



TITLE:

Regional evaluation of childhood acute lymphoblastic leukemia genetic susceptibility loci among Japanese

AUTHOR(S):

Urayama, Kevin Y.; Takagi, Masatoshi; Kawaguchi, Takahisa; Matsuo, Keitaro; Tanaka, Yoichi; Ayukawa, Yoko; Arakawa, Yuki; ... Mizutani, Shuki; Matsuda, Fumihiko; Manabe, Atsushi

CITATION:

Urayama, Kevin Y. ...[et al]. Regional evaluation of childhood acute lymphoblastic leukemia genetic susceptibility loci among Japanese. Scientific Reports 2018, 8: 789.

ISSUE DATE:

2018-01-15

URL:

<http://hdl.handle.net/2433/236113>

RIGHT:

© The Author(s) 2018. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

SCIENTIFIC REPORTS

OPEN

Regional evaluation of childhood acute lymphoblastic leukemia genetic susceptibility loci among Japanese

Kevin Y. Urayama^{1,2}, Masatoshi Takagi³, Takahisa Kawaguchi⁴, Keitaro Matsuo⁵, Yoichi Tanaka⁶, Yoko Ayukawa¹, Yuki Arakawa⁷, Daisuke Hasegawa⁸, Yuki Yuza⁹, Takashi Kaneko⁹, Yasushi Noguchi¹⁰, Yuichi Taneyama¹¹, Setsuo Ota¹², Takeshi Inukai¹³, Masakatsu Yanagimachi^{3,14}, Dai Keino¹⁵, Kazutoshi Koike¹⁶, Daisuke Toyama¹⁷, Yozo Nakazawa¹⁸, Hidemitsu Kurosawa¹⁹, Kozue Nakamura²⁰, Koichi Moriwaki²¹, Hiroaki Goto²², Yujin Sekinaka²³, Daisuke Morita¹⁸, Motohiro Kato²⁴, Junko Takita²⁵, Toshihiro Tanaka^{26,27}, Johji Inazawa²⁷, Katsuyoshi Koh⁷, Yasushi Ishida²⁸, Akira Ohara²⁹, Shuki Mizutani³, Fumihiko Matsuda⁴ & Atsushi Manabe⁸

Genome-wide association studies (GWAS) performed mostly in populations of European and Hispanic ancestry have confirmed an inherited genetic basis for childhood acute lymphoblastic leukemia (ALL), but these associations are less clear in other races/ethnicities. DNA samples from ALL patients (aged 0–19 years) previously enrolled onto a Tokyo Children’s Cancer Study Group trial were collected during 2013–2015, and underwent single nucleotide polymorphism (SNP) microarray genotyping resulting in 527 B-cell ALL for analysis. Cases and control data for 3,882 samples from the Nagahama Study Group and Aichi Cancer Center Study were combined, and association analyses across 10 previous GWAS-identified regions were performed after targeted SNP imputation. Linkage disequilibrium (LD) patterns in Japanese and other populations were evaluated using the varLD score based on 1000 Genomes

¹Department of Social Medicine, National Center for Child Health and Development, Tokyo, Japan. ²Graduate School of Public Health, St. Luke’s International University, Tokyo, Japan. ³Department of Pediatrics and Developmental Biology, Tokyo Medical and Dental University, Tokyo, Japan. ⁴Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. ⁵Division of Molecular and Clinical Epidemiology, Aichi Cancer Center Research Institute, Aichi, Japan. ⁶Department of Clinical Pharmacy, Center for Clinical Pharmacy and Sciences, School of Pharmacy, Kitasato University, Tokyo, Japan. ⁷Department of Hematology/Oncology, Saitama Children’s Medical Center, Saitama, Japan. ⁸Department of Pediatrics, St. Luke’s International Hospital, Tokyo, Japan. ⁹Department of Hematology/Oncology, Tokyo Metropolitan Children’s Medical Center, Tokyo, Japan. ¹⁰Department of Pediatrics, Japanese Red Cross Narita Hospital, Chiba, Japan. ¹¹Department of Hematology/Oncology, Chiba Children’s Hospital, Chiba, Japan. ¹²Department of Pediatrics, Teikyo University Chiba Medical Center, Chiba, Japan. ¹³Department of Pediatrics, University of Yamanashi, Yamanashi, Japan. ¹⁴Department of Pediatrics, Yokohama City University Graduate School of Medicine, Yokohama, Japan. ¹⁵Department of Pediatrics, St. Marianna University School of Medicine, Kawasaki, Japan. ¹⁶Division of Pediatric Hematology and Oncology, Ibaraki Children’s Hospital, Mito, Japan. ¹⁷Division of Pediatrics, Showa University Fujigaoka Hospital, Yokohama, Japan. ¹⁸Department of Pediatrics, Shinshu University School of Medicine, Matsumoto, Japan. ¹⁹Department of Pediatrics, Dokkyo Medical University, Tochigi, Japan. ²⁰Department of Pediatrics, Teikyo University Hospital, Tokyo, Japan. ²¹Department of Pediatrics, Saitama Medical Center, Saitama Medical University, Saitama, Japan. ²²Division of Hematology/Oncology & Regenerative Medicine, Kanagawa Children’s Medical Center, Yokohama, Japan. ²³Department of Pediatrics, National Defense Medical College, Saitama, Japan. ²⁴Children’s Cancer Center, National Center for Child Health and Development, Tokyo, Japan. ²⁵Department of Pediatrics, The University of Tokyo Hospital, Tokyo, Japan. ²⁶Department of Human Genetics and Disease Diversity, Tokyo Medical Dental University, Tokyo, Japan. ²⁷Bioresource Research Center, Tokyo Medical and Dental University, Tokyo, Japan. ²⁸Pediatric Medical Center, Ehime Prefectural Central Hospital, Matsuyama, Japan. ²⁹Department of Pediatrics, Toho University, Tokyo, Japan. Correspondence and requests for materials should be addressed to K.Y.U. (email: urayama-k@ncchd.go.jp)

data. Risk associations for *ARID5B* (rs10821936, OR = 1.84, $P = 6 \times 10^{-17}$) and *PIP4K2A* (rs7088318, OR = 0.76, $P = 2 \times 10^{-6}$) directly transferred to Japanese, and the *IKZF1* association was detected by an alternate SNP (rs1451367, OR = 1.52, $P = 2 \times 10^{-6}$). Marked regional LD differences between Japanese and Europeans was observed for most of the remaining loci for which associations did not transfer, including *CEBPE*, *CDKN2A*, *CDKN2B*, and *ELK3*. This study represents a first step towards characterizing the role of genetic susceptibility in childhood ALL risk in Japanese.

