# Genome-wide assessment of genetic risk loci for childhood acute lymphoblastic leukemia in Japanese patients

Acute lymphoblastic leukemia (ALL) is the most common malignancy among children in industrialized countries, with a peak incidence between 2 and 5 years of age.[1] The early onset of this cancer and heterogeneity in incidence by race and ethnicity implicates the influence of inherited genetic susceptibility in which evidence from genome-wide association studies (GWAS) of childhood ALL have identified several genomic regions associated with risk.[2] To date, the identification of risk loci has been driven by studies conducted in populations of Hispanic or European ancestry, with a paucity of genome-wide studies performed in Asian populations.[3] Pursuit of potential population-specific loci through genome-wide assessment and characterization of known loci across diverse populations is important to advance our understanding of inherited genetic variation in the risk childhood ALL.

Our previous study of targeted loci conducted within the Tokyo Children Cancer Study Group (TCCSG) showed that risk associations for single nucleotide polymorphism (SNP) in *ARID5B*, *IKZF1* and *PIP4K2A* transfer to the Japanese population.[4] As a next step, this current study included two independent GWAS series assembled through TCCSG and the Japanese Pediatric Leukemia/Lymphoma Study Group (JPLSG), including a total of 1,088 cases and 5,315 controls, in the first comprehensive evaluation of genetic variation in the risk of childhood ALL in Japanese. The first series (TCCSG GWAS) comprised patients from the TCCSG clinical network,[4, 5] and included childhood ALL patients diagnosed at age 19 years or younger prior to 2012 (N=621) from outpatient clinic visits between 2013 and 2015 through a convenience sampling approach. Controls comprised adult participants from the Nagahama Prospective Cohort for Comprehensive Human Bioscience (the Nagahama Study) (N=1,846) and the Hospital-based Epidemiologic Research Program at Aichi Cancer Center (HERPACC) Study (N=2,170).[6, 7] The second series (JPLSG GWAS) comprised childhood B-cell precursor (BCP) ALL patients (N=572) aged 1 to 19 years newly diagnosed between 2012 and 2018 through the nationwide ALL-B12 clinical study (*registry: UMIN000009339*).[5, 8] Controls comprised a subset of participants from the Nagahama Study (N=1,924). DNA were extracted from saliva samples (TCCSG) or peripheral blood at remission (JPLSG) and were genotyped with the Illumina HumanCoreExome and OmniExpress microarrays, respectively. Institutional review board approvals were obtained from St. Luke's International University and the major collaborating centers.

We performed quality control (QC) steps separately for the TCCSG and JPLSG cases and controls followed by additional QC filters after merging the case-control series for the TCCSG and JPLSG GWAS separately. After sample and SNP exclusions based on a standard QC approach (*Online Supplementary Figure S1*), a total of 258,069 and 481,270 directly genotyped SNP was available for the TCCSG GWAS (540 cases and 3,714 controls) and JPLSG GWAS (548 cases and 1,601 controls), respectively. Genome-wide SNP imputation was performed using ShapeIT2 and Minimac4 with an in-house Japanese haplotype reference panel. After post-imputation QC, bi-allelic loci shared between the TCCSG and JPLSG case-control series resulted in data for a total of 6,446,781 SNP for both GWAS.

Patients included in the TCCSG and JPLSG series comprised predominately of B-cell ALL (TCCSG, 93%; JPLSG, 100%), greater numbers of males (TCCSG, 52%; JPLSG, 58%) than females, and showed the majority to be between 1 and 6 years of age (TCCSG, 69%; JPLSG, 57%). We first performed a discovery analysis in the TCCSG series and observed a novel association represented by SNP rs116977518 (odds ratio [OR] =1.99, $P$=4.2x10$^{-9}$) at 1q24.1 (intergenic, proximity to *FMO8P*) and an association at a known region represented by rs4245595 (OR=1.84, $P$=3.4x10$^{-17}$) located at 10q21.2 (*ARID5B*) (Table 1; *Online Supplementary Figure S2A*). An association with the previously identified *IKZF1* region was also found (rs77563422, OR=1.62, $P$=9.5x10$^{-8}$). Only SNP in *ARID5B* (rs4245595, OR=1.82, $P$=2.0x10$^{-10}$) and *IKZF1* (rs77563422, OR=1.44, $P$=0.002) replicated in the JPLSG series (Table 1). Confirmation is still necessary for the putative risk locus at 1q24.1. This locus is located adjacent to the *FMO8P* and *FMO9P* pseudogenes, and contains expression quantitative trait loci (eQTL) in blood for the deoxyuridine triphosphatase pseudogene 6 (*DUTP6*) gene as documented in the Genotype-Tissue Expression (GTEx) portal. In a gene expression profiling study of tonsil squamous cell carcinoma, *DUTP6*, along with other pseudogenes and small nuclear RNA, were found to be upregulated in blood mononuclear cells of patients compared to controls.[9] Interestingly, the leading SNP in this region, rs116977518, is rare or not present in most other racial and ethnic populations.

