ethnicity 4) residency in the study area. For this study of age at diagnosis, eight children diagnosed with infant leukemia (age <1 year) were excluded from analyses. Additionally, nine children with constitutional trisomy 21 (complete or partial) were excluded from analyses due to differing distributions of age at diagnosis among cases with syndrome-associated ALL. Analyses were limited to cases with BALL.

Determination of immuno-phenotype and cytogenetic profiles

Immuno-phenotype was determined for ALL cases using flow cytometry profiles. Those expressing CD2, CD3, CD4, CD5, CD7, or CD8 (≥20%) were classified as T-lineage and those positive for CD10 or CD19 (≥20%) were classified as B-lineage, as previously described (8). Ploidy was determined using FISH or G-banding, and *TEL-AML1* translocations were identified by fusion of the *ETV6* and *RUNX1* loci.

To identify *RAS* mutations, genomic DNA was extracted from patient bone marrow samples using the QIAamp DNA Mini Kit (QIAGEN, USA, Valencia, CA). PCR amplifications of exon 1 of the *KRAS* and *NRAS* genes were performed in separate reactions using the following sets of oligonucleotide primers: *KRAS*_F: GGTCCTGCACCAGTAATATGC, *KRAS*_R: CTTAAGCGTCGATGGAGGAG; and *NRAS*_F: TCCGACAAGTGAGAGACAGG, *NRAS*_R: TGGAAGGTCACACTAGGGTTT. Bidirectional sequencing of PCR products was carried out on a 3730xl DNA Analyzer (Applied Biosystems), with sequence data compared to the human reference genome to confirm presence/absence of *KRAS* or *NRAS* exon 1 mutations.

Multiplex ligation-dependent probe amplification (MLPA) was carried out using the SALSA MLPA probemix P335-B1 ALL-IKZF1 (MRC Holland, The Netherlands) to identify somatic loss of genes commonly deleted in ALL. MLPA was performed as previously described for B-ALL samples with sufficient bone marrow DNA available (17). Analysis of MLPA data was carried out using the Coffalyser.NET fragment analysis software (MRC Holland) and any alteration found in >10% of cases (*CDKN2A*, *IKZF1*, and *PAX5*) was included in age at diagnosis analyses.

Genotyping

Genome-wide genotyping was performed by the UC Berkeley School of Public Health Genetic Epidemiology and Genomics Laboratory using the Illumina Human OmniExpress v1 array. DNA was extracted from dried bloodspots collected at birth and archived by the California Department of Public Health Genetic Diseases Screening Program. Newborn bloodspots were available for 87% of interviewed participants.

Ten duplicate samples were used to verify genotype reproducibility, with average concordance across all genotypes > 99.99%. Samples were excluded from analyses when genotyping call rates were <98%. Samples were screened for cryptic relatedness using 10,000 unlinked SNPs, and were excluded if the proportion of the genome identical-by-descent was >0.15. Samples with contradictory sex information (reported vs. genotyped sex) were excluded from further analyses.

SNPs with genotyping call rates <98% were not included in analyses. In order to exclude poorly genotyped SNPs, any SNP with a Hardy-Weinberg Equilibrium (HWE) p-value $<1\times10^{-5}$ (in a set of 454 Hispanic controls) was eliminated from analyses. In addition, SNPs were excluded from analysis if they had minor allele frequencies <2.0%. All quality-control was performed using Plink v1.07 (18).

Calculation of individuals' ancestral components

From the full set of SNPs genotyped in the Hispanic cases, a linkage-reduced set of 63,303 autosomal SNPs, evenly distributed across the genome, was extracted. This SNP set was produced using a window size of 50 SNPs, sliding the window in 5-SNP steps, with a Variance-Inflation-Threshold of 1.43 (corresponding to R^2=0.30). This same linkage-reduced set of SNPs was extracted From the Human Genome Diversity Project (HGDP), previously generated using an Illumina 650K array (19). All SNPs passed previously mentioned filters for quality-control. These 63,303 SNPs included 129 empirically tested ancestry-informative markers (AIMs), chosen to distinguish between Native American, African and European populations as previously described (20).