Scrutiny of the human genome through evaluation of common genetic variants has revealed hundreds of disease susceptibility loci. In childhood acute lymphoblastic leukemia (ALL), six regions that have replicated in several populations are now considered known susceptibility loci (likely representing associations with *ARID5B*, *IKZF1*, *CEBPE*, *CDKN2A*, *PIP4K2A*, and *GATA3*), with the majority of the evidence supported through studies conducted in populations of European and Hispanic descent^{1,2}. Gains in statistical power achieved by recent meta-analyses of childhood ALL genome-wide association studies (GWAS) have resulted in the identification of risk-associated single nucleotide polymorphisms (SNPs) of comparatively lower allele frequencies and estimated magnitude of effects including those tagging the *CDKN2B*, *LHPP*, and *ELK3* genes^{3,4}. Furthermore, the concept that genetic susceptibility studies may potentially reveal race/ethnicity-specific associations was demonstrated by a recent GWAS conducted in a Chinese population which implicated a role for the *WWOX* gene that had not been observed in the numerous previous studies conducted in populations of European and Hispanic ancestry⁵.

Success of GWAS in identifying true disease-associated loci have largely been due to the consistently high standards in methodological rigor of the approach including, strict quality control for genotype data, attention to issues of statistical power and sample size, criteria for genome-wide significance, and integrating components of independent validation and/or functional evaluation of loci⁶. However, as is the case with other complex diseases, it is well-recognized that the known GWAS 'hits' in childhood ALL account for only a small proportion of the total estimated heritability⁷. Based on data from populations of European ancestry, it has been estimated that the currently known childhood ALL associated risk loci account for about 19 percent of the additive heritable risk, not accounting for potential impact of epistasis or gene-environment interactions⁴. Adding to this issue of 'missing heritability' is the realization that we currently know even less about the nature of childhood ALL genetic susceptibility in other populations, particularly Asians⁸.

The effects of known genetic susceptibility loci have yet to be fully confirmed in populations of non-European ancestry. Targeted validation attempts based on the same SNPs originally identified in mostly non-Hispanic whites and Hispanics have been performed in Chinese and other Asian populations, but findings have been inconsistent^{9,10}. Assuming the same causal variant is operative across populations, a lack of association in Asians can be attributed to study flaws, but more likely due to reduced statistical power as a result of differences in allele frequency, strength of linkage disequilibrium (LD) with the causal variant(s), and/or the role of environmental exposures in affecting risk. Thus, a comprehensive characterization of genetic variation across the targeted genetic loci is required for an appropriate validation attempt in different populations.

To address this current gap in knowledge, we initiated an effort through the Tokyo Children's Cancer Study Group (TCCSG) to assemble resources for genomic investigation of germline contributions to childhood ALL susceptibility and outcomes. Here, we report results of our targeted analysis of previous GWAS-identified childhood ALL risk loci in a large Japanese population. We evaluated the transferability of risk associations of specific SNP loci in Japanese and interpreted the finding in the context of quantified differences in LD within those loci across populations. Furthermore, we analyzed regional SNP data in order to identify alternate SNPs which may potentially confer stronger association in Japanese.

Materials and Methods

Study Population. In collaboration with a large network of 23 hospitals participating in the Tokyo Children's Cancer Study Group (TCCSG), previously diagnosed childhood ALL patients visiting for a routine follow-up between 2013 and 2015 were invited to participate in this study. The TCCSG network includes nearly all clinical centers that diagnose and treat childhood ALL within the seven prefectures that comprise the Kanto and immediately surrounding regions^{11,12}. Patients were considered eligible if they were 19 years of age or younger at the time of ALL diagnosis, enrolled onto a TCCSG treatment protocol, and self-identified as Japanese. Due to the nature of this sampling scheme, the study population comprised a survivorship population of childhood ALL patients.

Upon obtaining written informed consent, saliva samples using the Oragene Saliva DNA Self-Collection Kit (4 years of age and older) or Assisted Collection Kit (less than 4 years of age) (DNA Genotek, Ottawa, Canada) were collected from the patients with instruction by the attending physician or nurse during the follow-up outpatient visit. The collected samples were shipped at room temperature to a central laboratory (Tokyo Medical and Dental University) for processing, DNA extraction, and storage.

Controls comprised a subset of adult participants enrolled in two ongoing epidemiological studies of lifestyle-related chronic diseases in Japan, the Nagahama Study Group¹³ and Aichi Cancer Center Study¹⁴, in which large-scale genome-wide SNP genotyping had already been performed. The Nagahama Study is a community-based prospective cohort study comprising a representative sample of residents of Nagahama City in Shiga Japan¹⁵. The Aichi Cancer Center Study comprised a hospital-based cohort of non-cancer outpatient visitors¹⁶. Despite the name, the Aichi Cancer Center resembles a general hospital that does not require physician referral in which the majority of outpatients present with no abnormal findings by clinical examination. Population substructure across regions of Japan does exist; most notably between populations of Okinawa and the other main islands of Japan collectively. Although cases and controls were recruited from different regions of the main island, simulation studies have shown only minimal genomic inflation potential when considering

these two subpopulations¹⁷. A history of childhood leukemia was not assessed in controls; however, the rarity of this disease suggests that any previous diagnosis of childhood leukemia in controls would have a minimal effect on the results of this study.

This study protocol was approved by the institutional review boards of St. Luke's International Hospital, Tokyo Medical and Dental University, Kyoto University, and all collaborating hospitals involved in patient recruitment. Written informed consent was obtained from the parents of each participant together with a written assent by the child where possible. Patients aged 16 to 19 years were asked to provide written informed consent together with parental consent; those aged 20 or older did not require parental consent. This study was conducted in accordance with the Declaration of Helsinki.

Genotyping and Quality Control. DNA extraction from childhood ALL patients' saliva samples were performed using the Oragene prepIT DNA Extraction Kit (DNA Genotek) based on the manufacturer's instruction. The approximately 2 mL saliva samples obtained from the Oragene Self-Collection Kits yielded, on average, a total of about 50 µg of genomic DNA.