Next, we performed a discovery analysis in the JPLSG GWAS, and observed an association with the known *ARID5B* region (rs4506592, OR=1.85, $P$=5.7x10$^{-11}$), along with another region at 6q23.1 in the sterile α motif domain containing 3 (*SAMD3*) gene (rs137991838, OR=0.21, $P$=1.9x10$^{-8}$) (Table 1; *Online Supplementary Figure S2B*). The novel *SAMD3* SNP association did not appear to replicate in the overall TCCSG case-control series, but limiting to only B-cell ALL showed a reduced risk (rs137991838, OR=0.67, $P$=0.046). The *SAMD3* gene exhibits the highest expression levels in lymphoid tissues and blood.[10] It belongs to the sterile α motif (SAM) domain superfamily in

which the characteristic SAM domain suggests involvement in diverse protein-protein interactions important in assembly, regulation, and localization of functional elements.[11] The leading SNP, rs137991838, is unique to the Japanese population and resides within a region that contain eQTL for *SAMD3* in lymphoblastoid cell lines according to GTEx and RegulomeDB. Chromosomal aberrations of the 6q23 region are known to be common across a diverse range of tumor types, including hematologic malignancies.[12]

In a genome-wide SNP meta-analysis of the TCCSG and JPLSG GWAS combined, three SNP representing regions with genome-wide significant associations included rs77563422 (*IKZF1*, OR=1.55, $P$=5.9x10$^{-10}$), and two uncorrelated SNP in *ARID5B* separated by about 38 kb ($r^2$=0.07), rs2393784 (OR=1.52, $P$=6.3x10$^{-13}$) and rs7896246 (OR=1.83, $P$=1.4x10$^{-25}$) (Table 2; Figure 1). Replication opportunities of discovery

**Table 1.** Results of top single nucleotide polymorphisms ($P$<1.0×10$^{-6}$) associated with childhood acute lymphoblastic leukemia risk observed in the TCCSG and JPLSG genome-wide association analyses.

| SNP | Chr: position[a] | Gene | Allele[b] | Discovery stage (genome-wide) | | | | Replication in TCCSG or JPLSG | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Cases MAF | Controls MAF | OR (95% CI)[c] | *P* | Cases MAF | Controls MAF | OR (95% CI)[c] | *P* |
| **TCCSG GWAS** | | | | | | | | | | | |
| rs116977518 | 1:166529240 | - | C / T | 0.14 | 0.08 | 1.99 (1.59-2.48) | 4.2x10$^{-9}$ | 0.11 | 0.08 | 1.14 (0.85-1.54) | 0.387 |
| rs2516643 | 6:30505103 | - | A / G | 0.31 | 0.39 | 0.67 (0.58-0.78) | 1.9x10$^{-7}$ | 0.36 | 0.37 | 1.01 (0.84-1.22) | 0.907 |
| rs77563422 | 7:50454209 | *IKZF1* | G / C | 0.23 | 0.16 | 1.62 (1.36-1.93) | 9.5x10$^{-8}$ | 0.23 | 0.16 | 1.44 (1.14-1.81) | 0.002 |
| rs4245595 | 10:63722895 | *ARID5B* | C / T | 0.50 | 0.35 | 1.84 (1.59-2.12) | 3.4x10$^{-17}$ | 0.48 | 0.34 | 1.82 (1.51-2.19) | 2.0x10$^{-10}$ |
| rs79649658 | 15:28063650 | *OCA2* | G / A | 0.01 | 0.03 | 0.22 (0.11-0.45) | 4.5x10$^{-7}$ | 0.03 | 0.02 | 1.50 (0.85-2.66) | 0.166 |
| **JPLSG GWAS** | | | | | | | | | | | |
| rs151235219 | 3:43088058 | *GASK1A* | C / A | 0.06 | 0.02 | 3.63 (2.22-5.91) | 2.4x10$^{-7}$ | 0.04 | 0.03 | 1.06 (0.72-1.56) | 0.757 |
| rs137991838 | 6:130620077 | *SAMD3* | G / C | 0.02 | 0.05 | 0.21 (0.12-0.39) | 1.9x10$^{-8}$ | 0.04 | 0.04 | 0.75 (0.52-1.09) | 0.118 |
| rs4506592 | 10:63727187 | *ARID5B* | A / G | 0.47 | 0.33 | 1.85 (1.53-2.23) | 5.7x10$^{-11}$ | 0.49 | 0.34 | 1.81 (1.57-2.09) | 2.0x10$^{-16}$ |

