Title Page:

Title of the paper: Genetic markers in a multi-ethnic sample for childhood acute lymphoblastic leukemia risk

Running title: Genetic markers for childhood leukemia

Name and affiliation of each author (in order):
Dr. Amy E. Kennedy - Florida International University (Kennedy AE)
Dr. Kala Y. Kamdar - Baylor College of Medicine (Kamdar KY)
Dr. Philip J. Lupo - Baylor College of Medicine (Lupo PJ)
Dr. M. Fatih Okcu - Baylor College of Medicine (Okcu MF)
Dr. Michael E. Scheurer - Baylor College of Medicine (Scheurer ME)
Dr. Mehmet Tevfik Dorak - Florida International University (Dorak MT)

Corresponding author information:

Dr. Mehmet Tevfik Dorak[*], MD, PhD
Florida International University
Robert Stempel College of Public Health & Social Work, Department of Epidemiology
11200 SW 8th Street
AHC II 595-1
Miami, FL 33199
dorakm@hope.ac.uk
Phone: 305-348-0118
Fax: 305-348-4901

[*]Present address:
Professor of Health Sciences
School of Health Sciences, HCA EW 208
Liverpool Hope University
Liverpool L16 9JD
United Kingdom
dorakm@hope.ac.uk
Phone: +44 151 291 3236

Other author's information:

Dr. Amy E. Kennedy**, PhD, MPH
Florida International University
Robert Stempel College of Public Health & Social Work, Department of Epidemiology
11200 SW 8th Street
AHC II 595-1
Miami, FL 33199
akenn001@fiu.edu
Phone: 305-348-0118

Fax: 305-348-4901

**Present address:
National Institutes of Health
National Cancer Institute
Epidemiology and Genomics Research Program
Division of Cancer Control and Population Sciences
9609 Medical Center Drive, Rm. 4E114, MSC 9763
Rockville, MD 20850
Telephone: (240) 276-6723
Fax: (240) 276-7921
E-mail: amy.kennedy@nih.gov

Dr. Kala Y. Kamdar, MD, MS
Baylor College of Medicine
Department of Pediatrics, Section of Hematology-Oncology
One Baylor Plaza
MS-BCM305
Houston, TX 77030
kykamdar@txch.org
Phone: 832-824-4163
Fax: 832-825-4732

Dr. Philip J. Lupo, PhD, MPH
Baylor College of Medicine
Department of Pediatrics, Section of Hematology-Oncology
One Baylor Plaza
MS-BCM305
Houston, TX 77030
Philip.Lupo@bcm.edu
Phone: 713-798-2960
Fax: 713-798-3658

Dr. M. Fatih Okcu, MD, MPH
Baylor College of Medicine
Department of Pediatrics, Section of Hematology-Oncology
One Baylor Plaza
MS-BCM320
Houston, TX 77030
mfokcu@txch.org
Phone: 832-822-4242
Fax: 832-825-4202

Dr. Michael E. Scheurer, PhD, MPH
Baylor College of Medicine
Department of Pediatrics, Section of Hematology-Oncology

One Baylor Plaza
MS-BCM305
Houston, TX 77030
scheurer@bcm.edu
Phone: 713-798-5547
Fax: 713-798-8711

ABSTRACT

Genome-wide association studies have identified multiple risk loci for childhood acute lymphoblastic leukemia (ALL), but mostly in European/White populations despite Hispanics having a greater risk. We re-examined SNPs of known associations with childhood ALL and known HLA region lymphoma risk markers in a multi-ethnic population. Significant associations were found in two *ARID5B* variants (rs7089424 and rs10821936). We replicated a strong risk association in non-Hispanic White males with rs2395185, a protective marker for lymphoma. Another HLA region marker, rs2647012, showed a risk association among Hispanics only, while a strong protective association was found with rs1048456, a follicular lymphoma risk marker. Our study validated this new case-control sample by confirming genetic markers associated with childhood ALL, and yielded new associations with lymphoma markers. Despite positive results, our study did not provide any clues to why Hispanics have a higher susceptibility to childhood leukemia, suggesting that environmental factors may have a strong contribution.

INTRODUCTION

Over a third of cancers in children are leukemias [1], with acute lymphoblastic leukemia (ALL) being the most common [2-4]. Approximately 3.8 individuals (less than 14 years of age) per 1,000,000 are diagnosed with childhood ALL per year in the United States [5]. The racial disparity in the incidence of childhood ALL has been well established [6,7]. In the United States, incidence of ALL is greatest in Hispanic children, followed by non-Hispanic Whites, Asians, and then Blacks [7-9].