Genome-wide SNP genotyping was attempted on 621 patient samples using the Illumina HumanCoreExome-12 v1.1 BeadChip (San Diego, CA) which contained probes for approximately 550,000 SNPs. Existing control data were genotyped previously using variable versions of the same Illumina HumanCoreExome BeadChip. Quality control steps were conducted within cases and each of the two different control sample series separately. SNPs were excluded if the genotype call rate was below 99%, the distribution of genotypes clearly deviated from that expected by Hardy-Weinberg equilibrium (HWE) ($P < 1 \times 10^{-6}$), or the minor allele frequency was less than 0.01. Samples were excluded if showed a genotyping success rate of less than 95% (51 cases and 4 controls) and relatedness based on an identity-by-descent analysis (1 case and 119 controls). In addition, principal components analysis (PCA) based on a genome-wide subset of SNPs in low LD (pruned at $r^2 < 0.1$) that passed quality control steps was performed on a known ethnically homogeneous population of Japanese ancestry (International HapMap Project) together with cases and controls. The PCA was conducted using the EIGENSTRAT 2.0 software package and outlier samples were excluded (2 cases and 5 controls)¹⁸. In result, after quality control steps and excluding 40 T-cell ALL patients, the final population for analysis included a total of 527 Japanese B-cell ALL cases and 3,882 controls with data available for 171,547 SNPs that were overlapping across the genotyped case and control series.

Targeted SNP imputation was performed on the combined case-control dataset for 10 genomic regions reported in previous childhood ALL GWAS (Table 1) using ShapeIT2¹⁹ and Minimac3²⁰, and the 1000 Genomes Project Phase III Version 5 as the reference population²¹. Poorly imputed SNPs defined by an $R^2 < 0.5$ were excluded from the analyses. Considering the gene and its broad surrounding region (about 100-kb flanking) for each locus, a total of 113 SNPs were excluded among 14,457 total SNPs imputed across the 10 regions. On average, about 0.8 percent of SNPs per locus were excluded based on this quality control threshold. Due to restrictions stipulated by the institutional review board approvals, data were not be made publicly available, but may be available on request in compliance with the policies and procedures of the TCCSG.

Statistical Analysis. We first tested the association between childhood ALL and 16 SNPs across the 10 genes (Table 1) identified in previous GWAS. SNPs for evaluation were selected based on the strongest result reported from the first study to report the association. Multiple SNPs tagging the same genomic region were selected if the SNP was examined across several studies. We examined the role of additional genetic variation across the entire span of the 10 targeted genes, including a 10-kb flanking region on both ends. The association between each genetic variant and risk of childhood ALL was estimated by the odds ratio (OR) per allele and 95% confidence intervals (CI) using multiple logistic regression assuming a log-additive genetic model. Genome-wide association analysis of the 171,547 SNPs showed evidence of genomic inflation ($\lambda > 1.10$); all analyses were adjusted for 10 PCA eigenvectors ($\lambda = 1.05$). For the test of specific previously reported GWAS SNPs, a nominal p-value of less than 0.05 was considered statistically significant. For the examination of other potentially associated SNPs across the genomic regions, to account for multiple comparisons in the presence of LD between SNPs, we calculated adjusted p-values based on 10,000 permutations of case-control status and considered p-values below a family-wise type I error rate threshold of 0.05 to be statistically significant. Analyses were conducted using PLINK²² and SAS software version 9 (SAS, Cary, NC). The LocusZoom web-based resource was used to generate plots of association results by genomic region²³.

Differences across race/ethnic populations in regional patterns of LD flanking a 10-kb region on both ends of the SNPs were quantified using the variation in LD (varLD) score applied to the 1000 Genomes Phase 3 data²¹. The varLD score is an algorithm based on comparing regional patterns of correlation previously developed by Teo *et al.* to quantify differences in LD within defined regions^{24,25}. With the exception of the *WWOX* locus, the Japanese (JPT) population was compared to the combined population of European ancestry (EUR); for the *WWOX* locus, JPT was compared to the combined Han Chinese and Southern Han Chinese (CHB-CHS) representing the population in which the locus was originally identified. Permutation procedures were performed to determine Monte Carlo statistical significance by comparing the estimated varLD score to the null distribution of varLD scores after successive re-sampling of the two populations from the combined data²⁵; 10,000 iterations were performed. Since 9 genomic loci were tested (*CDKN2A-CDKN2B* were evaluated as one region), an empirical p-value of less than 0.006 was considered statistically significant for the varLD evaluation. All statistical tests were two-sided.

Results

Association analyses were performed on a total of 527 B-cell ALL cases and 3,882 controls. Median age at ALL diagnosis was 4.5 years (range: 0.3–16.8 years). The risk of childhood B-cell ALL associated with 16

Gene/SNP ^a	Chr: Position ^b	Alleles	Cases (MAF) n = 527	Controls (MAF) n = 3,882	OR (95% CI) ^c	P
ARID5B (10q21.2)						
rs10994982	10:63,710,104	G/A	0.583	0.466	1.58 (1.37–1.82)	3.83 × 10 ⁻¹⁰
rs10821936	10:63,723,577	T/C	0.500	0.345	1.84 (1.60–2.13)	6.04 × 10 ⁻¹⁷
rs7089424	10:63,752,159	T/G	0.489	0.344	1.76 (1.53–2.03)	7.77 × 10 ⁻¹⁵
IKZF1 (7p12.2)						
rs11978267	7:50,466,304	A/G	0.092	0.097	1.13 (0.89–1.44)	0.321
rs4132601	7:50,470,604	T/G	0.091	0.092	1.19 (0.93–1.51)	0.164
CEBPE (14q11.2)						
rs4982731	14:23,585,333	T/C	0.136	0.133	1.11 (0.91–1.37)	0.306
rs2239633	14:23,589,057	G/A	0.452	0.467	0.93 (0.80–1.07)	0.282
CDKN2A, CDKN2B (9p21.3)						
rs3731217	9:21,984,661	A/C	0.191	0.183	1.11 (0.93–1.33)	0.251
rs662463	9:22,030,438	G/A	0.011	0.013	0.85 (0.45–1.62)	0.625
rs17756311	9:22,053,895	G/A	0.010	0.012	0.88 (0.45–1.71)	0.695
PIP4K2A (10p12.2)						
rs10828317	10:22,839,628	T/C	0.331	0.390	0.76 (0.65–0.88)	3.03 × 10 ⁻⁴
rs7088318	10:22,852,948	A/C	0.333	0.390	0.76 (0.65–0.88)	2.43 × 10 ⁻⁴
GATA3 (10p14)						
rs3824662	10:8,104,208	C/A	0.339	0.307	1.15 (1.00–1.33)	0.058
LHPP (10q26.13)						
rs35837782	10:126,293,309	G/A	0.414	0.393	1.06 (0.92–1.22)	0.442
ELK3 (12q23.1)						
rs4762284	12:96,612,762	T/A	0.415	0.430	0.93 (0.80–1.07)	0.303
WVOX (16q23.1)						
rs1121404	16:79,089,869	T/C	0.412	0.383	1.04 (0.90–1.19)	0.623

Table 1. Previously identified genetic variants from genome-wide association studies and risk of childhood B-cell ALL in Japanese. Abbreviations: Chr, chromosome; CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphism. ^aSNPs showing the strongest associations in previous genome-wide association studies (GWAS) within the identified region was selected. If multiple GWAS conducted in the same racial/ethnic population reported different SNPs, but tagged the same genomic region, SNPs from the first report were selected. ^bGenomic positions are based on the human genome assembly GRCh37 coordinates. ^cOdds ratios and 95% confidence intervals were calculated using logistic regression assuming a log-additive genetic model and adjusting for 10 PCA eigenvectors.