[a]Chromosome and genomic positions are based on GRCh37/hg19. [b]Minor/major allele. [c]Odds ratio and 95% confidence intervals associated with the minor allele were calculated using logistic regression assuming a log-additive genetic model and adjusting for genetic ancestry; a $P$<5.0x10$^{-8}$ was considered genome-wide significant. TCCSG: Tokyo Children Cancer Study Group; JPLSG: Japanese Pediatric Leukemia/Lymphoma Study Group; GWAS: genome-wide association study; Chr: chromosome; CI: confidence interval; MAF: minor allele frequency; OR: odds ratio; SNP: single nucleotide polymorphism.

**Table 2.** Meta-analysis of TCCSG and JPLSG genome-wide association analysis results and independent replication in CCRLP.

| SNP | Chr: position[a] | Gene | Allele[b] | Cases MAF | Controls MAF | Meta-analysis (TCCSG + JPLSG) | | Replication (CCRLP) | | Combined | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | OR (95% CI)[d] | $P_{meta}$ | OR (95% CI)[d] | *P* | OR (95% CI)[d] | *P* |
| **Genome-wide significant SNP ($P$<5.0×10$^{-8}$ in meta-analysis)** | | | | | | | | | | | |
| rs77563422 | 7:50454209 | *IKZF1* | C / G | 0.23 | 0.16 | 1.55 (1.35-1.78) | 5.9x10$^{-10}$ | 1.64 (1.34-2.00) | 2.3x10$^{-6}$ | 1.58 (1.41-1.77) | 3.7x10$^{-15}$ |
| rs2393784 | 10:63685638 | *ARID5B* | G / A | 0.45 | 0.34 | 1.52 (1.35-1.70) | 6.3x10$^{-13}$ | 0.88 (0.72-1.07) | 0.194 | 1.33 (1.20-1.46) | 2.1x10$^{-8}$ |
| rs7896246 | 10:63724390 | *ARID5B* | G / A | 0.49 | 0.34 | 1.83 (1.63-2.05) | 1.4x10$^{-25}$ | 1.58 (1.32-1.89) | 6.4x10$^{-7}$ | 1.76 (1.60-1.93) | 1.3x10$^{-30}$ |
| **Suggestive SNP ($P$<5×10$^{-6}$ in meta-analysis)** | | | | | | | | | | | |
| rs116977518 | 1:166529240 | - | T / C | 0.14 | 0.08 | 1.63 (1.37-1.95) | 7.6x10$^{-8}$ | 1.35 (0.79-2.33) | 0.289 | 1.60 (1.35-1.90) | 5.0x10$^{-8}$ |
| rs61459905 | 2:60133425 | - | C / T | 0.32 | 0.27 | 1.39 (1.23-1.57) | 1.4x10$^{-7}$ | 0.91 (0.76-1.09) | 0.304 | 1.21 (1.10-1.34) | 1.6x10$^{-4}$ |
| rs4596201 | 4:160682930 | - | T / G | 0.04 | 0.02 | 2.14 (1.59-2.89) | 6.0x10$^{-7}$ | 1.14 (0.73-1.77) | 0.555 | 1.51 (1.24-1.85) | 3.6x10$^{-4}$ |
| rs11753269 | 6:166558778 | - | A / G | 0.03 | 0.01 | 2.70 (1.81-4.04) | 1.3x10$^{-6}$ | 1.09 (0.71-1.65) | 0.705 | 1.75 (1.31-2.34) | 1.7x10$^{-4}$ |
| rs142498523 | 7:46050938 | - | A / C | 0.03 | 0.01 | 2.52 (1.75-3.64) | 7.8x10$^{-7}$ | 1.04 (0.49-2.20) | 0.925 | 2.13 (1.53-2.96) | 7.4x10$^{-6}$ |
| rs9641181 | 7:95638299 | *DYNC1I1* | G / A | 0.04 | 0.06 | 0.56 (0.44-0.72) | 4.8x10$^{-6}$ | 0.86 (0.44-1.67) | 0.647 | 0.59 (0.47–0.74) | 9.1x10$^{-6}$ |
| rs142757968 | 9:12688552 | - | T / A | 0.02 | 0.01 | 2.77 (1.81-4.24) | 2.7x10$^{-6}$ | NA | NA | 2.77 (1.81-4.24) | 2.7x10$^{-6}$ |
| rs1331625 | 9:135059487 | *NTNG2* | C / G | 0.06 | 0.09 | 0.59 (0.47-0.74) | 4.9x10$^{-6}$ | 0.87 (0.64-1.20) | 0.405 | 0.67 (0.56-0.81) | 2.1x10$^{-5}$ |
| rs4750853 | 10:129193311 | *DOCK1* | G / A | 0.24 | 0.30 | 0.73 (0.64-0.83) | 8.6x10$^{-7}$ | 0.97 (0.78-1.21) | 0.787 | 0.78 (0.70-0.87) | 1.1x10$^{-5}$ |

