

Published in final edited form as:

Leukemia. 2013 December ; 27(12): . doi:10.1038/leu.2013.130.

Associations between genome-wide Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children

KM Walsh^{1,2,8}, AP Chokkalingam^{3,8}, L-I Hsu³, C Metayer³, AJ de Smith⁴, DI Jacobs⁵, GV Dahl⁶, ML Loh⁷, IV Smirnov², K Bartley³, X Ma⁵, JK Wiencke², LF Barcellos³, JL Wiemels⁴, and PA Buffler³

KM Walsh: Kyle.Walsh@ucsf.edu

¹Program in Cancer Genetics, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, CA, USA

²Division of Neuroepidemiology, Department of Neurological Surgery, University of California, San Francisco, CA, USA

³School of Public Health, University of California, Berkeley, CA, USA

⁴Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA

⁵Yale School of Public Health, Yale School of Medicine, New Haven, CT, USA

⁶Division of Pediatric Hematology/Oncology, Department of Pediatrics, Stanford School of Medicine, Stanford, CA, USA

⁷Department of Pediatrics, Benioff Children's Hospital, University of California, San Francisco, CA, USA

Hispanic children have a 10–30% greater incidence rate of ALL than non-Hispanic whites, and nearly double the rate observed in African-Americans.¹ Ethnic differences in ALL incidence may be explained by population-level differences in the frequency of genetic risk factors, including those first discovered in genome-wide association studies of European-ancestry populations.^{2–5} As Hispanics are an admixed population with European, African and Native American ancestry, differences in ALL incidence observed in Hispanics may be attributable to genetic risk factors associated with Native American ancestry.

Increased Native American ancestry has been linked to increased risk of relapse among Hispanic children with ALL,⁶ but no study has yet investigated the contribution of genome-wide Native American ancestry to ALL incidence. Using genome-wide SNP data from 298 Hispanic children with B cell ALL and 456 matched controls from the California Childhood Leukemia Study (CCLS), we investigated whether genome-wide Native-American ancestry was associated with increased risk of B-cell ALL. Additionally, we assessed whether the risk alleles at loci identified in genome-wide association studies of European-ancestry populations (*IKZF1*, *CDKN2A*, *PIP4K2A*, *ARID5B*, *CEBPE*) were more common in individuals with greater levels of Native American ancestry. Finally, we quantified the contribution of these validated risk loci to the increased ALL incidence observed in Hispanics relative to populations of European or African ancestry.

© 2013 Macmillan Publishers Limited All rights reserved

⁸The first two authors should be regarded as joint first authors.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Supplementary Information accompanies this paper on the *Leukemia* website (<http://www.nature.com/leu>)

Study participants were Hispanic children from the CCLS, whose recruitment and enrollment procedures have been described in detail previously (Supplementary Table S1).^{7,8} Cytogenetic characteristics of included cases are shown in Supplementary Table S1. DNA was isolated from dried bloodspots collected at birth and archived by the California Department of Public Health. Samples were genotyped using the Illumina OmniExpress platform, assaying 730 525 single-nucleotide polymorphism (SNPs). Samples with genotyping call rates <98%, with discordant sex information (reported versus genotyped sex), or showing evidence of cryptic relatedness were excluded from analyses. To exclude poorly genotyped SNPs, SNPs with genotyping call rates <98% or Hardy–Weinberg Equilibrium P -value $<1 \times 10^{-5}$ in controls were removed from analyses.

A linkage-reduced set of 63 303 autosomal SNPs, evenly distributed across the genome, was extracted from the case-control data and the Human Genome Diversity Project (HGDP) data. The genetic structure of study subjects was evaluated using Structure v2.3.1 to estimate percent membership in three distinct founder populations: sub-Saharan African, European and Native American.⁹ Founder population allele frequencies were defined using SNP data from 372 unrelated HGDP individuals, including 111 Africans, 107 Native Americans and 154 Europeans.¹⁰

Logistic regression was used to determine if Native American ancestry was associated with case-status, with adjustment for sex, age and risk SNPs (where indicated). Logistic regression was also used to determine if these SNPs were associated with case-status, after adjustment for sex and age. We report results for the five SNPs (one in each risk locus) that achieved genome-wide significance in a previously published genome-wide association study and which were successfully genotyped on our Illumina platform. Although the array data provides genotypes for additional SNPs in these regions, we believed it important to analyse Native American ancestry in relation to risk loci first identified in populations of European-ancestry.