Research continually tries to determine the etiology of childhood leukemias. While there are a few known risk factors associated with childhood ALL [10-15], more than 90% are of unknown etiology [11].Increased birth weight is an established risk factor shown to increase risk of childhood ALL [16-18]. Environmental factors are also involved, and may work in conjunction with genetic factors to cause many cancers [10].

Genome-wide association studies (GWAS) [19-22] have identified multiple risk loci showing significant associations with childhood ALL. Most significantly, variants located within the *ARID5B, IKZF1*, and *CEBPE* genes have robust risk associations. Most of these studies, however, have only evaluated risk among those of European ancestry. Two multiethnic studies conducted by Xu *et al*. looked at various risk loci in both African American and Hispanic American populations [8,22], discovering that some markers are universal across races/ethnicities, while others are race/ethnic-specific.

We hypothesized that the variation in ethnic/racial susceptibility to childhood ALL has a genetic basis. We re-examined five previously discovered single nucleotide polymorphisms (SNPs) of known associations to leukemia (rs7089424, rs10821936, rs10994982, rs4132601, and rs2239633), along with three HLA region susceptibility markers for lymphomas, since

lymphoma and ALL both stem from lymphoid cells. These included rs2395185, a marker of HLA DRB4 lineage [23] which has previously shown associations with major leukemias including childhood ALL [24,25] and other diseases including Hodgkin lymphoma [26], lung cancer [27], rheumatoid arthritis [28], asthma [29], and ulcerative colitis [30-32]; rs10484561, which has been shown to be a strong risk marker in follicular lymphoma (FL) [33]; and rs2647012, which is a protective marker for FL [34]. These SNPs were examined in a multi-ethnic sample (non-Hispanic Whites, Hispanic Whites, and Blacks) from Houston, Texas to assess their association with ALL.

SUBJECTS AND METHODS

Study population

Institutional Review Board (IRB) approval was obtained at both the Baylor College of Medicine (BCM) and Florida International University prior to the start of the study. The case-control study was comprised of 161 incident childhood ALL cases and 231 healthy frequency-matched controls for gender, contemporaneously recruited at the Texas Children's Cancer Center in Houston, TX from 2007 to 2012. The children were less than 18 years of age at diagnosis, and exclusion criteria for both cases and controls were refusal to participate in the study and the diagnosis of any other disease or cancer. Subjects or their parents provided informed consent for provision of epidemiological data with a questionnaire and a biological sample. The DNA samples were extracted from saliva or peripheral blood samples at BCM. Race/ethnicity was determined by the responses provided on the questionnaire. Parents were requested to state the race (White, Black/African American, Asian, American Indian/Alaska Native, or Native Hawaiian/Other Pacific Islander) and ethnicity (Hispanic/Latino or non-Hispanic/Latino) of the

child. The questionnaire also collected the race/ethnicity of the parents, which was used to verify the response. Information on clinical subtype of ALL was collected from medical records.

Genotyping

The main features for the SNPs genotyped are shown in Table I. Pre-developed TaqMan® SNP Genotyping Assays (LifeTech, Foster City, CA) were used for all of the SNPs we examined. Genotyping was achieved using the Bio-Rad CFX96 real-time PCR machine (Hercules, CA). The TaqMan assays consist of singleplex reactions carried out in ninety-six well plates. Each plate contained two no template controls (NTCs), a positive control, and random replicate samples. Bio-Rad SsoFast™ Probes Supermix, a 2x reaction buffer which contains the necessary components for running the PCR; Sso7d-fusion polymerase, dNTPs, $MgCI_2$, and stabilizers, was used with the TaqMan Assay. PCR amplifications were performed using the manufacturer's suggestion of 20 μL total volume and with the following PCR thermal cycling conditions: enzyme activation at 95^oC for two minutes, and 49 cycles of denaturation at 95^oC for 5 seconds followed by annealing and extension at 61^oC for 5 seconds. Bio-Rad CFX Manager software (version 3.0) was used for data acquisition and genotype assignment.

To adjust for heterogeneity in our sample, especially in Hispanics, we used two ancestry-informative markers, AIMs, to control for potential confounding caused by population stratification and to avoid spurious associations [35]. The two AIMs we used, rs285 and rs2891, have been previously used in Hispanic populations to account for the differences in genetic ancestry [36,37]. The risk associations were adjusted by each of the AIMs besides adjustment by self-declared race/ethnicity.