SNPs (representing 10 genes) reported in previous GWAS was evaluated in this Japanese population (Table 1). The *ARID5B* SNPs showed strong evidence of an association with the highest risk observed for rs10821936 (OR = 1.84, 95% CI = 1.60–2.13, $P = 6.04 \times 10^{-17}$). Of the remaining loci, the 2 correlated *PIP4K2A* SNPs evaluated showed an association with childhood ALL (rs10828317, OR = 0.76, 95% CI = 0.65–0.88, $P = 3.03 \times 10^{-4}$) as well. The *GATA3* rs3824662 association was only suggestive (OR = 1.15, 95% CI = 1.00–1.33, $P = 0.058$), but was further supported by the presence of a nearby SNP in LD (rs2275806, $r^2 = 0.72$) that showed a stronger association (OR = 1.20, 95% CI = 1.04–1.38, $P = 0.011$). The *WVOX* SNP, rs1121404, recently identified to be associated with childhood ALL in Chinese, showed no association in Japanese (OR = 1.04, 95% CI = 0.90–1.19, $P = 0.623$).

Risk allele frequencies and association estimates across various races/ethnicities are presented in Table 2. Among the loci identified through GWAS, only *ARID5B* SNPs showed a consistent association across the race/ethnic populations despite marked differences in allele frequencies. Although only marginally significant in Chinese (rs7088318, OR = 1.23, $P = 0.047$), the *PIP4K2A* association also showed consistency across populations. Primary SNPs first reported in populations of European ancestry for *IKZF1* (rs4132601 and rs11978267), then subsequently replicated in Hispanics and African Americans, showed no association in both Chinese and Japanese. The risk allele frequencies for the SNPs in Japanese (approximately 0.10) are markedly lower than frequencies observed in the original GWAS populations (approximately 0.20–0.30). Risk-associated SNPs recently identified in *LHPP* and *ELK3* in Europeans have not yet been reported in other populations. In Japanese, rs35837782 in *LHPP* and rs4762284 in *ELK3* did not show an association.

Using available SNP data across all 10 genetic loci including 10-kb flanking regions on both ends of the target genes, B-cell ALL risk associations were identified for alternate SNPs in *IKZF1* (rs1451367, OR = 1.52, 95% CI = 1.28–1.80, $P = 1.9 \times 10^{-6}$) (Table 3 and Fig. 1). For the two genetic loci where the SNP associations directly transferred to Japanese, rs4245595 in *ARID5B* (OR = 1.86, $P = 2.1 \times 10^{-17}$) and rs12146350 in *PIP4K2A* (OR = 0.72, $P = 2.7 \times 10^{-5}$) showed slightly stronger p-values, and both were in strong LD ($r^2 > 0.90$) in Japanese with the originally reported respective SNPs. The rs4245595 *ARID5B* SNP is also in strong LD ($r^2 > 0.96$) with the recently reported functional SNP, rs7090445 (OR = 1.85, $P = 3.1 \times 10^{-17}$), identified by Studd *et al.* in which they showed influences on enhancer activity and *RUNX3* binding²⁶. For the remaining genetic loci, alternate SNPs with

Gene/SNP	European Ancestry				Hispanic Ancestry				African Ancestry				Chinese Ancestry			Japanese Ancestry ^d		
	RAF ^a	OR ^b	P	Ref	RAF ^a	OR ^b	P	Ref	RAF ^a	OR ^b	P	Ref	RAF ^a	OR ^b	P	RAF ^a	OR ^b	P
ARID5B																		
rs10821936	0.33	1.91	1 × 10 ⁻¹⁵	²⁷	0.47	1.95	4 × 10 ⁻¹¹	⁴²	0.24	1.52	4 × 10 ⁻³	⁴²	0.38	1.43	5 × 10 ⁻⁴	0.35	1.84	6 × 10 ⁻¹⁷
IKZF1																		
rs4132601	0.28	1.69	1 × 10 ⁻¹⁹	²⁸	0.27	1.46	1 × 10 ⁻³	³²	—	—	—	—	0.14	1.20	0.197	0.09	1.19	0.164
rs11978267	0.27	1.69	9 × 10 ⁻¹¹	²⁷	0.26	1.31	0.01	⁴²	0.19	1.59	5 × 10 ⁻³	⁴²	—	—	—	0.10	1.13	0.321
CEBPE																		
rs2239633	0.52	1.34	2.9 × 10 ⁻⁷	²⁸	0.61	1.35	6.6 × 10 ⁻³	³²	—	—	—	—	0.64	1.18	0.117	0.53	1.08	0.282
rs4982731	0.28	1.29	9.1 × 10 ⁻⁶	⁴²	0.39	1.58	2.3 × 10 ⁻⁶	⁴²	0.38	1.13	0.410	⁴²	—	—	—	0.13	1.11	0.306
CDKN2A, CDKN2B																		
rs3731217	0.86	1.41	3 × 10 ⁻¹¹	³⁴	0.88 ^c	1.76	5 × 10 ⁻³	³²	—	—	—	—	0.80	1.04	0.769	0.82	0.90	0.251
rs662463	0.10	1.48	2 × 10 ⁻¹⁰	³	0.07	1.45	0.034	³	0.12	1.55	3 × 10 ⁻³	³	—	—	—	0.01	0.85	0.625
rs17756311	0.09	1.43	3 × 10 ⁻⁵	⁴²	0.06	1.36	0.100	⁴²	0.10	1.12	0.620	⁴²	—	—	—	0.01	0.88	0.695
PIP4K2A																		
rs7088318	0.59	1.25	5 × 10 ⁻⁶	⁴²	0.75	1.42	9 × 10 ⁻³	⁴²	0.39	1.65	1 × 10 ⁻³	⁴²	0.58	1.23	0.047	0.61	1.32	2 × 10 ⁻⁴
GATA3																		
rs3824662	0.17	1.31	9 × 10 ⁻¹²	⁴⁰	—	1.23	0.046	⁴³	—	—	—	—	0.32	1.32	0.013	0.31	1.15	0.058
LHPP																		
rs35837782	0.62	1.21	1 × 10 ⁻¹¹	⁴	—	—	—	—	—	—	—	—	—	—	—	0.61	0.94	0.442
ELK3																		
rs4762284	0.29	1.19	8 × 10 ⁻⁹	⁴	—	—	—	—	—	—	—	—	—	—	—	0.57	1.08	0.303
WVOX																		
rs1121404	—	—	—	—	—	—	—	—	—	—	—	—	0.28	1.38	5 × 10 ⁻¹⁰	0.38	1.04	0.623