Correlations between Native American ancestry and number of risk alleles in *IKZF1*, *CDKN2A*, *PIP4K2A*, *ARID5B* and *CEBPE* were assessed using Pearson's correlation coefficient. The contribution of known susceptibility loci to ethnic incidence rate ratios were calculated according to varying genotypic relative risks and ethnic group allele frequencies using previously described methods.¹¹ Additional information on samples, genotyping and statistical procedures is available in the Supplementary Methods.

A total of 297 cases and 454 controls passed all quality control filters. Four SNPs identified as ALL risk factors in previous genome-wide association studies were significantly associated with ALL risk in our Hispanic sample (Supplementary Table S2). The strongest association was at rs7089424 in *ARID5B* (odds ratio (OR) = 2.33, 95% confidence interval (CI): 1.85–2.92, $P = 2.6 \times 10^{-14}$). As previously reported,^{2–4} this effect was stronger in hyperdiploid cases (OR = 2.91, 95% CI: 2.05–4.12, $P = 2.1 \times 10^{-10}$). SNP rs2239633 in *CEBPE* was also more strongly associated with hyperdiploid B-cell ALL (OR = 2.07, 95% CI: 1.44–2.98, $P = 8.9 \times 10^{-5}$) than with B-cell ALL not stratified by subtype (OR = 1.35, 95% CI: 1.09–1.68, $P = 6.6 \times 10^{-3}$). Although rs7088318 in *PIP4K2A* was not statistically significantly associated with B-cell ALL risk in our sample (OR = 1.16, 95% CI: 0.92–1.49, $P = 0.21$), the association approached significance among hyperdiploid cases (OR = 1.37, 95% CI: 0.96–1.96, $P = 0.084$). Risk alleles at rs4132601 (*IKZF1*) and rs3731217 (*CDKN2A*) were also strongly associated with B-cell ALL risk in our case-control sample (OR = 1.46, 95% CI: 1.16–1.83, $P = 1.3 \times 10^{-3}$ and OR = 1.76, 95% CI: 1.17–2.65, $P = 4.6 \times 10^{-3}$, respectively).

Compared with controls, cases had higher levels of Native American ancestry and lower levels of European ancestry (Supplementary Table S1 and Supplementary Figure S1). After adjustment for age, sex and percent African ancestry, each 20% increase in Native American ancestry was associated with a 1.20-fold increase in risk of B-cell ALL (OR = 1.20, 95% CI: 1.00–1.45, $P = 0.048$) (Supplementary Table S2). The association between genome-wide Native American ancestry and ALL risk was modestly attenuated when controlling for genotype at rs3731217 (*CDKN2A*), rs7088318 (*PIP4K2A*) and rs2239633 (*CEBPE*) (1, 2.5 and 4.2% decreases, respectively), and was further attenuated when conditioned on genotype at rs7089424 (*ARID5B*, 6.6% decrease) (Supplementary Table S2). These SNPs, in particular rs7089424, may contribute to the observed association between Native American ancestry and ALL risk.

Further support for this was shown when correlations were calculated between Native American ancestry and number of risk alleles at the five ALL risk SNPs. The number of risk alleles at four of these SNPs was positively and significantly correlated with increased Native American ancestry (Table 1). The strongest of these associations were with *ARID5B* and *PIP4K2A* SNPs ($r = 0.13$, $P = 6.0 \times 10^{-4}$ and $r = 0.18$, $P = 2.1 \times 10^{-5}$, respectively). The number of risk alleles at rs3731217 (*CDKN2A*) and rs2239633 (*CEBPE*) was also positively correlated with increased Native American ancestry ($r = 0.11$, $P = 3.6 \times 10^{-3}$ and $r = 0.081$, $P = 0.027$, respectively). These associations were consistent when analyses were restricted to control subjects, indicating that these associations reflect population structure, independent of case-status (Table 1).

We next assessed whether these risk loci contribute to the increased ALL incidence observed in Hispanics relative to populations of European or African ancestry (Table 2). Interestingly, the risk allele of rs3731217 in *CDKN2A* has an allele frequency of 100% in Native Americans. Despite the absence of the minor (protective) allele in this population, this SNP explains only a small proportion of the increased B-cell ALL risk observed in Hispanics compared with European or African-ancestry populations.