[a]Chromosome and genomic positions are based on GRCh37/hg19. [b]Minor/major allele. [c]Odds ratio and 95% confidence intervals associated with the minor allele were calculated using logistic regression assuming a log-additive genetic model and adjusting for genetic ancestry; a $P$<5.0x10$^{-8}$ was considered genome-wide significant. TCCSG: Tokyo Children Cancer Study Group; JPLSG: Japanese Pediatric Leukemia/Lymphoma Study Group; CCRLP: California Cancer Records Linkage Poject; Chr: chromosome; CI: confidence interval; MAF: minor allele frequency; OR: odds ratio; SNP: single nucleotide polymorphism; NA: not applicable.

results were pursued within cases (N=318) and controls (N=5,107) of East Asian ancestry from the California Cancer Records Linkage Project (CCRLP), a study based on the birth population of California previously reported.[13] The associations were confirmed in this CCRLP replication series except for rs2393784 in *ARID5B* (Table 2). Conditional analysis of the two *ARID5B* SNP showed attenuation in effect size for both loci, but evidence of independent associations remained (rs2393784, OR=1.22, $P$=2.1x10$^{-3}$; rs7896246, OR=1.69, $P$=1.6x10$^{-16}$). rs2393784 is located about 38 kb upstream in intron 2, a SNP in LD (rs6479778) has been shown associated with both ALL relapse and disease risk in a US population.[14] *ARID5B* SNP associations represent some of the most consistently observed in childhood ALL susceptibility, all of which suggest a role for variation in intronic regions and thus, mechanisms that involve gene regulation through affecting RNA splicing, transcription factor binding, and other processes. In a UK study, fine-mapping in high-hyperdiploid ALL cases and controls identified two plausibly casual SNP in LD, one of which is the same top hit identified in the current study (rs7896246).[15]

The association between *IKZF1* and ALL risk has been confirmed repeatedly for rs4132601 and rs11978267 in populations of European, Hispanic, and African ancestry, but has been less clear for East Asians.[3] For both SNP, East Asians exhibit among the lowest allele frequencies (MAF~0.08), and previous studies in this population may have been hampered by statistical power. ALL associations replicated for both rs4132601 and rs11978267 ($P$<0.01). We also identified a genome-wide significant region in *IKZF1* (rs77563422), which is uncorrelated with the known risk locus ($r^2$<0.01), and results conditioning on the presence of rs4132601 resulted in a stronger effect

