Age-related differences of genetic susceptibility to patients with acute lymphoblastic leukemia

Qing Hao[1,2,3,*], Minyuan Cao[4,*], Chunlan Zhang[5,*], Dandan Yin[4], Yuelan Wang[4], Yuanxin Ye[4], Shan Zhao[4], Yunfan Yang[5], Ke-Ling Chen[6], Binwu Ying[4], Lanlan Wang[4], Yiguan Zhang[3], Caigang Xu[5], Yiping Zhu[7], Yu Wu[5], Ju Gao[7], Jun-Ning Zhao[3], Yan Zhang[1], Xiaoxi Lu[7]

[1]Department of Thoracic Oncology, Cancer Center, West China Hospital, Sichuan University, Chengdu, China
[2]College of Pharmaceutical Sciences, Southwest Medical University, Luzhou, China
[3]Sichuan Center for Translational Medicine of Traditional Chinese Medicine, Institute of Translational Pharmacology, Sichuan Academy of Chinese Medicine Sciences, Chengdu, China
[4]Department of Laboratory Medicine, West China Hospital, Sichuan University, Chengdu, China
[5]Department of Hematology, West China Hospital, Sichuan University, Chengdu, China
[6]Digestive Surgery Institute, West China Hospital, Sichuan University, Chengdu, China
[7]Department of Hematology/Oncology, West China Second Hospital, Sichuan University, Chengdu, China
*Equal contribution

Correspondence to: Jun-Ning Zhao, Yan Zhang, Xiaoxi Lu; email: zarmy@189.cn, https://orcid.org/0000-0002-8397-5350; zhang.yan@scu.edu.cn, lu_wcums@scu.edu.cn

ABSTRACT

Inherited predispositions to acute lymphoblastic leukemia have been well investigated in pediatric patients, but studies on adults, particularly Chinese patients, are limited. In this study, we conducted a genome-wide association study in 466 all-age Chinese patients with Acute lymphoblastic leukemia (ALL) and 1,466 non-ALL controls to estimate the impact of age on ALL susceptibility in the Chinese population. Among the 17 reported loci, 8 have been validated in pediatric and 1 in adult patients. The strongest association signal was identified at *ARID5B* locus and gradually decreased with age, while the signal at *GATA3* exhibited the opposite trend and significantly impact on adult patients. With genome-wide approaches, germline variants at 2q14.3 rank as the top inherited predisposition to adult patients (e.g., rs73956024, $P = 4.3 \times 10^{-5}$) and separate the genetic risk of pediatric vs. adult patients ($P = 3.6 \times 10^{-6}$), whereas variants at 15q25.3 (e.g., rs11638062) have a similar impact on patients in different age groups (overall $P = 2.9 \times 10^{-7}$). Our analysis highlights the impact of age on genetic susceptibility to ALL in Chinese patients.

INTRODUCTION

Acute lymphoblastic leukemia (ALL) is a deadly malignancy for both children and adults, mainly observed in children aged between 2 and 5 years old [1]. Although over 80% of pediatric patients can experience 5-year survival after conventional chemotherapeutics [2], the survival rate is still low in adult patients [3, 4]. Genomically, higher prevalence of ALL genetic subtypes

with poor prognosis (e.g., BCR-ABL) and lower subtypes with favorable outcomes (e.g., hyperdiploidy) were presented in adults, therefore, it can partially explain the different survival rate between pediatric and adult patients [4, 5]. However, the genetic determinant for age-specific leukemogenesis is still controversial.

On the other hand, germline variants can strongly influence both susceptibility and treatment outcomes

of ALL. Through genome-wide association studies (GWAS), multiple inherited predispositions to ALL in pediatric patients have been identified, including single nucleotide polymorphisms (SNPs) at loci of *ARID5B*, *IKZF1*, *GATA3,* etc. [6–20]. Association of some SNPs are greatly impacted by clinical characteristics (e.g., age, ethnicity, and subtypes) [8, 13–17, 19, 21], which can be validated by a series of replication studies in independent patient cohorts [22–29]. Recently, with the GWASs performed in specific ethnicity and meta-analysis with a large sample size conducted based on GWASs, four novel loci were identified including Hispanic specific locus (i.e., rs2836365 in *ERG*) [13], and hyperdiploid subtype specific locus (e.g., rs210143 in *BAK1*) [17]. Therefore, a total of 17 pediatric ALL susceptibility loci with genome-wide significance have been reported after several years of investigations. However, studies on inherited predispositions to ALL risk for Chinese patients are limited [30], particularly validations for the recently reported loci. Besides, there are relatively fewer studies on adult patients [31], especially on adolescents and young adults [10]. We thus conducted GWAS in all-age Chinese patients with ALL, to evaluate the reported GWAS signals and screen novel genetic variants, which have age-related different effects on ALL.