Table 2. Summary of genetic variants and childhood ALL risk associations across races/ethnicities identified through genome-wide association studies. Abbreviations: RAF, risk allele frequency; OR, odds ratio; Ref, reference; SNP, single nucleotide polymorphism. ^aFrequency of the allele in controls conferring an increased risk of childhood ALL. ^bOdds ratios indicate the risk associated with the each additive increase in risk conferring allele. ^cRisk allele frequency obtained from the Human Genome Diversity Project (HGDP) ^dCalculations pertain to the designated risk allele indicated in the primary report. Risk estimate results may not be greater than 1.0 in East Asians.

a nominal p-values of less than 0.05 were identified, but were not statistically significant after adjustment for the number of SNPs tested across the respective regions.

To examine whether the non-transferability of association may be due to population differences in regional LD structure, varLD scores were calculated using 1000 Genomes Project Phase 3 data for Japanese, Han Chinese, and populations of European ancestry (Table 3). The regions surrounding the *ARID5B* and *PIP4K2A* SNPs, loci that directly replicated in Japanese, did not show statistically significant evidence of regional LD differences between populations of European ancestry and Japanese. Regions surrounding the *IKZF1* SNPs also showed minimal evidence of regional LD differences, but the previous GWAS-identified SNP associations did not directly transfer to Japanese. However, alternate statistically significant SNPs within *IKZF1* were identified (described above). With the exception of *LHPP* and *WVOX*, the four additional genetic loci in which the association did not transfer to Japanese showed strong evidence of regional LD structure differences based on varLD evaluations.

Discussion

Aided by successful validation across multiple populations, the genome-wide association analysis approach has led to the identification of several genetic loci involved in childhood ALL risk^{1,2}. However, there is still uncertainty about the role of these loci and consistency of specific SNP associations in East Asians with the majority of robust studies being performed primarily in populations of European and Hispanic ancestries. In our targeted evaluation of 16 previous GWAS-reported SNPs, we observed that the risk associations of those in *ARID5B* and *PIP4K2A* directly transfer to the Japanese population. The involvement of *IKZF1* is also supported by the identification of alternate associated SNPs in proximity to the originally reported loci, and the *GATA3* locus appears suggestive. However, this leaves the associations in the six remaining genes without clear evidence for a role in childhood ALL risk in East Asians. Examination of regional varLD scores showed that significant differences in LD between Japanese and the population in which the association was first reported were commonly observed in genes where the risk association did not transfer. Rather than concluding that the association is not present in Japanese, the varLD observations suggest that the associations may be obscured by differences in LD patterns and that other strategies are necessary to further clarify the role of the remaining six loci that did not transfer to this population.

Childhood ALL SNP associations in *ARID5B* first reported concurrently in studies performed in populations of European ancestry in the United States²⁷ and the United Kingdom²⁸ have been widely validated across multiple race/ethnic population²⁹, now including Japanese. The risk-conferring minor allele frequency of rs10821936

Locus	VarLD Evaluation (EUR vs JPN) ^a			Alternate SNP Association in Japanese within Gene						
	GWAS SNPs	Region (10-kb flanking) ^b	P _{adj} ^c	SNP (allele)	Position	Ca-MAF	Co-MAF	OR ^d	P _{nominal}	P _{adj} ^e
Loci associated in Japanese										
<i>ARID5B</i>	rs10994982, rs10821936, rs7089424	63,700,104–63,762,159	0.031	rs4245595 (T/C)	63,722,895	0.50	0.35	1.86	2 × 10 ⁻¹⁷	9 × 10 ⁻⁵
<i>IKZF1</i>	rs11978267, rs4132601	50,456,304–50,480,604	0.018	rs1451367 (C/T)	50,477,661	0.23	0.17	1.52	2 × 10 ⁻⁶	4 × 10 ⁻⁴
<i>PIP4K2A</i>	rs10828317, rs7088318	22,829,628–22,862,948	0.109	rs12146350 (G/C)	22,843,111	0.32	0.39	0.72	3 × 10 ⁻⁵	2 × 10 ⁻³
Association uncertain										
<i>CEBPE</i>	rs4982731, rs2239633	23,575,333–23,599,057	1 × 10 ⁻⁴	rs4981457 (G/A)	23,578,089	0.36	0.39	0.86	0.041	0.640
<i>CDKN2A/B</i>	rs3731217, rs17756311	21,974,661–22,063,895	1 × 10 ⁻⁴	rs21986535 (A/G)	21,986,535	0.04	0.03	1.28	0.169	0.965
<i>GATA3</i>	rs3824662	8,094,208–8,114,208	1 × 10 ⁻⁴	rs2275806 (A/G)	8,095,340	0.41	0.36	1.20	0.011	0.310
<i>LHPP</i>	rs35837782	126,283,309–126,303,309	0.053	rs113148868 (G/A)	126,275,824	0.21	0.18	1.24	0.014	0.862
<i>ELK3</i>	rs4762284	96,602,762–96,622,762	8 × 10 ⁻³	rs2075362 (A/G)	96,606,889	0.38	0.41	0.90	0.145	0.997
<i>WWOX</i> ^a	rs1121404	79,079,869–79,099,869	0.187	rs2738652 (A/T)	78,494,356	0.37	0.42	0.78	6 × 10 ⁻⁴	0.617

Table 3. Comparison of linkage disequilibrium of ALL-associated genomic regions between populations of European and Japanese ancestry and alternate SNP associations in Japanese. Abbreviations: Ca, cases; Co, controls; EUR, European ancestry; GWAS, genome-wide association study; JPN, Japanese ancestry; kb, kilobase; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphism; varLD, variation in linkage disequilibrium. ^aUsing 1000 Genomes Project Phase 3 data, regional LD patterns in Japanese were compared to patterns in the population for which the SNP association was first reported. *WWOX* varLD evaluations were performed compared to the Han Chinese population. ^bRegions included the evaluated SNPs and an additional 10-kb span flanking both ends. Genomic positions are based on the human genome assembly GRCh37 coordinates. ^cAdjusted p-values (P_{adj}) were based on 10,000 permutations of the combined data comprising the two populations being compared. An adjusted p-value of less than 0.006 based on a Bonferroni correction was considered statistically significant since 9 genomic regions were being tested. ^dOdds ratios indicate the risk associated with the each additive increase in minor allele. ^eAdjusted p-values (P_{adj}) were based on 10,000 permutations of case-control status and considered p-values below a family-wise type I error rate threshold of 0.05 to be statistically significant.