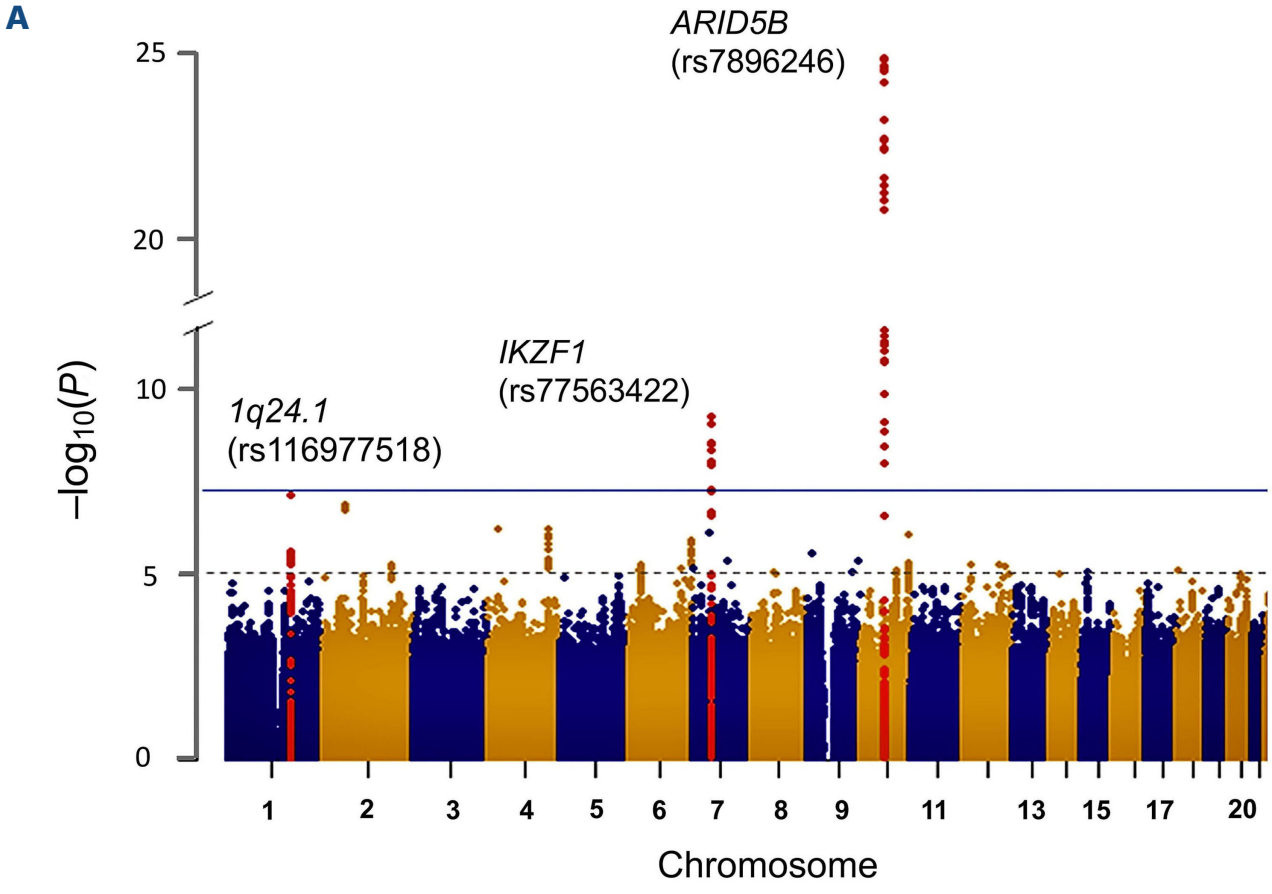


**Figure 1. Results of meta-analysis combining the TCCSG and JPLSG genome-wide association analysis results**. (A) The Manhattan plot shows the -log10 ($P$ value) of the logistic regression analysis plotted against the chromosomal position (GRCh37/hg19). The blue line indicates the genome-wide significant $P$ value threshold of 5x10$^{-8}$, and the dotted line was the threshold (1x10$^{-6}$) for considering results as suggestive. (B) A quantile-quantile plot of the observed *versus* expected distribution of the -log10 ($P$ value) from the analysis. TCCSG: Tokyo Children Cancer Study Group; JPLSG: Japanese Pediatric Leukemia/Lymphoma Study Group.

size and significance for both variants (rs77563422, OR=1.61, $P$=2.6x10$^{-11}$; rs4132601, OR=1.49, $P$=5.5x10$^{-5}$). SNP rs77563422 is rare in populations of European ancestry and is located in a different intronic region about 16 kb upstream of the other known variants.