MATERIALS AND METHODS

Subject and genotyping

Peripheral blood was obtained from 1,466 non-ALL controls, as well as 466 B-linage ALL patients (381 childhood [0-14 yrs] and 85 adult patients [14-68 yrs] who were treated with standard protocol in West China Hospital of West China Second Hospital (e.g., CCGC-ALL2015, registered in http://www.chictr.org.cn/ with ID: ChiCTR-IPR-14005706). Clinical information was obtained from the record system at our hospitals, including gender, age at diagnosis, and molecular subtypes. Fusion-based molecular subtypes were determined by fluorescence in situ hybridization.

A total of 811,852 SNPs were genotyped with Precision Medicine Research Array (ThermoFisher) and filtered based on minor allele frequency, call-rate, Hardy-Weinberg equilibrium, etc., according to previous standard steps [8]. Subsequently, imputation was conducted by well-established methods (i.e., Michigan Imputation Server [32]) with the filtered SNPs. After setting $r^2 = 0.5$ as a cutoff threshold, 9,466,286 SNPs were finally used for subsequent association analysis.

Statistical analysis

Four GWAS approaches were conducted: all patients *vs.* non-ALL control, pediatric patients (< 14yrs) *vs.* non-ALL controls, adult patients *vs.* non-ALL controls, and pediatric patients *vs.* adult patients. For the reported loci, subtype-specific associations were evaluated. For statistical analysis, the association of SNP genotypes with the indicated phenotypes (e.g., ALL susceptibility of all-age patients) were estimated by comparing the genotype frequency between ALL cases and non-ALL controls, or different age groups with logistic regression model after adjusting for gender and the top three principal components. *P* value, odds ratio (OR) and 95% confidence interval (95% CI) was estimated by using PLINK (version 1.90) [33].

RESULTS

To investigate the inherited predispositions to ALL in the Chinese population, GWAS was performed with all imputed SNPs after stringently filtering. A total of 1,466 non-ALL controls and 466 B-linage ALL patients were included in this study with baseline characteristics illustrated in Table 1, including 381 pediatric (< 14 yrs] and 85 adult patients (≥ 14yrs). For childhood ALL (< 14yrs), only one locus (i.e., *ARID5B*) reached genome-wide significance ($P < 5 \times 10^{-8}$) (Figure 1), suggesting no novel strong genetic predisposition to ALL in the Chinese population with current sample size. Subsequently, we retrieved the association results for all 21 SNPs at 17 reported loci from ALL patients and non-ALL controls. In pediatric patients, a total of 7 SNPs at 6 loci were significantly associated with ALL susceptibility regardless of molecular subtypes, including SNPs at *ARID5B*, *IKZF1*, *BMI-PIP4K2A*, *CEBPE*, *CDKN2B-AS1*, and *BAK1* (Table 2). The top signal at *ARID5B* was rs7090445 ($P = 2.7 \times 10^{-14}$, Odds ratio [OR] = 1.96, [95% CI: 1.60-2.24]), which was reported as the potential causal variant for *ARID5B* locus. rs11770117 ($P = 0.03$, OR = 1.27 [1.01-1.62]) at *IKZF1* locus exhibited independent association with ALL susceptibility even after adjusting for the top signal at *IKZF1* locus (i.e., rs11978267, $P = 0.003$, OR = 1.40 [1.12-1.75]) with $P_{adjust} = 0.04$, which is consistent with previous observation in Hispanics [8]. Additionally, rs4266962 ($P = 0.01$, OR = 1.64 [1.11-2.43]) also exhibited significance independent of rs7088318 at *BMI-PIP4K2A* locus. Among three recently identified loci with a large sample size of the Caucasian population, only rs210143 ($P = 0.007$, OR = 1.41 [1.10-1.81]) at *BAK1* locus can be validated in our Chinese patients (Table 2), indicating ethnicity-specific and shared mechanism of leukemogenesis. However, in adult patients, only SNPs at *GATA3* locus (rs3824662) exhibited significant association with ALL susceptibility ($P = 0.0005$, OR = 1.79 [1.29-2.50]) (Table 2), suggesting the different genetic basis of adult from pediatric patients.