in Japanese is similar to that of Europeans (MAF ~ 0.35), but is significantly higher in Hispanics (MAF ~ 0.45) and lower in populations of African ancestry (MAF ~ 0.20). Interestingly, this pattern is similar to the relative population differences in incidence of childhood ALL and evidence supports a role for this locus in partially explaining this difference. Based on available data from St. Jude Children's Research Hospital and descriptive statistics from the Surveillance, Epidemiology, and End Results Program in the US, it was estimated that about 30% of the observed racial differences in ALL incidence may be attributable to the higher frequency of the rs10821936 risk allele in non-Hispanic whites compared to blacks³⁰. Characterization of genetic ancestry of the Children's Oncology Group Hispanic population showed increasing rs10821936 risk allele frequencies with increasing percentages of Native American ancestry³¹. Building on this observation, the California Childhood Leukemia Study reported increasing proportions of Native American ancestry to be associated with increasing risk of childhood ALL and showed that *ARID5B* contributes directly to the higher incidence in Hispanics compared to non-Hispanic whites³². However, the contribution of *ARID5B* is less clear in relation to Japanese given that this SNP has similar frequency and magnitude of effect as non-Hispanic whites despite known differences in incidence between the two populations.

Although consistently replicating in populations of European ancestry⁹, similar to studies performed in Chinese⁵, the *IKZF1* SNP association did not transfer to the Japanese population. Comparison of LD patterns based on varLD score between Europeans and Japanese did not show evidence of marked difference across an approximately 25-kb region comprising the previously reported SNPs. However, the allele frequency of the SNPs are considerably lower in East Asians at about 0.10 or less compared to close to 0.30 in Europeans and Hispanics. The ability to analyze the effect of other SNPs across the flanking regions led to the identification of an alternate associated SNP (rs1451367) located within about 10-kb that is common in East Asians (MAF ~ 0.20), but rare in Europeans (MAF < 0.01). This suggests that variation in *IKZF1* is also associated with risk of childhood ALL in Japanese; however, it cannot be concluded yet whether the SNP associations are representing the same causal locus across the populations.

Based on the results of the current analysis, evidence for childhood ALL risk associations with GWAS-identified SNPs in *CEBPE*, *CDKN2A*, *CDKN2B*, *LHPP*, *ELK*, and *WWOX* is lacking. Associations represented by other SNPs potentially tagging a causal locus within these genes were also not apparent. While the evidence is still limited, results could be influenced by differences in a gene-environment effect across populations not appropriately captured, or it may be possible that certain common SNPs identified in GWAS may be representing associations with rare causal variant(s) on the same haplotype background of the GWAS-identified tag SNP³³. If rare causal variants are at play, even modest differences in haplotype structure of the regions may significantly affect detection potential, or it is possible that the variants may not be present in Japanese. As an example, the *CDKN2A* risk association originally identified through GWAS based on the common variant rs3731217³⁴ was

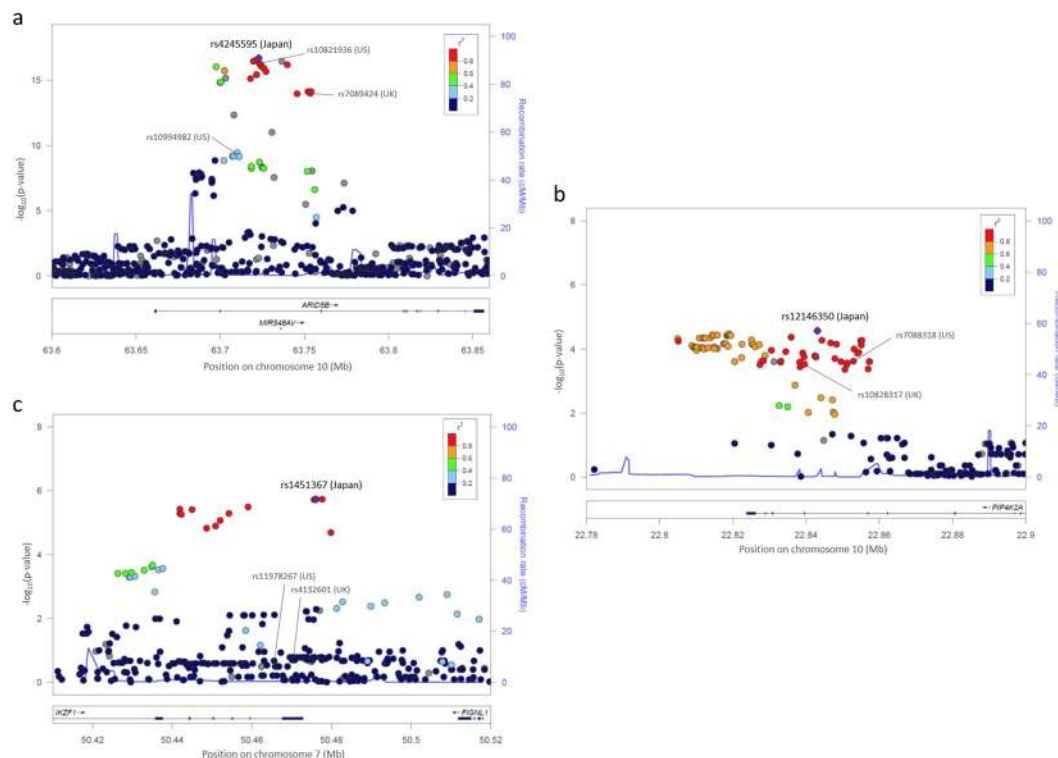


Figure 1. Regional plot of results of the association between SNPs in (a) *ARID5B*, (b) *PIP4K2A*, and (c) *IKZF1* and risk of childhood B-cell ALL. Multiple logistic regression was performed assuming a log-additive genetic model and adjusting for 10 principal components eigenvectors. The $-\log_{10}$ (p-value) for each SNP are plotted against their chromosomal position. The purple diamond (◆) indicates the strongest associated SNP in the region and the colors of the dots (●) represent the degree of linkage disequilibrium (based on r^2) in relation to that index SNP in Japanese. Recombination rates (cM/Mb) overlay the plots. Coordinates are based on human genome assembly GRCh37 build.

recently shown to be explained by a rare high-impact coding variant (rs3731249)^{35–37}. This variant is present in about two percent of Europeans, but is not present in Japanese. With the exception of *LHPP*, all loci for which the associations did not transfer showed evidence of differences in genetic architecture between Japanese and Europeans based on varLD score, whereas those that transferred did not show marked differences. In line with the common disease/common variant hypothesis³⁸, if the GWAS associations are instead tagging a common causal variant and assuming this variant is operative as a risk locus in Japanese as well, we would have expected the regional SNP coverage and statistical power of the current study to be sufficient to detect the association signal. The lack of association suggests a need for future studies to consider characterization of rare variants in order to fully understand the nature of these GWAS loci in Japanese.