Statistical analysis

Statistical analyses were performed using Stata v.11 (StataCorp, College Station, TX). Pearson's X^2, Student's t-test (for means) or median test (for medians) were used to compare characteristics between the cases and controls. Logistic regression methods were used to calculate crude and adjusted odds ratios (OR) and 95% confidence intervals (CIs). All statistical tests were two-tailed, and the threshold for statistical significance was set at $P \leq 0.05$. The ORs, with 95% CIs, were used as a measure of effect size. Genotype counts were tested for Hardy–Weinberg equilibrium (HWE) in controls for each SNP. By default, we used the additive genetic model to assess associations by Cochrane-Armitage trend test. Due to its previous association being in the recessive model, rs2395185 was analyzed also for the recessive model association. Ethnic- and gender-specific associations were calculated through stratified analyses after grouping subjects as non-Hispanic White, Hispanic White, Blacks, and Others. Associations in the overall sample were assessed by adjustment for race/ethnicity as well as by Mantel-Haenszel analysis after stratification. None of the SNPs were located in coding regions, therefore we used RegulomeDB (Stanford University, Palo Alto, CA) to determine their scores for regulatory effects.

RESULTS

All cases and controls were genotyped for the eight candidate SNPs and two ancestry-informative markers (AIMs). Genotype call rates were greater than 95% for both cases and controls. Table II shows characteristics of the case-control sample. The case samples included 86 males (53%) and 75 females (47%). Out of these cases, 66 identified themselves as non-Hispanic White, 72 as Hispanic White, 17 as Black, and 6 as "other." The group labeled "other" included those identifying themselves as Asian, Native American, or other. The healthy controls included 130 males (56%) and 101 females (44%), who had visited a pediatric clinic at Texas

Children's Hospital for a non-disease related reason. Forty-nine were classified as non-Hispanic White, 98 as Hispanic White, and 78 as Black. The distribution of ethnic background was different between cases and controls mainly due to the infrequency of childhood ALL in Blacks. Because of this difference, results were adjusted for ethnic background or stratified analyses were performed when necessary. Mean birth weight was not different between cases and controls. Eighty-eight percent of the cases were diagnosed with early precursor B (early pre-B) ALL subtype, and associations did not change in effect size depending on the molecular subtype. There were no significant genotype associations found within the Black subpopulation of the sample and each of the SNPs genotyped (results not shown), possibly due to the small number of cases in the population sample.

GWAS risk markers

Results including genotype frequencies and ORs are described in Tables III-IV. Analyses yielded significant associations with some of the genetic markers similar to previous reports. In total, three *ARID5B* SNPs were examined for associations. Two of the *ARID5B* SNPs, rs7089424 and rs10821936, showed expected risk associations, while rs10994982 did not show an association (Table III). The SNP rs7089424 had an overall OR per allele (OR_{allele}) of 1.69 (P>0.001). The association showed a somewhat stronger risk in the non-Hispanic subgroup (OR_{allele}=2.11, P=0.01), compared with Hispanics (OR_{allele}=1.61, P=0.02). Similarly, rs10821936 had an overall OR_{allele}=1.48 (P=0.05). Adjustments of the analyses for race/ethnicity did not change the observed results. Adjustment for AIMs did not appreciably alter the results either.

Hardy-Weinberg equilibrium was violated in controls for *IKZF1* rs4132601, which could lead to spurious results, and was therefore excluded. There was no significant association found for *IKZF1* rs4132601 or for *CEBPE* rs2239633.

HLA region lymphoma risk markers

The SNP located in the HLA-DR region, rs2395185, showed a weak, non-significant risk overall (Table IV). This SNP is an exclusive marker for the HLA-DRB4 (DR53) lineage [23]. Since this lineage was shown to be a risk marker for childhood ALL in European samples, but only in males [24], we examined rs2395185 association in males. The non-Hispanic White male group had an OR_{allele} of 2.79 (P=0.016). The OR reached 6.21 (95% CI=0.70-54.96) for homoyzgosity for the variant allele, which corresponds to the original association [24]. The known protective marker for follicular lymphoma, rs2647012, showed a statistically significant association in Hispanics OR_{allele} =2.21 (P=0.007), but not in non-Hispanics ($P_{interaction}$= 0.003 for ethnicity). The significance remained after adjusting for both rs23951885 and rs10484561. The variant rs10484561 was shown to be a strong protective marker in this study, opposite of what was found in follicular lymphoma [33]. Using the recessive model, the variant allele showed a strong association, in the overall sample with OR_{rec}=0.19 (P=0.009) after adjustment for race/ethnicity, and OR_{rec}=0.17 (P=0.004) in Mantel-Haenszel analysis.

DISCUSSION

Despite being well established, the racial disparity in the incidence of childhood ALL is not always addressed in genetic association studies. Most GWAS, until recently, have identified risk loci using only European-origin populations. Variant polymorphisms located within the *ARID5B*, *IKZF1*, and *CEBPE* genes have reported strong risk associations in multiple studies [19-22, 38-40]. Our study provides some confirmation of previously discovered genetic markers associated with childhood ALL, which also validated our case-control set for further exploration.