Table 1. Clinical characteristics of all-age ALL patients.

Clinical features	Childhood (N=381)	Adult (N=85)
Age at diagnosis (years); median ± SD	4.7± 3.1	32± 12.8
Gender		
Male, N (%)	210 (55.1%)	48 (56.5%)
Female, N (%)	171 (44.9%)	37 (43.5%)
WBC (× 10^9/L); median ± SD	10.2 ± 84.8	20.63 ± 99.0
Molecular subtype		
ETV6-RUNX1	76	1
TCF3-PBX1	18	0
BCR-ABL (Ph$^+$)	16	46
Hyperdiploid*	31	0

*Hyperdiploid subtype was not determined in 161 out of 466 patients, because it is not routinely tested before 2016 in pediatric ALL and never tested in adults.

Since the significance of some GWAS signals is greatly impacted by clinical features, we thus estimated their associations with ALL susceptibility in Chinese patients considering ethnicity, age, and molecular subtype. The best example for ethnic specificity is the causal missense variant (i.e., rs3731249) in *CDKN2A*. Risk allele frequency (RAF) of rs3731249 is absent (0%) in our cohort, which is consistent with that in a public database (i.e., 0% in East Asian *vs.* 3.3% in Caucasians according to gnomAD [34]) (Figure 2A), and thus perfectly explain the racial difference at this locus. The insignificance of rs17481869 at 2q22.3 locus and rs76925697 at 9q21.31 locus can also be explained by ethnicity specific risk allele frequency. In pediatric ALL patients, the novel

Hispanic-specific ALL risk signal at *ERG* locus exhibited marginally significant association with ALL susceptibility in Chinese patients (P = 0.07, OR = 1.18 [0.99-1.41]), and reached statistical significance in patients without common fusion (P = 0.02, OR = 1.29 [1.04-1.61]), which has been validated in previous study [28]. However, rs1121404 (P = 0.34, OR = 1.13 [0.88-1.45]) at *WWOW* identified in Chinese specific GWAS cannot be validated in our cohort. Association analysis was next performed in different genetic subtypes. Although association of rs7088318 at *BMI-PIP4K2A* locus with ALL susceptibility did not reach statistical significance in the whole patient cohort (P = 0.21, OR = 1.11 [0.94-1.31]), risk allele of this SNP was enriched in
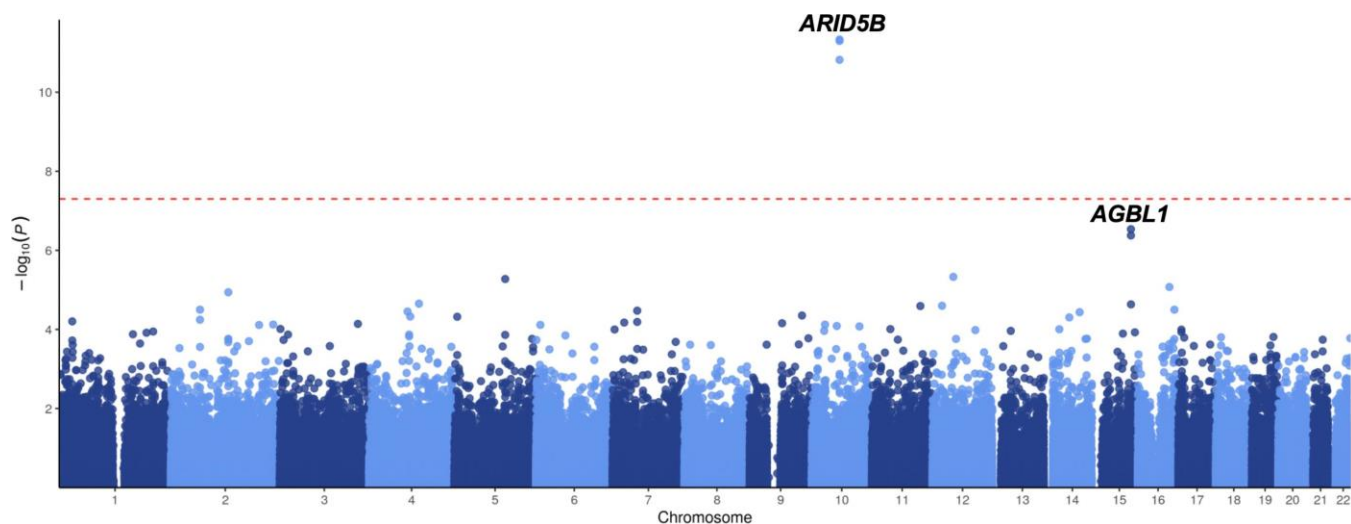