Certain limitations inherent to this study may have also affected the results. Although our study was limited to B-cell lineage ALL similar to most previous GWAS, availability of molecular subtype data was incomplete for a large proportion of the patients. While heterogeneity by subtype in the magnitude of risk has been observed for several of the loci, effects exclusive to a specific subtype have not been clearly demonstrated and is likely not the reason for the lack of association observed. One exception may be the *GATA3* risk locus identified in a GWAS of Ph-like ALL³⁹ and another study that observed the association specifically in non-hyperdiploid B-cell ALL that lack the *ETV6-RUNX1* fusion⁴⁰. Results for the *GATA3* variant (rs3824662) in the current study were suggestive of an association among the total B-cell ALL series, but requires further evaluation in a subtype specific analysis for confirmation. Also, access to patients for recruitment into this study was through the outpatient mechanism which resulted in a study population of surviving patients. This may have led to over-representation of patients of certain disease profiles; however a 80 to 85 percent survival rate of childhood ALL, as reported by the TCCSG⁴¹, suggests that the effect may have been minimal since our objectives focused on validating known GWAS hits, those of which were originally identified using general ALL patient populations that comprised of the most common ALL subtypes (versus a sequencing-based design targeting rare subtypes of poor prognosis). Finally, our data included imputed genotypes to enhance the coverage of genetic variation across the targeted genomic regions. Despite stringent quality control measures and advances in imputation methodologies, uncertainty still exists and may have introduced non-differential misclassification of genotypes and a reduction in statistical power to detect associations.

In this targeted evaluation of SNPs across regions previously identified in GWAS of childhood ALL, we showed that variation in *ARID5B*, *IKZF1*, *PIP4K2A*, and possibly *GATA3* contribute to the genetic susceptibility of childhood B-cell ALL in Japanese. There is a need to account for population-specificity in producing accurate

risk prediction estimates based on inherited genetic variation. Thus, this analysis serves as the first step towards characterizing the role of genetic variation in the susceptibility to childhood ALL in the Japanese population. Identification of potential novel loci, perhaps specific to the East Asian population or those more detectable due to enhanced LD with a causal locus and/or allele frequency differences, may be possible through a genome-wide association analysis after expansion of this population for increased statistical power.