Of the three *ARID5B* SNPs, rs7089424 and rs10821936 showed significant risk associations. The *ARID5B* gene is involved in transcriptional regulation during embryonic

development [38]. Overexpression of the gene in particular leukemias have led some to speculate that variations within the gene may affect B-lineage development, and increase susceptibility to B-lineage leukemia [20]. The marker rs7089424 was associated with a stronger risk of leukemia in the non-Hispanic subgroup compared with Hispanics, replicating recent findings from case-control studies using Hispanic populations [8,22].

In the recent multi-ethnic GWAS by Xu *et al*., rs10821936 was found to be a significant risk marker across all ethnicities. Xu *et al*. noted that the risk allele frequencies for rs10821936 increased in order by race incident rates: Black/African Americans, non-Hispanic/European American, and Hispanic Americans [22]. Our results showed a similar trend with an increasing risk allele frequency in cases of Blacks, non-Hispanic Whites and Hispanics. The multi-ethnic GWAS reported that rs10821936 was highly correlated with Native American genetic ancestry [22], substantiating their previous observations that Native American ancestry correlates with higher risk of relapse in Hispanics, and leading to speculation that this may be a factor with the increased risk of childhood ALL for Hispanic children who have a high proportion of Native American ancestry [41].

The SNP rs4132601, located in the Ikaros family zinc finger 1 (*IKZF1*) gene, is associated with increased risk of childhood ALL in multiple studies [19,21,39,40,42]. The Ikaros proteins are known to be involved with lymphocyte development and differentiation [19], and deletions are frequent and associated with unfavorable prognosis in B-cell precursor ALL [19,43]. One study found this variant to be a significant risk marker amongst non-Hispanic Whites, but not in Hispanics, despite similar allele frequencies [40]. Chokkalingam*et al*. hypothesized that that this marker's association may be due to linkage disequilibrium with a functional variant, and because of admixture in Hispanic populations the linkage disequilibrium

may vary [40]. Our study was unable to examine this SNP, due to Hardy-Weinberg disequilibrium found in controls after stratification for race/ethnicity.

The present study confirmed some, but not all previous findings of GWAS. With the *ARID5B* risk SNPs, there was heterogeneity even between the first two GWAS reports [19,20]. The modest sample size we had also reduced the statistical power of our study. Nevertheless, confirmed results validated the present case-control sample for further genetic association studies.

The relevance of lymphoma-associated polymorphisms in childhood ALL was assessed by genotyping rs2395185, rs10484561, and rs2647012. The SNP near the *HLA-DRA* gene, rs2395185, is a marker for the *HLA-DRB4* (DR53) lineage [23]. The *HLA-DRB4* lineage or its marker SNP have been previously shown as a risk marker in lung cancer [27], asthma [29], rheumatoid arthritis [28], type I diabetes [44], adult acute myeloblastic leukemia [45], chronic myeloid leukemia [46], chronic lymphoid leukemia [47-49] and in childhood ALL (males only)[24,50,51], and as a protective marker for non-Hodgkin lymphoma [26], and ulcerative colitis [30-32,52]. The DRB4/DR53 lineage has been shown previously to have a risk association with childhood ALL, with male specificity, within a European sample via HLA typing [24]. The first GWAS association of rs2395185 was with ulcerative colitis [30,32]. The variant allele, T, was later found to be a protective marker in a GWAS examining risk factors for classical Hodgkin lymphoma [26], and most recently a risk marker factor in Asian females for lung cancer [27]. Our results replicated the strong male specificity of the risk for childhood ALL, specifically in non-Hispanic White males, with no association in Hispanics. The DRB4 lineage has unique features, such as lower expression levels of HLA-DR molecules, poor interaction with CD4, disrupted intracellular transport, and possibly contains extra amount of DNA in the DR/DQ

region which may contribute to this risk association in childhood ALL [25]. One important finding of the present study is that the risk modifiers of lymphoma showed associations in opposite directions in childhood ALL.

The variant rs2647012, a marker for DRB3/DRB5 lineages [53] and a protective marker for lymphoma [34], showed a statistically significant risk association among Hispanics, but not in non-Hispanics ($P_{\text{interaction}} = 0.003$ for ethnicity). Our study appears to be the first looking at a Hispanic population with this SNP, and further studies are warranted to determine if this inverse relationship of risk exists in non-Hispanic and Hispanic Whites. Variant rs10484561 showed signs of being a protective marker for childhood ALL, opposite to the follicular lymphoma findings [33].