Figure 1. GWAS results of ALL susceptibility in all-age Chinese patients. Association between SNPs and ALL was evaluated in 466 ALL cases and 1,466 non-ALL controls. P value was estimated by logistic regression test and -log₁₀ P (*y*-axis) were plotted against the respectively chromosomal position of each SNP (*x*-axis). Only genotyped but not imputed SNPs were illustrated.

Table 2. Association of the GWAS hits with ALL susceptibility in Chinese patients.

SNP ID	Position*	Genes	Childhood all (N = 381)		ETV6-RUNX1 (N = 77)		no fusion B-ALL (N = 219)		Adult (N = 85)		Age difference
			P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
rs17481869	2:146124454	*ZEB2*	1	NA	1	NA	1	NA	1	NA	1
rs886285	5:131765206	*C5orf56*	0.85	1.02 (0.86-1.21)	0.84	1.04 (0.74-1.45)	0.29	1.12 (0.91-1.37)	0.29	1.20 (0.86-1.66)	0.5
rs210143	6:33546930	***BAK1***	**0.007**	1.41 (1.10-1.81)	0.41	1.22 (0.76-1.94)	**0.017**	1.45 (1.07-1.96)	0.15	1.33 (0.90-1.96)	**0.007**
rs11978267	7:50466304	***IKZF1***	**0.003**	1.40 (1.12-1.75)	0.09	1.45 (0.95-2.21)	**0.006**	1.45 (1.11-1.88)	0.55	1.14 (0.73-1.78)	0.41
rs11770117	7:50473763	***IKZF1***	**0.03**	1.27 (1.01-1.62)	0.39	1.23 (0.77-1.97)	**0.01**	1.51 (1.10-2.05)	0.34	1.25 (0.79-1.96)	0.89
rs28665337	8:130194104	*MYC*	0.27	1.46 (0.74-2.90)	0.43	1.60 (0.50-5.18)	0.35	1.29 (0.76-2.19)	0.54	1.39 (0.49-3.95)	0.32
rs3731249	9:21970916	*CDKN2A*	1	NA	1	NA	1	NA	1	NA	1
rs10811641	9:22014137	***CDKN2B-AS1***	**0.013**	1.24 (1.04-1.47)	**0.024**	1.45 (1.05-2.01)	0.13	1.17 (0.96-1.42)	0.7	1.07 (0.76-1.49)	0.31
rs17756311	9:22053895	*CDKN2A*	0.4541	1.36 (0.61-3.06)	0.42	1.77 (0.45-7.01)	0.34	1.45 (0.68-3.10)	0.37	1.90 (0.47-7.69)	0.76
rs76925697	9:83747371	*TLE1*	1	NA	1	NA	1	NA	1	NA	1
rs3824662	10:8104208	***GATA3***	0.1526	1.14 (0.95-1.35)	0.73	1.06 (0.76-1.49)	**0.032**	1.25 (1.02-1.54)	**0.0005**	1.79 (1.29-2.50)	**0.05**
rs4266962	10:22341574	***BMI-PIP4K2A***	**0.013**	1.64 (1.11-2.43)	0.98	1.01 (0.52-1.95)	**0.013**	1.98 (1.16-3.38)	0.89	1.05 (0.54-2.01)	0.55
rs7088318	10:22852948	***BMI-PIP4K2A***	0.21	1.11 (0.94-1.31)	0.46	1.13 (0.82-1.57)	**0.05**	1.21 (1.00-1.49)	0.12	1.61 (0.88-2.95)	0.8
rs7090445	10:63723577	***ARID5B***	2.7×10^{-14}	1.96 (1.60-2.24)	**0.002**	1.69 (1.21-2.34)	8.2×10^{-13}	2.29 (1.85-2.82)	0.36	1.16 (0.84-1.60)	**0.003**
rs35837782	10:126293309	***LHPP***	0.09	1.17 (0.98-1.40)	0.56	1.10 (0.79-1.55)	**0.044**	1.23 (1.00-1.51)	0.67	1.08 (0.77-1.51)	0.33
rs4762284	12:96612762	*ELK3*	0.72	1.03 (0.87-1.23)	0.43	1.15 (0.81-1.61)	0.35	1.10 (0.90-1.36)	0.71	1.07 (0.75-1.52)	0.61
rs2239630	14:23589349	***CEBPE***	**0.007**	1.29 (1.07-1.54)	**0.016**	1.53 (1.08-2.15)	**0.013**	1.30 (1.06-1.60)	0.21	1.26 (0.88-1.80)	**0.033**
rs1121404	16:79089869	*WWOX*	0.34	1.13 (0.88-1.45)	0.88	1.03 (0.72-1.48)	0.7	1.04 (0.83-1.31)	0.28	1.34 (0.78-2.31)	0.56
rs2290400	17:38066240	*IKZF3*	0.51	1.07 (0.88-1.29)	0.84	1.04 (0.72-1.51)	0.64	1.06 (0.84-1.32)	0.6	1.10 (0.77-1.56)	0.79
rs10853104	17:47092076	*IGF2BP1*	0.53	1.08 (0.84-1.39)	0.5	1.17 (0.74-1.86)	0.96	1.01 (0.75-1.36)	0.81	1.08 (0.59-1.95)	0.3
rs2836365	21:39768274	***ERG***	0.069	1.18 (0.99-1.41)	0.39	1.17 (0.82-1.67)	**0.022**	1.29 (1.04-1.61)	0.94	1.02 (0.70-1.47)	0.88