References

1. Moriyama, T., Relling, M. V. & Yang, J. J. Inherited genetic variation in childhood acute lymphoblastic leukemia. *Blood* **125**, 3988–3995, <https://doi.org/10.1182/blood-2014-12-580001> (2015).
2. Urayama, K. Y., Chokkalingam, A. P., Manabe, A. & Mizutani, S. Current evidence for an inherited genetic basis of childhood acute lymphoblastic leukemia. *Int J Hematol* **97**, 3–19, <https://doi.org/10.1007/s12185-012-1220-9> (2013).
3. Hungate, E. A. *et al.* A variant at 9p21.3 functionally implicates CDKN2B in paediatric B-cell precursor acute lymphoblastic leukaemia aetiology. *Nat Commun* **7**, 10635, <https://doi.org/10.1038/ncomms10635> (2016).
4. Vijayakrishnan, J. *et al.* A genome-wide association study identifies risk loci for childhood acute lymphoblastic leukemia at 10q26.13 and 12q23.1. *Leukemia*, <https://doi.org/10.1038/leu.2016.271> (2016).
5. Shi, Y. *et al.* Identification of a novel susceptibility locus at 16q23.1 associated with childhood acute lymphoblastic leukemia in Han Chinese. *Hum Mol Genet* **25**, 2873–2880, <https://doi.org/10.1093/hmg/ddw112> (2016).
6. Hunter, D. J. Lessons from genome-wide association studies for epidemiology. *Epidemiology* **23**, 363–367, <https://doi.org/10.1097/EDE.0b013e31824da7cc> (2012).
7. Gibson, G. Hints of hidden heritability in GWAS. *Nat Genet* **42**, 558–560, <https://doi.org/10.1038/ng0710-558> (2010).
8. Urayama, K. Y. & Manabe, A. Genomic evaluations of childhood acute lymphoblastic leukemia susceptibility across race/ethnicities. *[Rinsho ketsueki] The Japanese Journal of Clinical Hematology* **55**, 2242–2248 (2014).
9. Dai, Y. E., Tang, L., Healy, J. & Sinnett, D. Contribution of polymorphisms in IKZF1 gene to childhood acute leukemia: a meta-analysis of 33 case-control studies. *PLoS One* **9**, e113748, <https://doi.org/10.1371/journal.pone.0113748> (2014).
10. Sun, J. *et al.* Association between CEBPE Variant and Childhood Acute Leukemia Risk: Evidence from a Meta-Analysis of 22 Studies. *PLoS One* **10**, e0125657, <https://doi.org/10.1371/journal.pone.0125657> (2015).
11. Ishida, Y. *et al.* Secondary cancers among children with acute lymphoblastic leukaemia treated by the Tokyo Children's Cancer Study Group protocols: a retrospective cohort study. *Br J Haematol* **164**, 101–112, <https://doi.org/10.1111/bjh.12602> (2014).
12. Kato, M. *et al.* Long-term outcome of 6-month maintenance chemotherapy for acute lymphoblastic leukemia in children. *Leukemia* **31**, 580–584, <https://doi.org/10.1038/leu.2016.274> (2017).
13. Muro, S. *et al.* Relationship Among Chlamydia and Mycoplasma Pneumoniae Seropositivity, IKZF1 Genotype and Chronic Obstructive Pulmonary Disease in A General Japanese Population: The Nagahama Study. *Medicine* **95**, e3371, <https://doi.org/10.1097/MD.0000000000003371> (2016).
14. Seow, W. J. *et al.* Association between GWAS-identified lung adenocarcinoma susceptibility loci and EGFR mutations in never-smoking Asian women, and comparison with findings from Western populations. *Hum Mol Genet* **26**, 454–465, <https://doi.org/10.1093/hmg/ddw414> (2017).
15. Izuhara, Y. *et al.* Mouth breathing, another risk factor for asthma: the Nagahama Study. *Allergy* **71**, 1031–1036, <https://doi.org/10.1111/all.12885> (2016).
16. Inoue, M. *et al.* Epidemiology of pancreatic cancer in Japan: a nested case-control study from the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC). *International journal of epidemiology* **32**, 257–262 (2003).
17. Yamaguchi-Kabata, Y. *et al.* Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet* **83**, 445–456, <https://doi.org/10.1016/j.ajhg.2008.08.019> (2008).
18. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909, <https://doi.org/10.1038/ng1847> (2006).
19. O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nat Genet* **48**, 817–820, <https://doi.org/10.1038/ng.3583> (2016).
20. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–1287, <https://doi.org/10.1038/ng.3656> (2016).
21. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74, <https://doi.org/10.1038/nature15393> (2015).
22. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575, <https://doi.org/10.1086/519795> (2007).
23. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337, <https://doi.org/10.1093/bioinformatics/btq419> (2010).
24. Ong, R. T. & Teo, Y. Y. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. *Bioinformatics* **26**, 1269–1270, <https://doi.org/10.1093/bioinformatics/btq125> (2010).
25. Teo, Y. Y. *et al.* Genome-wide comparisons of variation in linkage disequilibrium. *Genome Res* **19**, 1849–1860, <https://doi.org/10.1101/gr.092189.109> (2009).
26. Studd, J. B. *et al.* Genetic and regulatory mechanism of susceptibility to high-hyperdiploid acute lymphoblastic leukaemia at 10p21.2. *Nat Commun* **8**, 14616, <https://doi.org/10.1038/ncomms14616> (2017).
27. Trevino, L. R. *et al.* Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet* **41**, 1001–1005, <https://doi.org/10.1038/ng.432> (2009).
28. Papaemmanuil, E. *et al.* Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet* **41**, 1006–1010, <https://doi.org/10.1038/ng.430> (2009).
29. Zeng, H. *et al.* Associations between AT-rich interactive domain 5B gene polymorphisms and risk of childhood acute lymphoblastic leukemia: a meta-analysis. *Asian Pac J Cancer Prev* **15**, 6211–6217 (2014).
30. Yang, W. *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* **24**, 894–896, <https://doi.org/10.1038/leu.2009.277> (2010).
31. Xu, H. *et al.* ARID5B genetic polymorphisms contribute to racial disparities in the incidence and treatment outcome of childhood acute lymphoblastic leukemia. *J Clin Oncol* **30**, 751–757, <https://doi.org/10.1200/JCO.2011.38.0345> (2012).
32. Walsh, K. M. *et al.* Associations between genome-wide Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children. *Leukemia* **27**, 2416–2419, <https://doi.org/10.1038/leu.2013.130> (2013).
33. Wang, K. *et al.* Interpretation of association signals and identification of causal variants from genome-wide association studies. *Am J Hum Genet* **86**, 730–742, <https://doi.org/10.1016/j.ajhg.2010.04.003> (2010).
34. Sherborne, A. L. *et al.* Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nat Genet* **42**, 492–494, <https://doi.org/10.1038/ng.585> (2010).
35. Vijayakrishnan, J. *et al.* The 9p21.3 risk of childhood acute lymphoblastic leukaemia is explained by a rare high-impact variant in CDKN2A. *Sci Rep* **5**, 15065, <https://doi.org/10.1038/srep15065> (2015).

36. Walsh, K. M. *et al.* A Heritable Missense Polymorphism in CDKN2A Confers Strong Risk of Childhood Acute Lymphoblastic Leukemia and Is Preferentially Selected during Clonal Evolution. *Cancer Res* **75**, 4884–4894, <https://doi.org/10.1158/0008-5472.CAN-15-1105> (2015).
37. Xu, H. *et al.* Inherited coding variants at the CDKN2A locus influence susceptibility to acute lymphoblastic leukaemia in children. *Nat Commun* **6**, 7553, <https://doi.org/10.1038/ncomms8553> (2015).
38. Chakravarti, A. Population genetics—making sense out of sequence. *Nat Genet* **21**, 56–60, <https://doi.org/10.1038/4482> (1999).
39. Perez-Andreu, V. *et al.* Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat Genet* **45**, 1494–1498, <https://doi.org/10.1038/ng.2803> (2013).
40. Migliorini, G. *et al.* Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype. *Blood* **122**, 3298–3307, <https://doi.org/10.1182/blood-2013-03-491316> (2013).
41. Tsuchida, M. *et al.* Long-term results of Tokyo Children's Cancer Study Group trials for childhood acute lymphoblastic leukemia, 1984–1999. *Leukemia* **24**, 383–396, <https://doi.org/10.1038/leu.2009.260> (2010).
42. Xu, H. *et al.* Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. *J Natl Cancer Inst* **105**, 733–742, <https://doi.org/10.1093/jnci/djt042> (2013).
43. Walsh, K. M. *et al.* GATA3 risk alleles are associated with ancestral components in Hispanic children with ALL. *Blood* **122**, 3385–3387, <https://doi.org/10.1182/blood-2013-08-524124> (2013).

Acknowledgements

This work was supported by funding from St. Luke's Life Science Institute (Tokyo, Japan), the Children's Cancer Association of Japan, and the Japan Leukemia Research Fund. Genotyping was partially supported by Japan Society for the Promotion of Science KAKENHI Grant number 26253041. We would like to thank the patients and families participating in this study, and staff of the collaborating hospitals for their various contributions. This study made use of data from the 1000 Genomes Project (<http://www.internationalgenome.org/data>) and the International HapMap Project.

Author Contributions

K.Y.U., A.M., M.T., S.M., Y.I. and F.M. conceived and designed the study; Y.T., Y.A., D.H., Y.Y., T.K., Y.N., Y.T., S.O., T.I., M.Y., D.K., Kaz.K., D.T., Y.N., H.K., K.N., Ko.M., H.G., Y.S., D.M., M.K., J.T., Kat.K., Y.I., A.O. and A.M. were involved in patient recruitment and sample collection; M.T., Ke.M., T.K., F.M., T.T. and J.I. led the laboratory analyses and generation of genomic data; K.Y.U. and T.K. and conducted the statistical analysis and bioinformatics evaluations; K.Y.U. drafted the first version of the manuscript. All authors critically reviewed and edited the manuscript for intellectual content and gave final approval of the final version.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018