We were able to confirm associations for known risk loci representing *ARID5B, IKZF1, DDC, CEBPE, PIP4K2A, GATA3, IKZF3*, and 8q24.21, with some showing a different leading SNP in Japanese (*Online Supplementary Table S1*). There are several reasons why certain associations may not have been detected, including insufficient statistical power due to lower allele frequencies and/or effect sizes, unavailable SNP data in sufficient LD with the causal locus, and analyses without similar subtype specificity as the original study. An overall limitation of the current study included limited access to molecular subtype data for this analysis. Notably, it is possible that the recruitment strategy of the TCCSG series may have over-represented patients with higher survival probabilities and specific molecular subtype profiles which could have affected replication attempts for loci that show subtype-specificity. In addition, the CCRLP replication population represented a broadly defined group of cases and controls of East Asian ancestry, and differing genetic substructure between Japanese and others of East Asian origins needs consideration in interpreting the failure to replicate.

In this first case-control GWAS effort in Japanese, we confirmed the strong ALL risk associations with *ARID5B* and *IKZF1* variation, and we report two putative ALL risk associations suggesting a role for the 1q24.1 region and *SAMD3*, but confirmation is necessary. Together with also characterizing the effects of known risk loci in Japanese, we expect this study to aid efforts in understanding the heritability of childhood ALL in this population, a key step for elucidating the causes of this devastating disease.

## Authors

Mayumi Hangai,[1,2*] Takahisa Kawaguchi,[3*] Masatoshi Takagi,[4] Keitaro Matsuo,[5] Soyoung Jeon,[6] Charleston W.K. Chiang,[6,7] Andrew T. Dewan,[8] Adam J. de Smith,[6] Toshihiko Imamura,[9] Yasuhiro Okamoto,[10] Akiko M. Saito,[11] Takao Deguchi,[12] Michiaki Kubo,[13] Yoichi Tanaka,[14] Yoko Ayukawa,[1] Toshinari Hori,[15] Kentaro Ohki,[16] Nobutaka Kiyokawa,[16] Takeshi Inukai,[17] Yuki Arakawa,[18] Makiko Mori,[18] Daisuke Hasegawa,[19] Daisuke Tomizawa,[12] Hiroko Fukushima,[20] Yuki Yuza,[21] Yasushi Noguchi,[22] Yuichi Taneyama,[23] Setsuo Ota,[24] Hiroaki Goto,[25] Masakatsu Yanagimachi,[25] Dai Keino,[25] Kazutoshi Koike,[26] Daisuke Toyama,[27] Yozo Nakazawa,[28] Kozue Nakamura,[29] Koichi Moriwaki,[30] Yujin Sekinaka,[31] Daisuke Morita,[28] Shinsuke Hirabayashi,[32] Yosuke Hosoya,[19] Yuri Yoshimoto,[33] Hiroki Yoshihara,[19] Miwa Ozawa,[19] Shinobu Kobayashi,[1] Naho Morisaki,[1] Tshewang Gyeltshen,[34] Osamu Takahashi,[34] Yukinori Okada,[35,36,37] Makiko Matsuda,[38] Toshihiro Tanaka,[38] Johji Inazawa,[39] Junko Takita,[40] Yasushi Ishida,[41] Akira Ohara,[42] Catherine Metayer,[43] Joseph L. Wiemels,[6] Xiaomei Ma,[8] Shuki Mizutani,[4] Katsuyoshi Koh,[18] Yukihide Momozawa,[13] Keizo Horibe,[11] Fumihiko Matsuda,[3] Motohiro Kato,[2#] Atsushi Manabe[32#] and Kevin Y. Urayama[1,34#]