Percent membership of study participants in the three Hispanic founder populations (Native American, sub-Saharan African and European) was evaluated using Structure v2.3.1 (21). Percent membership can be superior to computing eigenvectors in circumstances where >2 ancestral populations contribute to the genomic structure of a population, as in Hispanics, because it distinguishes membership in one group from membership in all others simultaneously (22). SNP data from 372 HGDP individuals was used to define founder population allele frequencies. This included 154 Europeans (Caucasus Mountain Russian, Italian, Sardinian, French, Orkney, Basque), 111 Africans (Yoruban, Bioko, Mbuti, Bantu, San, Mandinka) and 107 Native Americans (Surui, Mayan, Karitana, Colombian, Pima).

Selection of SNPs in validated ALL risk loci

Previous GWAS conducted in populations of European-ancestry have identified ALL risk loci in six regions (10, 11). Here we report associations with age at diagnosis for six SNPs that were successfully genotyped on the OmniExpress platform and which attained genome-wide significance in a previously published GWAS. Thus, we report associations between age at diagnosis and one only SNP in each risk locus.

Statistical analyses

Relationships between heritable ALL risk factors and age at diagnosis were assessed one at a time in multivariable linear regression models, adjusted for: sex, income, African ancestry, hyperdiploidy, and *TEL-AML1* status. Ancestral proportion (*e.g.* %European or %Native American) was modeled as a continuous covariate, ranging from 0-1. Coefficients were calculated to correspond to the difference in age at diagnosis (in months) experienced for each 20% increase in an ancestral proportion. Relationships between somatic alterations and age at diagnosis were assessed simultaneously in a single multivariable linear regression model, adjusted for: sex, income, African ancestry and European ancestry. For all linear regression models, residual plots, including normal probability plots, were examined for departures from normality, excess skew, and kurtosis.

The relationship between age at diagnosis and ALL risk SNPs was assessed using an allelic additive model, in which regression coefficients represent the effect of each additional copy of the risk allele (23). To examine the effect of the total burden of inherited risk SNPs on age at diagnosis, the number of risk alleles per individual at all six independent GWAS risk

loci were summed to produce a global genetic risk score (range 0-12). All reported P-values are 2-sided and all statistical tests were conducted using SAS v9.1.3.

RESULTS

After excluding one sample with discordant reported vs. genotyped sex, 280 Hispanic children with B-cell ALL remained for analyses. Tests for cryptic relatedness did not reveal any excess allele sharing among study subjects. Cases were more likely to be male than female, and had a mean age at diagnosis of 5.3 years. Demographic and cytogenetic characteristics of participants are shown in Table 1, including distribution of genomic ancestry (Table 1 and Figure 1), and are consistent with previously reported distributions in Hispanics (2, 8).

In multivariable regression analyses of heritable factors previously associated with ALL risk, each 20% increase in European ancestry was associated with a statistically significant six month younger age at diagnosis (95% CI = 0.36-11.57 months, P = 0.037). Correspondingly, each 20% increase in Native American ancestry was associated with a statistically significant six month older age at diagnosis (Table 2). Adjustment for *TEL-AML1* status, hyperdiploidy, *CDKN2A* deletion and *IKZF1* deletion did not substantially attenuate these ancestry associations (5.25 month decrease, P = 0.079 and 5.26 month increase, P = 0.079). This is expected, as ancestry did not differ by patient gender or by molecular subgroups (Table S1). Associations between genetic ancestry and age at diagnosis were modestly strengthened when modeled with adjustment for the six ALL risk loci identified by previous GWAS (6.45 month decrease, P = 0.028 and 6.50 month increase, P = 0.029).

With the exception of rs3731217 in *CDKN2A* and rs3824662 in *GATA3*, carrying an increased number of risk alleles at the known ALL risk loci was associated with a younger age at diagnosis. However, none of these SNP associations reached statistical significance (Table 2), nor was the global genetic risk score significantly associated with age at diagnosis (95% CI = -3.37-2.26 months, P = 0.70). Associations between SNPs and age at diagnosis were adjusted for sex, percent African ancestry, and the major cytogenetic subtypes. These associations remained null when analyses were stratified by ploidy or by *TEL-AML1* status.