*Chromosomal locations are based on hg19. Bold genes indicate that SNPs at this locus exhibit significance with ALL susceptibility at least in one association test. P values and ORs were estimated by the logistic regression test, and bold P values indicate P<0.05. OR, odds ration; CI, confidence interval.

B-ALL patients with no common fusions (*P* = 0.05, OR = 1.21 [1.00-1.49]). Similarly, rs35837782 at hyperdiploid-specific *LHPP* locus exhibited marginally significant association with ALL susceptibility in the whole cohort (*P* = 0.09, OR = 1.17 [0.98-1.40]), but

achieved statistical significance in B-ALL patient with no common fusions (*P* = 0.04, OR = 1.23 [1.00-1.51]), which is also observed for rs2836365 in *ERG* (*P* = 0.02, OR = 1.29 [1.04-1.61]). Consistent with the observation in Hispanics [13], an enriched risk allele of rs2836365
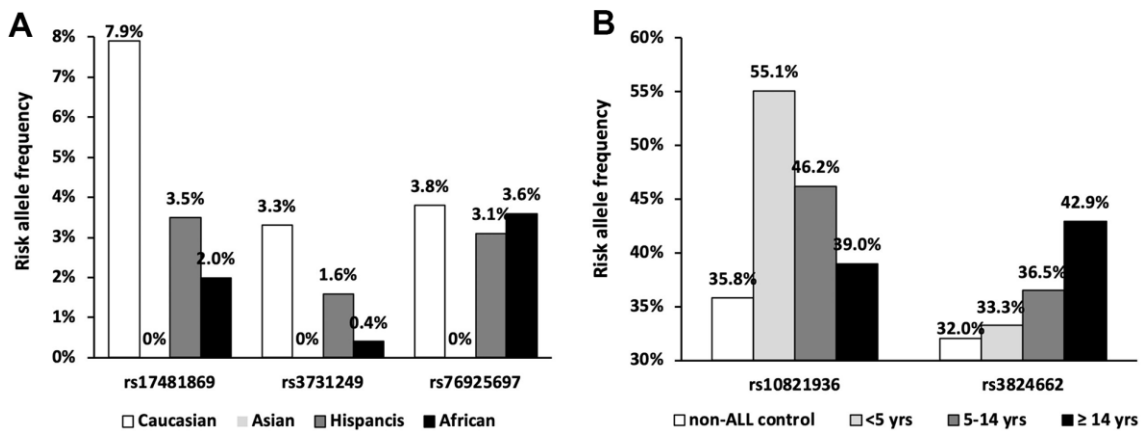