[1]Department of Social Medicine, National Center for Child Health and Development, Tokyo, Japan; [2]Department of Pediatrics, the University of Tokyo, Tokyo, Japan; [3]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; [4]Department of Pediatrics and Developmental Biology, Tokyo Medical and Dental University, Tokyo, Japan; [5]Division of Cancer Epidemiology and Prevention, Aichi Cancer Center Research Institute, Nagoya, Japan; [6]Center for Genetic Epidemiology, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; [7]Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA; [8]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CO, USA; [9]Department of Pediatrics, Kyoto Prefectural University of Medicine, Kyoto, Japan; [10]Department of Pediatrics, Graduate School of Medical and Dental Sciences, Kagoshima University, Kagoshima, Japan; [11]Clinical Research Center, National Hospital Organization Nagoya Medical Center, Nagoya, Japan; [12]Children's Cancer Center, National Center for Child Health and Development, Tokyo, Japan; [13]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; [14]Division of Medicinal Safety Science, National Institute of Health Sciences, Kawasaki, Japan; [15]Department of Pediatrics, Aichi Medical University Hospital, Nagoya, Japan; [16]Department of Pediatric Hematology and Oncology Research, National Center for Child Health and Development, Tokyo, Japan; [17]Department of Pediatrics, University of Yamanashi, Yamanashi, Japan; [18]Department of Hematology/Oncology, Saitama Children's Medical Center, Saitama, Japan; [19]Department of Pediatrics, St. Luke's International Hospital, Tokyo, Japan; [20]Department of Child Health, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan; [21]Department of Hematology/Oncology, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; [22]Department of Pediatrics, Japanese Red Cross Narita Hospital, Chiba, Japan; [23]Department of Hematology/Oncology, Chiba Children's Hospital, Chiba, Japan; [24]Department of Pediatrics, Teikyo University Chiba Medical Center, Chiba, Japan; [25]Division of Hematology/Oncology, Kanagawa Children's Medical Center, Yokohama, Japan; [26]Division of Pediatric Hematology and Oncology, Ibaraki Children's Hospital, Mito, Japan; [27]Division of Pediatrics, Showa University Fujigaoka Hospital, Yokohama, Japan; [28]Department of Pediatrics, Shinshu University School of Medicine, Matsumoto, Japan; [29]Department of Pediatrics, Teikyo University Hospital, Tokyo, Japan; [30]Department of Pediatrics, Saitama Medical Center, Saitama Medical University, Saitama, Japan; [31]Department of Pediatrics, National Defense Medical College, Saitama, Japan; [32]Department of Pediatrics, Hokkaido University, Sapporo, Japan; [33]Department of Pediatrics, National Center for Global Health and Medicine, Tokyo, Japan; [34]Graduate School of Public Health, St. Luke's International University, Tokyo, Japan; [35]Department of Statistical Genetics, Graduate School of Medicine, Osaka University, Osaka, Japan; [36]Department of Genome Informatics, Graduate School of Medicine, the University of Tokyo,

Tokyo, Japan; [37]Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; [38]Department of Human Genetics and Disease Diversity, Tokyo Medical Dental University, Tokyo, Japan; [39]Department of Molecular Cytogenetics, Tokyo Medical and Dental University, Tokyo, Japan; [40]Department of Pediatrics, Kyoto University Graduate School of Medicine, Kyoto, Japan; [41]Pediatric Medical Center, Ehime Prefectural Central Hospital, Matsuyama, Japan; [42]Department of Pediatrics, Toho University, Tokyo, Japan and [43]School of Public Health, University of California Berkeley, Berkeley, CA, USA

*MH and TK contributed equally as first authors.
#KYU, AM and MK contributed equally as senior authors.

Correspondence:
K.Y. URAYAMA - kevurayama@gmail.com

https://doi.org/10.3324/haematol.2023.282914

Received: February 14, 2023.
Accepted: October 16, 2023.
Early view: October 26, 2023.



**Disclosures**

No conflicts of interest to disclose.

**Contributions**

MH, TK, MT, KM, TI, YO, OT, JLW, XM, CM, YI, AO, SM, KK, YM, KH, FM, MK, AM, and KYU conceived and designed the study. MH, MT, KM, YT, YA, TH, KO, NK, TI, TI, YO, AMS, TD, YA, MM, DH, DT, HF, YY, YN, YU, SO, HG, MY, DK, KK, DT, YN, KN, KM, YS, DM, SH, YH, YY, HY, MO, JLW, XM, CM, JT, YI, AO, SM, KK, MK, KH, AM, and KYU were involved in patient recruitment and sample and data collection. HM, TK, MT, KM, YA, KO, NK, MK, NM, SK, SJ, CWKC, ATD, AJD, TG, YO, TT, JI, YM, FM, and KYU contributed to the laboratory analyses and assembly of genomic data. HM, TK, SJ, CWKC, ATD, AJD, and KYU conducted the statistical analysis and bioinformatics evaluations. HM, TK and KYU 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.

**Acknowledgments**

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.