In regression analyses investigating the influence of cytogenetic abnormalities on age at diagnosis in Hispanic patients with B-cell ALL, the presence of the *TEL-AML1* mutation was associated with a statistically significant 24.4 month reduction in age at diagnosis (95% CI = 11.5-37.4 months, $P = 2.0 \times 10^{-4}$). The presence of hyperdiploidy (51+ chromosomes) was also associated with a statistically significant 12.4 month reduction in age at diagnosis (95% CI = 2.8- 21.9 months, P = 0.011). *CDKN2A* deletion was associated with a statistically significant 19.7 month older age at diagnosis (95% CI = 8.4-31.0 months, $P = 7.4 \times 10^{-4}$). Similarly, *IKZF1* deletion was associated with a statistically significant 18.1 month older age at diagnosis (95% CI = 3.9-32.2 months, P = 0.012). No significant associations were observed between *RAS* mutation or *PAX5* deletion and age at diagnosis (Table 3).

DISCUSSION

This is the first study of childhood ALL to investigate the effects of genomic ancestry on age at diagnosis. Using genome-wide SNP data to assess genomic ancestry, each 20% increase in the proportion of an individual's genome that is of European origin was associated with an approximately 6-month younger age at diagnosis in Hispanic subjects. Importantly, we show that the effect of genomic ancestry on age at diagnosis is independent of population-level differences in the frequency of common ALL-associated somatic alterations (e.g. *TEL-AML1*, hyperdiploidy, *CDKN2A* deletion, *IKZF1* deletion) and known heritable ALL risk variants (*IKZF1, CDKN2A, PIP4K2A, GATA3, ARID5B* or *CEBPE*).

Although the regression model adjusted for differences in parental income, it remains possible that the observed association between ancestry and age at diagnosis is partly attributable to ancestry-related differences in access to healthcare among Hispanics in California. Due to the acute nature of ALL as a disease entity, we believe that access to healthcare is unlikely to explain the sizable ancestry-related differences in age at diagnosis. However, access to healthcare and socioeconomic factors that influence adherence to treatment protocols likely explain a more substantial proportion of the ancestry-related differences in overall survival that have been reported previously.

Population-based data from the United States indicate that Hispanic children with ALL are diagnosed at older ages than non-Hispanic whites. Even within U.S. Hispanics, SEER data indicate that ancestry may underlie more subtle differences in age at diagnosis, as children of Puerto Rican descent are diagnosed at younger ages than children of Mexican or Central American descent (5). Because previous studies have shown that Hispanic children carry an increased burden of ALL risk alleles, this may seem paradoxical (12, 14, 24). However, while Hispanic children have a greater burden of ALL risk alleles, they have also been reported to harbor a different frequency of certain ALL-associated cytogenetic abnormalities (2, 8).

Because *TEL-AML1* translocation and hyperdiploidy are associated with a significantly younger age at diagnosis in our data, and deletions of *CDKN2A* and *IKZF1* with significantly older age at diagnosis, ethnic differences in the distribution of ALL cytogenetic subtypes may underlie inter-population variability in age at diagnosis. However, within Hispanic subjects in our sample, ancestry did not associate with any common cytogenetic/molecular subtypes (*TELAML1*, hyperdiploidy, *RAS* mutation, *IKZF1* deletion, *CDKN2A* deletion, *PAX5* deletion). Thus, it is possible that unmeasured or yet unknown ALL subtypes may underlie the ancestry-associated age differences observed in our sample.

Both *TEL-AML1* translocation and hyperdiploidy were associated with younger age at diagnosis. This aligns well with current theories of leukemogenesis, as *TEL-AML1* translocation and hyperdiploidy appear to be initiating events and have been detected prenatally (25). The development of leukemia in prenatally-initiated cells may be influenced postnatally by risk alleles or environmental exposures which differ across individuals. There is compelling evidence that hematopoiesis is a nonlinear process, with the stem and progenitor cells which arise in different developmental windows possessing distinct

properties (26). Age-associated changes in the number and proliferative capacity of these cells may alter the potential for leukemogenesis (27). Ancestry-related variability in hematopoiesis may therefore influence the rate and timing of leukemia incidence both within and across populations.