It is now customary that the additive model is used to assess statistical associations of SNPs. While the additive model has sufficient power to detect associations in most situations, there are certain scenarios that it may not show statistical significance when in fact, there is an association. An association conforming to the strictly recessive model when allele frequency is low is one example [54-56]. For this reason, and also because the original rs2395185 association was a recessive model association, we also assessed this genetic model in HLA region associations. This approach consistently yielded larger effect sizes for HLA region SNPs, especially for rs2395185 as in previous studies. We are in favor of routine use of the recessive model analysis in exploration of associations in the HLA region.

Our study had a number of limitations. With childhood ALL being a rare disease, the sample size for our study was small, resulting in limited statistical power. The issue of self-reported ethnicity may be of concern. This method is common with population-based association studies, and residual confounding is often suspected. Even though it has been described that self-

reported ethnicity may be reliable [57], the heterogeneity within the Hispanic population is still a concern. A recent study conducted in a Spanish population was unable to replicate original risk associations found in Hispanic Americans, demonstrating the large heterogeneity in this high risk group [22,58].To adjust for heterogeneity in our sample, especially in Hispanics, we used two ancestry-informative markers, AIMs, to control for confounding caused by population stratification and to avoid spurious associations [35]. The AIMs have widely different allele frequencies in major human continental groups. The two AIMs we used, rs285 and rs2891, have been previously used in Hispanic populations to account for the differences in genetic ancestry [36,37]. Adjusting the risk associations by AIMs did not alter the results.

Our study did have a well-defined phenotype, with molecular ALL types determined. The use of a multi-ethnic sample population was a strength, especially for the ethnic disparity that exists in childhood ALL susceptibility. Another strength is the use of multiple genetic models, where appropriate, to determine associations that may remain undetectable by the exclusive use of the additive model association. The replication of known leukemia markers validated our sample set for further studies. This pilot study is part of an ongoing effort at BCM. Recruitment for the second phase is continuing together with clinical follow-up.

In summary, we validated a new multi-ethnic case-control set and also examined some new markers with their association with childhood ALL. The examination of lymphoma risk markers yielded associations in opposite directions for childhood ALL, and also confirmed a previously identified childhood ALL risk marker. Two HLA region associations were ethnicity-specific. Still, our study did not provide clues as to why Hispanics have a higher susceptibility to childhood ALL, suggesting that environmental factors may have to be incorporated in future studies to examine their role in this differential. Studies with information on environmental

exposures may help explain how gene-environment interactions contribute to childhood ALL

susceptibility and its variation among different populations.

POTENTIAL CONFLICTS OF INTERESTS

Disclosure forms provided by the authors are available with the full text of this article at www.informahealthcare.com/lal.

REFERENCES

[1]   Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011: The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. CA Cancer J Clin 2011;61:212-36.

[2]   Greaves M. Infection, immune responses and the aetiology of childhood leukaemia. Nat Rev Cancer 2006;6:193-203.

[3]   Mody R, Li S, Dover DC, et al. Twenty-five-year follow-up among survivors of childhood acute lymphoblastic leukemia: a report from the Childhood Cancer Survivor Study. Blood 2008;111:5515-23.

[4]   Sherborne AL, Hemminki K, Kumar R, et al. Rationale for an international consortium to study inherited genetic susceptibility to childhood acute lymphoblastic leukemia. Haematologica 2011;96:1049-54.

[5]   Howlader N NA, Krapcho M, Garshell J, et al. SEER Cancer Statistics Review, 1975-2010. Bethesda, MD: National Cancer Institute; 2013.

[6]   McNeil DE, Cote TR, Clegg L, Mauer A. SEER update of incidence and trends in pediatric malignancies: acute lymphoblastic leukemia. Med Pediatr Oncol 2002;39:554-7; discussion 552-3.

[7]   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;47:2365-70.

[8]   Xu H, Cheng C, Devidas M, et al. ARID5B genetic polymorphisms contribute to racial disparities in the incidence and treatment outcome of childhood acute lymphoblastic leukemia. J Clin Oncol 2012;30:751-7.

[9]   Swinney RM, Beuten J, Collier AB, 3rd, Chen TT, Winick NJ, Pollock BH, Tomlinson GE. Polymorphisms in CYP1A1 and ethnic-specific susceptibility to acute lymphoblastic leukemia in children. Cancer Epidemiol Biomarkers Prev 2011;20:1537-42.

[10]   Birch JM. Genes and cancer. Arch Dis Child 1999;80:1-3.

[11]   Pui CH. Childhood leukemias. N Engl J Med 1995;332:1618-30.