**Funding**

This work was support by funding from St. Luke's Life Science Institute (Tokyo, Japan), Japan Society for the Promotion of Science (JSPS) KAKENHI grant number 26253041), Japan Agency for Medical Research and Development (grant numbers 15km0305013h0101, 16km0405107h0004, 21kk0305014), the Children's Cancer Association of Japan, and the Japan Leukemia Research Fund. The HERPACC Study was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, Culture and Technology of Japan Priority Areas of Cancer (17015018), Innovative Areas (221S0001), and JSPS KAKENHI grants (JP16H06277 and 22H04923 [CoBiA], JP26253041, JP20K10463) and Grant-in-Aid for the Third Term Comprehensive 10-year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare of Japan. The work pertaining to the CCRLP replication series was supported by grants from the US National Institutes of Health (R01CA155461, R35GM142783). The biospecimens and/or data used in CCRLP were obtained from the California Biobank Program (SIS request#26), Section 6555(b), 17 CCR. The collection of cancer incidence data used in the CCRLP study was supported by the California Department of Public Health pursuant to California Health and Safety Code Section 103885, Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries, under cooperative agreement 5NU58DP003862-04/DP003862, the National Cancer Institute's Surveillance, Epidemiology and End Results Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the author(s) and do not necessarily reflect the opinions of the State of California, Department of Public Health, the National Institutes of Health, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors. This study used birth data obtained from the State of California Center for Health Statistics and Informatics. The California Department of Public Health is not responsible for the analyses, interpretations, or conclusions drawn by the authors regarding the birth data used in this publication.

**Data-sharing statement**

The datasets generated and analyzed during the current study are not publicly available due to privacy and ethical restrictions, but are available from the corresponding author on reasonable request.

# References

1. Hunger SP, Mullighan CG. Acute lymphoblastic leukemia in children. N Engl J Med. 2015;373(16):1541-1552.
2. Moriyama T, Relling MV, Yang JJ. Inherited genetic variation in childhood acute lymphoblastic leukemia. Blood. 2015;125(26):3988-3995.
3. Shi Y, Du M, Fang Y, et al. Identification of a novel susceptibility locus at 16q23.1 associated with childhood acute lymphoblastic leukemia in Han Chinese. Hum Mol Genet. 2016;25(13):2873-2880.

4. Urayama KY, Takagi M, Kawaguchi T, et al. Regional evaluation of childhood acute lymphoblastic leukemia genetic susceptibility loci among Japanese. Sci Rep. 2018;8(1):789.
5. Kato M, Manabe A. Treatment and biology of pediatric acute lymphoblastic leukemia. Pediatr Int. 2018;60(1):4-12.
6. Inoue M, Tajima K, Takezaki T, 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). Int J Epidemiol. 2003;32(2):257-262.
7. Terao C, Ota M, Iwasaki T, et al. IgG4-related disease in the Japanese population: a genome-wide association study. Lancet Rheumatol. 2019;1(1):e14-e22.
8. Koh K, Kato M, Saito AM, et al. Phase II/III study in children and adolescents with newly diagnosed B-cell precursor acute lymphoblastic leukemia: protocol for a nationwide multicenter trial in Japan. Jpn J Clin Oncol. 2018;48(7):684-691.
9. Marcussen M, Sonderkaer M, Bodker JS, et al. Oral mucosa tissue gene expression profiling before, during, and after radiation therapy for tonsil squamous cell carcinoma. PLoS One. 2018;13(1):e0190709.
10. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
11. Qiao F, Bowie JU. The many faces of SAM. Sci STKE. 2005;2005(286):re7.
12. Wang DM, Miao KR, Fan L, et al. Intermediate prognosis of 6q deletion in chronic lymphocytic leukemia. Leuk Lymphoma. 2011;52(2):230-237.
13. Jeon S, de Smith AJ, Li S, et al. Genome-wide trans-ethnic meta-analysis identifies novel susceptibility loci for childhood acute lymphoblastic leukemia. Leukemia. 2022;36(3):865-868.
14. 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(7):751-757.
15. Studd JB, Vijayakrishnan J, Yang M, Migliorini G, Paulsson K, Houlston RS. Genetic and regulatory mechanism of susceptibility to high-hyperdiploid acute lymphoblastic leukaemia at 10p21.2. Nat Commun. 2017;8:14616.