Younger age at diagnosis is strongly associated with favorable treatment outcomes for children with ALL. Hispanic children with ALL are frequently older and display less favorable cytogenetic profiles than non-Hispanic whites. Identifying genetic factors underlying differences in age at diagnosis and cytogenetic profiles across populations could help to explain the poorer survival experiences of Hispanic children. While increased Native American ancestry has been associated with poorer outcomes in previous studies of childhood ALL, *GATA3* variation is the only genetic factor shown to directly underlie this association (11). Our data indicate that the *GATA3* locus and somatic *IKZF1* deletions are unable to account for the effect ancestry has on age at diagnosis. This is consistent with results of a recent GWAS of Philadelphia chromosome– like ALL, which revealed that both *GATA3* variation and Native American ancestry independently contribute to risk of this refractory subtype (15). Importantly, we observe that the effect of genomic ancestry on age at diagnosis is independent of known heritable risk loci and somatic alterations, and may further mediate the poor survival experiences of Hispanic children with ALL.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institutes of Health grants: R25CA112355 (KMW), R01CA155461 (JLW, XM), and R01ES009137 (APC, CM, GVD, LFB, JLW, and PAB). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors gratefully acknowledge the clinical investigators at the following collaborating hospitals for help in recruiting patients: 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 Feusner), Kaiser Permanente Roseville (formerly Sacramento) (Drs. Kent Jolly and Vincent Kiley), Kaiser Permanente Santa Clara (Drs. Carolyn Russo, Alan Wong, and Denah Taggart), Kaiser Permanente San Francisco (Dr. Kenneth Leung), and Kaiser Permanente Oakland (Drs. Daniel Kronish and Stacy Month).

REFERENCES

1. Matasar MJ, Ritchie EK, Consedine N, Magai C, Neugut AI. Incidence rates of the major leukemia subtypes among US Hispanics, Blacks, and non-Hispanic Whites. Leuk Lymphoma. 2006; 47:2365–2370. [PubMed: 17107911]
2. Ceppi F, Brown A, Betts DR, Niggli F, Popovic MB. Cytogenetic characterization of childhood acute lymphoblastic leukemia in Nicaragua. Pediatr Blood Cancer. 2009; 53:1238–1241. [PubMed: 19672974]
3. Smith MA, Seibel NL, Altekruse SF, Ries LA, Melbert DL, O'Leary M, Smith FO, Reaman GH. Outcomes for children and adolescents with cancer: challenges for the twenty-first century. J Clin Oncol. 2010; 28:2625–2634. [PubMed: 20404250]
4. Yang JJ, Cheng C, Devidas M, Cao X, Fan Y, Campana D, Yang W, Neale G, Cox NJ, Scheet P, Borowitz MJ, Winick NJ, Martin PL, Willman CL, Bowman WP, Camitta BM, Carroll A, Reaman GH, Carroll WL, Loh M, Hunger SP, Pui CH, Evans WE, Relling MV. Ancestry and

pharmacogenomics of relapse in acute lymphoblastic leukemia. Nat Genet. 2011; 43:237–241. [PubMed: 21297632]