[12]   Belson M, Kingsley B, Holmes A. Risk factors for acute leukemia in children: a review. Environ Health Perspect 2007;115:138-45.

[13]   Pui CH, Robison LL, Look AT. Acute lymphoblastic leukaemia. Lancet 2008;371:1030-43.

[14]   Feller M, Adam M, Zwahlen M, Brazzola P, Niggli F, Kuehni C. Family characteristics as risk factors for childhood acute lymphoblastic leukemia: a population-based case-control study. PLoS One 2010;5.

[15]   Greaves MF. Aetiology of acute leukaemia. Lancet 1997;349:344-9.

[16]   Hjalgrim LL, Westergaard T, Rostgaard K, et al. Birth weight as a risk factor for childhood leukemia: a meta-analysis of 18 epidemiologic studies. Am J Epidemiol 2003;158:724-35.

[17]   Hjalgrim LL, Rostgaard K, Hjalgrim H, et al. Birth weight and risk for childhood leukemia in Denmark, Sweden, Norway, and Iceland. J Natl Cancer Inst 2004;96:1549-56.

[18]   Caughey RW, Michels KB. Birth weight and childhood leukemia: A meta-analysis and review of the current evidence. Int J Cancer 2009;124:2658-70.

[19]   Papaemmanuil E, Hosking FJ, Vijayakrishnan J, 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.

[20]   Trevino LR, Yang W, French D, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nat Genet 2009;41:1001-5.

[21]   Orsi L, Rudant J, Bonaventure A, et al. Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE). Leukemia 2012;26:2561-4.

[22]   Xu H, Yang W, Perez-Andreu V, et al. Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. J Natl Cancer Inst 2013;105:733-42.

[23]   Kennedy AE, Singh SK, Dorak MT. Re: genome-wide association study of classical hodgkin lymphoma and epstein-barr virus status-defined subgroups. J Natl Cancer Inst 2012;104:884-5.

[24]   Dorak MT, Lawson T, Machulla HK, Darke C, Mills KI, Burnett AK. Unravelling an HLA-DR association in childhood acute lymphoblastic leukemia. Blood 1999;94:694-700.

[25]   Dorak MT, Oguz FS, Yalman N, et al. A male-specific increase in the HLA-DRB4 (DR53) frequency in high-risk and relapsed childhood ALL. Leuk Res 2002;26:651-6.

[26]   Urayama KY, Jarrett RF, Hjalgrim H, et al. Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. J Natl Cancer Inst 2012;104:240-53.

[27] Lan Q, Hsiung CA, Matsuo K, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. Nat Genet 2012;44:1330-5.

[28] Plenge RM, Cotsapas C, Davies L, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 2007;39:1477-82.

[29] Li X, Ampleford EJ, Howard TD, et al. Genome-wide association studies of asthma indicate opposite immunopathogenesis direction from autoimmune diseases. J Allergy Clin Immunol 2012;130:861-8 e7.

[30] Asano K, Matsushita T, Umeno J, et al. A genome-wide association study identifies three new susceptibility loci for ulcerative colitis in the Japanese population. Nat Genet 2009;41:1325-9.

[31] Juyal G, Prasad P, Senapati S, et al. An investigation of genome-wide studies reported susceptibility loci for ulcerative colitis shows limited replication in north Indians. PLoS One 2011;6:e16565.

[32] Silverberg MS, Cho JH, Rioux JD, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet 2009;41:216-20.

[33] Conde L, Halperin E, Akers NK, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. Nat Genet 2010;42:661-4.

[34] Smedby KE, Foo JN, Skibola CF, et al. GWAS of follicular lymphoma reveals allelic heterogeneity at 6p21.32 and suggests shared genetic susceptibility with diffuse large B-cell lymphoma. PLoS Genet 2011;7:e1001378.

[35] Enoch MA, Shen PH, Xu K, Hodgkinson C, Goldman D. Using ancestry-informative markers to define populations and detect population stratification. J Psychopharmacol 2006;20:19-26.

[36] Ziv E, John EM, Choudhry S, et al. Genetic ancestry and risk factors for breast cancer among Latinas in the San Francisco Bay Area. Cancer Epidemiol Biomarkers Prev 2006;15:1878-85.

[37] Lee YL, Teitelbaum S, Wolff MS, Wetmur JG, Chen J. Comparing genetic ancestry and self-reported race/ethnicity in a multiethnic population in New York City. J Genet 2010;89:417-23.

[38] Healy J, Richer C, Bourgey M, Kritikou EA, Sinnett D. Replication analysis confirms the association of ARID5B with childhood B-cell acute lymphoblastic leukemia. Haematologica 2010;95:1608-11.