Previously identified risk alleles in *CEBPE*, *PIP4K2A* and *ARID5B* are also more common in Native American and Hispanic populations than in Europeans. SNP rs2239633 in *CEBPE* accounted for a 1.03-fold increased risk of B-cell ALL in Hispanics versus Caucasians (95% CI: 1.002–1.067). In addition, rs7089424 in *ARID5B* accounted for a 1.11-fold increased risk of B-cell ALL in Hispanics versus Caucasians (95% CI: 1.005–1.212) (Table 2). As this SNP is more strongly associated with hyperdiploid B-cell ALL than with other subtypes, it can explain an even larger proportion of the differences observed across populations in the incidence of this ALL subtype (Supplementary Table S3).

Our findings suggest that the increased risk of B-cell ALL observed in Hispanic populations is due, at least in part, to an effect of Native American ancestry. In our sample, each 20% increase in the proportion of an individual's genome that is of Native American origin conferred a 1.20-fold increased risk of B-cell ALL. Because increased Native American ancestry was also associated with known ALL risk alleles, even among controls, we believe the increased risk of ALL associated with increased Native American ancestry is not easily attributed to potential confounding factors.

Taken together, the risk alleles in *CDKN2A*, *PIP4K2A*, *CEBPE* and *ARID5B* may account for an important proportion of the ALL incidence differences observed across ethnicities. Although these variants are associated with ALL risk in numerous populations,^{5,12–14} their increased frequency in populations with Native American ancestry may result from a founder effect occurring during migration to the New World and genetic drift during subsequent population expansion.

As a corollary to the positive association between Native American ancestry and ALL risk, increased European ancestry is associated with decreased B-cell ALL risk in this Hispanic sample. However, were European ancestry protective, both Hispanic and African-American populations would be expected to have higher ALL incidence than European populations. As African-Americans have lower ALL incidence than Europeans, it appears the Native American component of Hispanic ancestry may be a risk factor, and not that the European component is a protective factor. This is further corroborated by our observations that known risk alleles in *CDKN2A*, *PIP4K2A*, *CEBPE* and *ARID5B* were all significantly associated with increased Native American ancestry.

In conclusion, we demonstrate that increased genome-wide Native American ancestry is associated with an increased risk of B-cell ALL in Hispanic children, and trace this to the effects of at least three genes. Additional questions remain as to whether the known risk loci can account for all of the increased B-cell ALL risk observed in Hispanics, or if additional risk loci can be identified through further study of this high-risk population.

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), R01CA126831 (JKW) and R01ES009137 (APC, LH, CM, GVD, MLL, KB, LFB, JLW, and PAB). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1. Yamamoto JF, Goodman MT. Patterns of leukemia incidence in the United States by subtype and demographic characteristics, 1997–2002. *Cancer Causes Control*. 2008; 19:379–390. [PubMed: 18064533]
2. Trevino 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; 41:1001–1005. [PubMed: 19684603]
3. 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; 41:1006–1010. [PubMed: 19684604]
4. Sherborne AL, Hosking FJ, Prasad RB, Kumar R, Koehler R, Vijayakrishnan J, et al. Variation in *CDKN2A* at 9p21. 3 influences childhood acute lymphoblastic leukemia risk. *Nat Genet*. 2010; 42:492–494. [PubMed: 20453839]
5. Xu H, Yang W, Perez-Andreu V, Devidas M, Fan Y, Cheng C, et al. Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations. *J Natl Cancer Inst*. 2013 e-pub ahead of print 9 March 2013.
6. Yang JJ, Cheng C, Devidas M, Cao X, Fan Y, Campana D, et al. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nat Genet*. 2012; 43:237–241. [PubMed: 21297632]
7. 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]
8. Aldrich MC, Zhang L, Wiemels JL, Ma X, Loh ML, Metayer C, et al. Cytogenetics of Hispanic and White children with acute lymphoblastic leukemia in California. *Cancer Epidemiol Biomarkers Prev*. 2006; 15:578–581. [PubMed: 16537719]
9. 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]

10. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 2008; 319:1100–1104. [PubMed: 18292342]
11. Jacobs DI, Walsh KM, Wrensch M, Wiencke J, Jenkins R, Houlston RS, et al. Leveraging ethnic group incidence variation to investigate genetic susceptibility to glioma: a novel candidate SNP approach. *Front Genet*. 2012; 3:203. [PubMed: 23091480]
12. Vijayakrishnan J, Sherborne AL, Sawangpanich R, Hongeng S, Houlston RS, Pakakasama S. Variation at 7p12.2 and 10q21.2 influences childhood acute lymphoblastic leukemia risk in the Thai population and may contribute to racial differences in leukemia incidence. *Leuk Lymphoma*. 2010; 51:1870–1874. [PubMed: 20919861]
13. Yang W, Trevino LR, Yang JJ, Scheet P, Pui CH, Evans WE, et al. ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence. *Leukemia*. 2010; 24:894–896. [PubMed: 20054350]
14. Xu H, Cheng C, Devidas M, Pei D, Fan Y, Yang W, et al. 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]

Table 1

Correlation coefficients for number of risk alleles at known ALL risk loci and percent membership in each of three ancestral populations among CCLS controls, cases and combined sample

	<u>rs4132601-G (IKZFL)</u>	<u>rs3731217-T (CDKN2A)</u>	<u>rs7088318-A (PIP4K2A)</u>	<u>rs7089424-C (ARID5B)</u>	<u>rs2239633-G (CEBPE)</u>					
	r	P	r	P	r	P				
<i>Controls only</i>										
% Native American Ancestry	0.0088	0.85	0.12	0.014	0.18	7.4×10^{-5}	0.13	0.0060	0.098	0.027
% European Ancestry	-0.0018	0.97	-0.11	0.018	-0.18	7.5×10^{-5}	-0.11	0.015	-0.097	0.039
% African Ancestry	-0.021	0.65	0.001	0.91	0.034	0.46	-0.027	0.56	-0.003	0.94
<i>Cases only</i>										
% Native American Ancestry	-0.12	0.040	0.070	0.23	0.16	0.0044	0.082	0.16	0.031	0.60
% European Ancestry	0.067	0.25	-0.058	0.32	-0.12	0.039	-0.077	0.19	-0.055	0.35
% African Ancestry	0.14	0.018	-0.029	0.62	-0.11	0.051	-0.0083	0.89	0.072	0.22
<i>Cases and controls</i>										
% Native American Ancestry	-0.035	0.35	0.11	0.0036	0.18	2.1×10^{-5}	0.13	0.00060	0.081	0.027
% European Ancestry	0.016	0.65	-0.10	0.0061	-0.16	4.6×10^{-5}	-0.12	0.0013	-0.087	0.017
% African Ancestry	0.054	0.14	-0.0037	0.92	-0.025	0.50	-0.080	0.83	0.033	0.37

Nominally significant *P*-values (<0.05) appear in bold.

Table 2

SNP effect size, risk allele frequency and contribution to B-cell ALL ethnic incidence rate ratios (IRR) by established susceptibility loci

Risk allele ^a	SNP effect size (95% CI) ^b	Risk allele frequency ^c				Hispanic-Caucasian IRR (95% CI)		Hispanic-African IRR (95% CI)	
		Caucasian	African	Hispanic	Native American				
rs4132601-G (<i>JKZF1</i>)	1.46 (1.16–1.83)	0.301	0.212	0.160	0.224	0.904 (0.843–0.953)	0.953 (0.912–0.987)		
rs3731217-T (<i>CDKN2A</i>)	1.76 (1.17–2.65)	0.863	0.907	0.880	1.000	1.003 (0.981–1.026)	0.974 (0.952–1.005)		
rs7088318-A (<i>PIP4K2A</i>)	1.16 (0.92–1.49)	0.588	0.243	0.727	0.939	1.041 (0.993–1.091)	1.163 (0.976–1.362)		
rs7089424-C (<i>ARID5B</i>)	2.33 (1.85–2.92)	0.304	0.235	0.380	0.570	1.110 (1.005–1.212)	1.236 (1.149–1.343)		
rs2239633-G (<i>CEBPE</i>)	1.35 (1.09–1.68)	0.522	0.832	0.610	0.612	1.031 (1.002–1.067)	0.893 (0.845–0.953)		

Abbreviations: CI, confidence interval; IRR, incidence rate ratio; SNP, single-nucleotide polymorphism.

^a SNPs have previously been reported to increase risk for B-cell ALL in a published genome-wide association study.

^b OR are derived from the CCLS Hispanic case-control study, comparing 297 B-cell ALL cases to 454 controls, adjusted for age, sex and the first five principal components.

^c Allele frequencies for Caucasians, Africans and Native Americans are from Human Genome Diversity Panel data. Allele frequencies for Hispanics are from HapMap data.