5. Goggins WB, Lo FF. Racial and ethnic disparities in survival of US children with acute lymphoblastic leukemia: evidence from the SEER database 1988-2008. Cancer Causes Control. 2012; 23:737–743. [PubMed: 22450738]
6. Kadan-Lottick NS, Ness KK, Bhatia S, Gurney JG. Survival variability by race and ethnicity in childhood acute lymphoblastic leukemia. JAMA. 2003; 290:2008–2014. [PubMed: 14559954]
7. Bhatia S, Sather HN, Heerema NA, Trigg ME, Gaynon PS, Robison LL. Racial and ethnic differences in survival of children with acute lymphoblastic leukemia. Blood. 2002; 100:1957–1964. [PubMed: 12200352]
8. Aldrich MC, Zhang L, Wiemels JL, Ma X, Loh ML, Metayer C, Selvin S, Feusner J, Smith MT, Buffler PA. Cytogenetics of Hispanic and White children with acute lymphoblastic leukemia in California. Cancer Epidemiol Biomarkers Prev. 2006; 15:578–581. [PubMed: 16537719]
9. Loh ML, Goldwasser MA, Silverman LB, Poon WM, Vattikuti S, Cardoso A, Neuberg DS, Shannon KM, Sallan SE, Gilliland DG. Prospective analysis of TEL/AML1-positive patients treated on Dana-Farber Cancer Institute Consortium Protocol 95-01. Blood. 2006; 107:4508–4513. [PubMed: 16493009]
10. Xu H, Yang W, Perez-Andreu V, Devidas M, Fan Y, Cheng C, Pei D, Scheet P, Burchard EG, Eng C, Huntsman S, Torgerson DG, Dean M, Winick NJ, Martin PL, Camitta BM, Bowman WP, Willman CL, Carroll WL, Mullighan CG, Bhojwani D, Hunger SP, Pui CH, Evans WE, Relling MV, Loh ML, Yang JJ. Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. J Natl Cancer Inst. 2013; 105:733–742. [PubMed: 23512250]
11. Migliorini G, Fiege B, Hosking FJ, Ma Y, Kumar R, Sherborne AL, da Silva Filho MI, Vijayakrishnan J, Koehler R, Thomsen H, Irving JA, Allan JM, Lightfoot T, Roman E, Kinsey SE, Sheridan E, Thompson P, Hoffmann P, Nothen MM, Muhleisen TW, Eisele L, Zimmermann M, Bartram CR, Schrappe M, Greaves M, Stanulla M, Hemminki K, Houlston RS. Variation at 10 p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype. Blood. 2013
12. Walsh KM, Chokkalingam AP, Hsu LI, Metayer C, de Smith AJ, Jacobs DI, Dahl GV, Loh ML, Smirnov IV, Bartley K, Ma X, Wiencke JK, Barcellos LF, Wiemels JL, Buffler PA. Associations between genome-wide Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children. Leukemia. 2013
13. Walsh KM, de Smith AJ, Chokkalingam AP, Metayer C, Dahl GV, Hsu LI, Barcellos LF, Wiemels JL, Buffler PA. Novel childhood ALL susceptibility locus BMI1-PIP4K2A is specifically associated with the hyperdiploid subtype. Blood. 2013; 121:4808–4809. [PubMed: 23744494]
14. Walsh KM, de Smith AJ, Chokkalingam AP, Metayer C, Roberts W, Barcellos LF, Wiemels JL, Buffler PA. GATA3 risk alleles are associated with ancestral components in Hispanic children with ALL. Blood. 2013 in-press.
15. Perez-Andreu V, Roberts KG, Harvey RC, Yang W, Cheng C, Pei D, Xu H, Gastier-Foster J, E S, Lim JY, Chen IM, Fan Y, Devidas M, Borowitz MJ, Smith C, Neale G, Burchard EG, Torgerson DG, Klussmann FA, Villagran CR, Winick NJ, Camitta BM, Raetz E, Wood B, Yue F, Carroll WL, Larsen E, Bowman WP, Loh ML, Dean M, Bhojwani D, Pui CH, Evans WE, Relling MV, Hunger SP, Willman CL, Mullighan CG, Yang JJ. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. Nat Genet. 2013
16. Ma X, Buffler PA, Layefsky M, Does MB, Reynolds P. Control selection strategies in case-control studies of childhood diseases. Am J Epidemiol. 2004; 159:915–921. [PubMed: 15128601]
17. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res. 2002; 30:e57. [PubMed: 12060695]
18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]

19. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM. Worldwide human relationships inferred from genome-wide patterns of variation. Science. 2008; 319:1100–1104. [PubMed: 18292342]