[39] Prasad RB, Hosking FJ, Vijayakrishnan J, et al. Verification of the susceptibility loci on 7p12.2, 10q21.2 and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood. Blood 2010;115:1765-7.

[40]   Chokkalingam AP, Hsu LI, Metayer C, et al. Genetic variants in ARID5B and CEBPE are childhood ALL susceptibility loci in Hispanics. Cancer Causes Control 2013;24:1789-95.

[41]   Yang JJ, Cheng C, Devidas M, et al. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. Nat Genet 2011;43:237-41.

[42]   Lautner-Csorba O, Gezsi A, Semsei AF, et al. Candidate gene association study in pediatric acute lymphoblastic leukemia evaluated by Bayesian network based Bayesian multilevel analysis of relevance. BMC Med Genomics 2012;5:42.

[43]   Mullighan CG, Downing JR. Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: recent insights and future directions. Leukemia 2009;23:1209-18.

[44]   Nakanishi K, Shima Y. Capture of type 1 diabetes-susceptible HLA DR-DQ haplotypes in Japanese subjects using a tag single nucleotide polymorphism. Diabetes Care 2010;33:162-4.

[45]   Seremetis S, Cuttner J, Winchester R. Definition of a possible genetic basis for susceptibility to acute myelogenous leukemia associated with the presence of a polymorphic Ia epitope. J Clin Invest 1985;76:1391-7.

[46]   Oguz FS, Kalayoglu S, Diler AS, et al. HLA system affects the age-at-onset in chronic myeloid leukemia. Am J Hematol 2003;73:256-62.

[47]   Machulla HK, Muller LP, Schaaf A, Kujat G, Schonermarck U, Langner J. Association of chronic lymphocytic leukemia with specific alleles of the HLA-DR4:DR53:DQ8 haplotype in German patients. Int J Cancer 2001;92:203-7.

[48]   Dyer PA, Ridway JC, Flanagan NG. HLA-A,B and DR antigens in chronic lymphocytic leukaemia. Dis Markers 1986;4:231-7.

[49]   Dorak MT, Machulla HK, Hentschel M, Mills KI, Langner J, Burnett AK. Influence of the major histocompatibility complex on age at onset of chronic lymphoid leukaemia. Int J Cancer 1996;65:134-9.

[50]   Dorak MT, Owen G, Galbraith I, et al. Nature of HLA-associated predisposition to childhood acute lymphoblastic leukemia. Leukemia 1995;9:875-8.

[51]   Dorak MT, Chalmers EA, Gaffney D, et al. Human major histocompatibility complex contains several leukemia susceptibility genes. Leuk Lymphoma 1994;12:211-22.

[52]   Latiano A, Palmieri O, Latiano T, et al. Investigation of multiple susceptibility loci for inflammatory bowel disease in an Italian cohort of patients. PLoS One 2011;6:e22688.

[53]   Satta Y, Mayer WE, Klein J. HLA-DRB intron 1 sequences: implications for the evolution of HLA-DRB genes and haplotypes. Hum Immunol 1996;51:1-12.

[54]   Zheng G, Joo J, Lin JP, Stylianou M, Waclawiw MA, Geller NL. Robust ranks of true associations in genome-wide case-control association studies. BMC Proc 2007;1 Suppl 1:S165.

[55]   Lettre G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. Genet Epidemiol 2007;31:358-62.

[56]   Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend tests for case-control studies of genetic markers: power, sample size and robustness. Hum Hered 2002;53:146-52.

[57]   Mez JB, Cole JW, Howard TD, et al. Evaluation of self-reported ethnicity in a case-control population: the stroke prevention in young women study. BMC Res Notes 2009;2:260.

[58]   Lopez-Lopez E, Gutierrez-Camino A, Martin-Guerrero I, Garcia-Orad A. Re: Novel Susceptibility Variants at 10p12.31-12.2 for Childhood Acute Lymphoblastic Leukemia in Ethnically Diverse Populations. J Natl Cancer Inst 2013.

Table I. Main features of SNPs analyzed

Gene	SNP	Chromosome nucleotide position[*]	Inclusion criteria	PMID[**]	Minor allele and frequency[†]	Location	RegulomeDB score[††]
ARID5B	rs7089424	chr10: 63752159	GWAS identified risk loci for childhood ALL	19684604, 22660188, 20042726	(G) 0.314	Intronic	3a
ARID5B	rs10821936	chr10: 63723577	GWAS identified risk loci for childhood ALL	19684603, 20054350, 22660188, 23512250, 22291082	(C) 0.318	Intronic	5
ARID5B	rs10994982	chr10: 63710104	GWAS identified risk loci for childhood ALL	19684603, 22660188	(A) 0.457	Intronic	NA
IKZF1	rs4132601	chr7: 50470604	GWAS identified risk loci for childhood ALL	19684604, 22660188, 20054350	(G) 0.306	3'-UTR	5
CEBPE	rs2239633	chr14: 23589057	GWAS identified risk loci for childhood ALL	19684604, 22660188	(A) 0.466	5'-upstream	4
HLA-DR region	rs2395185	chr6: 32433167	HLA-DRB4/DR53 lineage; ALL risk marker (in European males); Hodgkin lymphoma risk marker	10397736, 12008082, 22286212, 7909466	(T) 0.423	Intronic	6
HLA-DQB1 region	rs2647012	chr6: 32664458	HLA-DRB3/DRB5 lineage, protective marker for non-Hodgkin (follicular) lymphoma	21533074, 22911334, 23455380	(T) 0.381	Intergenic	6
HLA-DQA1 region	rs10484561	chr6: 32665420	HLA-DR1/DR10 lineage, risk marker in follicular lymphoma	20639881, 21533074, 23025665	(G) 0.084	Intergenic	6

[*]Genome Reference Consortium Human Build 37 patch release 10 (GRCh37.p10) used for nucleotide position (http://www.ncbi.nlm.nih.gov/SNP/)

[**]PubMed identifier number

[†]Minor allele frequencies are from a reference Caucasian population (U.S. residents of northern and western European ancestry) genotyped in HapMap project

[††]RegulomeDB scores range from 1 (most functional) to 5 (least functional) (6=other). Not all SNPs have a RegulomeDB score (http://regulome.stanford.edu/)

Table II. Characteristics of cases and controls

	Cases n=161	Controls n=231	*P* value
Ethnic background			
Non-Hispanic White	66	49	
Hispanic White	72	98	**<0.001**
Black	17	78	
Other[*]	6	6	
Gender			
Male	86	130	
Female	75	101	0.58
Ratio	1.15	1.29	
Birth weight (*grams*)			
Mean (SD)	3349.3 (584)	3263.3 (684)	0.23
Median (IQR)	3400 (760)	3311.5 (850)	0.48

[*]Other includes Asian, Native American, or any other ethnicities

Table III. SNP associations previously shown as ALL risk markers in GWAS (overall[*])

Gene	SNP	Minor Allele	Cases			Controls			Minor allele frequency		OR$_{allele}$[**] (95% CI)	P value
			A/A	A/B	B/B	A/A	A/B	B/B	Cases	Controls		
ARID5B	**rs7089424**	G	37	80	43	113	68	46	0.52	0.35	**1.69 (1.28-2.24)**	**<0.001**
ARID5B	**rs10821936**	C	39	72	44	104	69	50	0.52	0.38	**1.48 (1.12-1.95)**	**0.005**
ARID5B	**rs10994982**	A	55	78	25	90	78	43	0.41	0.39	1.00 (0.75-1.34)	0.992
IKZF1	**rs4132601**	G	76	68	16	141	56	29	0.31	0.25	1.19 (0.88-1.60)	0.269[†]
CEBPE	**rs2239633**	A	68	68	23	118	74	36	0.36	0.32	1.02 (0.76-1.36)	0.881

[*]ORs adjusted for self-reported race/ethnicity

[**]OR per allele (OR$_{allele}$) for the additive model

[†]Hardy-Weinberg disequilibrium in controls

Table IV. Association of HLAregion lymphoma susceptibility markers (overall[*])

| SNP | Gene | Minor allele | Cases | | | Controls | | | Minor allele frequency | | | | |
			A/A	A/B	B/B	A/A	A/B	B/B	Cases	Controls	OR$_{allele}$[**] (95% CI)	*P* value	OR$_{rec}$[†] (95% CI)	*P* value
rs2395185	HLA-DR region	T	69	68	22	120	72	25	0.35	0.28	1.27 (0.94-1.71)	0.127	1.1 (0.65-2.26)[††]	0.553
rs2647012	HLA-DQB1 region	A	78	64	16	116	47	30	0.3	0.28	1.09 (0.80-1.47)	0.595	0.57 (0.30-1.12)	0.103
rs10484561	HLA-DQA1 region	G	128	23	3	149	21	18	0.09	0.15	0.70 (0.46-1.06)	0.094	**0.19 (0.05-0.66)**	**0.009**

[*]ORs adjusted for self-reported ethnicity and race

[**]OR per allele (OR$_{allele}$) for the additive model

[†]OR recessive (OR$_{rec}$) for the variant homozygous genotype

[††]OR$_{allele}$=1.88 (*P*=0.003) in males; OR$_{allele}$=2.79 (*P*=0.016) in non-Hispanic White males