20. Galanter JM, Fernandez-Lopez JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, Via M, Hidalgo-Miranda A, Contreras AV, Figueroa LU, Raska P, Jimenez-Sanchez G, Zolezzi IS, Torres M, Ponte CR, Ruiz Y, Salas A, Nguyen E, Eng C, Borjas L, Zabala W, Barreto G, Gonzalez FR, Ibarra A, Taboada P, Porras L, Moreno F, Bigham A, Gutierrez G, Brutsaert T, Leon-Velarde F, Moore LG, Vargas E, Cruz M, Escobedo J, Rodriguez-Santana J, Rodriguez-Cintron W, Chapela R, Ford JG, Bustamante C, Seminara D, Shriver M, Ziv E, Burchard EG, Haile R, Parra E, Carracedo A. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. PLoS Genet. 2012; 8:e1002554. [PubMed: 22412386]

21. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003; 164:1567–1587. [PubMed: 12930761]

22. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet. 2006; 2:e190. [PubMed: 17194218]

23. Walsh KM, Rice T, Decker PA, Kosel ML, Kollmeyer T, Hansen HM, Zheng S, McCoy LS, Bracci PM, Anderson E, Hsuang G, Wiemels JL, Pico AR, Smirnov I, Molinaro AM, Tihan T, Berger MS, Chang SM, Prados MD, Lachance DH, Sicotte H, Eckel-Passow JE, Wiencke JK, Jenkins RB, Wrensch MR. Genetic variants in telomerase-related genes are associated with an older age at diagnosis in glioma patients: evidence for distinct pathways of gliomagenesis. Neuro Oncol. 2013; 15:1041–1047. [PubMed: 23733245]

24. Xu H, Cheng C, Devidas M, Pei D, Fan Y, Yang W, Neale G, Scheet P, Burchard EG, Torgerson DG, Eng C, Dean M, Antillon F, Winick NJ, Martin PL, Willman CL, Camitta BM, Reaman GH, Carroll WL, Loh M, Evans WE, Pui CH, Hunger SP, Relling MV, Yang JJ. ARID5B genetic polymorphisms contribute to racial disparities in the incidence and treatment outcome of childhood acute lymphoblastic leukemia. J Clin Oncol. 2012; 30:751–757. [PubMed: 22291082]

25. Wiemels JL, Kang M, Chang JS, Zheng L, Kouyoumji C, Zhang L, Smith MT, Scelo G, Metayer C, Buffler P, Wiencke JK. Backtracking RAS mutations in high hyperdiploid childhood acute lymphoblastic leukemia. Blood Cells Mol Dis. 2010; 45:186–191. [PubMed: 20688547]

26. Montecino-Rodriguez E, Dorshkind K. B-1 B cell development in the fetus and adult. Immunity. 2012; 36:13–21. [PubMed: 22284417]

27. Greaves M. Leukemogenesis and ageing: ‘fit for transformation’? Aging (Albany NY). 2011; 3:79–80. [PubMed: 21386130]

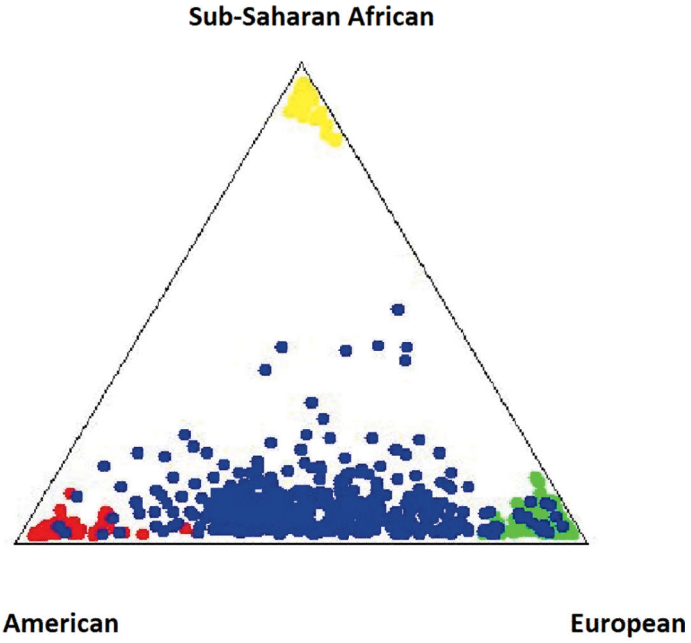

Figure 1. Triangle plot showing estimated admixture in Hispanic CCLS participants with B-cell ALL

Estimates were performed in Structure using 63,303 SNPs from the Human Genome Diversity Project on 111 sub-Saharan African (Yellow dots), 154 European (Green dots) and 107 Native American (Red dots) founders. The figure depicts ancestry in 280 Hispanic B-cell ALL cases (Blue dots).

Table 1

Demographic and tumor characteristics of Hispanic acute lymphoblastic leukemia cases from the California Childhood Leukemia Study appearing in the association analyses.

Sample Size	280
n (%)	
Male	151 (53.9)
B-cell ALL	280(100.0)
Hyperdiploid (51+ chromosomes)	97 (34.6)
TEL-AML1 +	40 (14.3)
RAS mutant	38(20.0*)
CDKN2A (p16) deleted	47 (27.0*)
IKZF1 deleted	28 (16.1*)
PAX5 deleted	39 (22.4*)
Mean (SD)	
Age in years	5.3 (3.2)
% Native American Ancestry	38.4(15.7)
% African Ancestry	6.8(5.8)
% European Ancestry	54.8 (16.2)
Global genetic risk score (risk allele sum)	7.5 (1.6)

* *KRAS* and *NRAS* mutation data were available for 191 patients. *CDKN2A, IKZF1* and *PAX5* data were available for 174 patients.

Table 2

Associations between heritable factors previously associated with ALL risk and age at diagnosis (in months) among Hispanic CCLS participants with B-cell ALL.

Heritable factor previously associated with ALL risk	Effect (95% CI)*	P*
%European ancestry	−5.97 (−0.36, −11.57)	**0.037**
%Native American ancestry	5.98 (0.37, 11.64)	**0.037**
rs4132601-G (*IKZF1*)	−1.06 (−7.68, 5.64)	0.75
rs3731217-T (*CDKN2A*)	4.44 (−9.12, 17.88)	0.52
rs3824662-A (*GATA3*)	2.59 (−3.80, 8.99)	0.52
rs7088318-A (*PIP4K2A*)	−0.85 (−8.04, 6.36)	0.82
rs7089424-C (*ARID5B*)	−2.28 (−8.88, 4.32)	0.50
rs2239633-G (*CEBPE*)	−2.96 (−9.34, 3.41)	0.36
Global genetic risk score (risk allele sum)	−0.56 (−3.37, 2.26)	0.70

*Effect size (measured in months) is generated from a multivariable linear regression model where “age at diagnosis” is the dependent variable, adjusting for: sex, income, %African ancestry, hyperdiploidy, and *TEL-AML1* status. Positive values indicate older age at diagnosis with an increasing number of risk alleles or with each 20% increase in % ancestry. Negative values indicate younger age at diagnosis with an increasing number of risk alleles or with each 20% increase in Native American or European ancestry. P-values are two-sided and are derived from these regression models (Ho: Beta=0).

Table 3

Associations between somatic ALL alterations and age at diagnosis (in months) among Hispanic CCLS participants with B-cell ALL.

Somatic ALL alteration	Effect (95% CI)*	P*
Hyperdiploidy (51+ chromosomes)	−12.4 (−21.9, −2.8)	**0.011**
TEL-AML1 + status	−24.4 (−37.4, −11.5)	**2.0×10^{-4}**
RAS mutation + (*KRAS/NRAS*)	0.76 (−12.8, 14.3)	0.91
CDKN2A (p16) deletion +	19.7 (8.4, 31.0)	**7.0×10^{-4}**
IKZF1 deletion +	18.1 (3.9, 32.2)	**0.012**
PAX5 deletion +	−0.38 (-13.1, 12.3)	0.95

*Effect size (measured in months) is generated from a multivariable linear regression model where “age at diagnosis” is the dependent variable and somatic ALL alterations are modeled jointly, adjusting for: sex, income, %African ancestry and %European ancestry. Positive values indicate that the somatic alteration is associated with an older age at diagnosis. Negative values indicate that the somatic alteration is associated with a younger age at diagnosis. P-values are two-sided and are derived from this regression model (Ho: Beta=0).