Figure 2: Frequency of the ethnic and age specific loci. (**A**) Risk allele frequencies were illustrated in SNPs with absent variant allele in Chinese population according to gnomAD database; (**B**) Risk allele frequencies of SNPs at *ARID5B* and *GATA3* loci in patients with different age group.

was observed in the *TCF3-PBX1* subtype at rs2836365. In contrast, *ETV6-RUNX1* subtype specific SNPs rs10853104 at *IGF2BP1* locus ($P = 0.53$, OR = 1.08 [0.84-1.39]) at 17q21.32 cannot be validated in our cohort even after considering different subtypes.

Next, we evaluated the impact of age on genetic predisposition. Among all the reported GWAS loci, four were significantly associated with age at diagnosis, namely *ARID5B*, *GATA3*, *BAK1* and *CEBPE* (Table 2). Particularly, all signals lost their association with ALL susceptibility in adults except rs3824662 at *GATA3* locus. As an example, the impact of rs10821936 at *ARID5B* locus gradually decreased from <5 yrs (RAF= 0.55, $P = 2.6 \times 10^{-12}$, OR = 2.27 [1.29-2.09]), 5-14 yrs, (RAF = 0.46, $P = 6.4 \times 10^{-5}$, OR = 1.63 [1.29-2.09]), to ≥14 yrs (RAF = 0.39, $P = 0.24$, OR = 1.23 [0.87-1.72]). Contrastly, rs3824662 at *GATA3* locus exhibited the opposite trend, with increased impact from <5 yrs (RAF = 0.33, $P = 0.43$, OR = 1.09 [0.87-1.36-2.09]), 5-14 yrs (RAF = 0.37, $P = 0.03$, OR = 1.31 [1.03-1.67]), and ≥14 yrs (RAF = 0.43, $P = 0.0005$, OR = 1.79 [1.29-2.50]) Figure 2B. To investigate the impact of age on ALL susceptibility, we compared allele frequencies of each SNP between pediatric and adult patients with GWAS approach by using logistic regression model ($\lambda=1.03$ for the quantile-quantile plot). Although no locus reached genome-wide significance, one novel locus was identified at 2q14.3 with the top signal of rs73956024

($P = 3.6 \times 10^{-6}$) (Figure 3A and Table 3). Additionally, to identify novel ALL susceptibility loci in different age of Chinese patients, we also preformed GWASs in pediatric and adult patients separately. Only *ARID5B* locus reached genome-wide significance in pediatric patients, but none for adult patients, probably because of the small sample size. Interestingly, signals at 2q14.3 locus, which was described above, also ranked the top in adult (e.g., rs73956024, $P = 4.5 \times 10^{-5}$, OR = 2.31 [1.55-3.45]) but not significant in pediatric patients (Figure 3B and Table 3), with RAF of 0.08 in pediatric patients compared with 0.21 and 0.10 in adult patients and non-ALL controls, respectively. Moreover, rs11638062 at 15q25.3 has impact on ALL susceptibility in both pediatric ($P = 4.2 \times 10^{-5}$, OR = 1.67 [1.31-2.14]) and adult patients ($P = 4.7 \times 10^{-6}$, OR = 2.53 [1.70-3.77]), and the overall susceptibility of this locus reached marginally genome-wide significance for all-age patients ($P = 2.9 \times 10^{-7}$, OR = 1.80 [1.44-2.25]) (Figures 1, 3C and Table 3). However, validation in independent patient cohorts is needed due to the small sample size of our study.

DISCUSSION

Most inherited dispositions to ALL have been revealed in Caucasians through genome-wide approaches but limited in Chinese patients. Although subsequent validations have been proceeded for the early identified loci (e.g., *ARID5B*), the impact of the novel loci identified recently
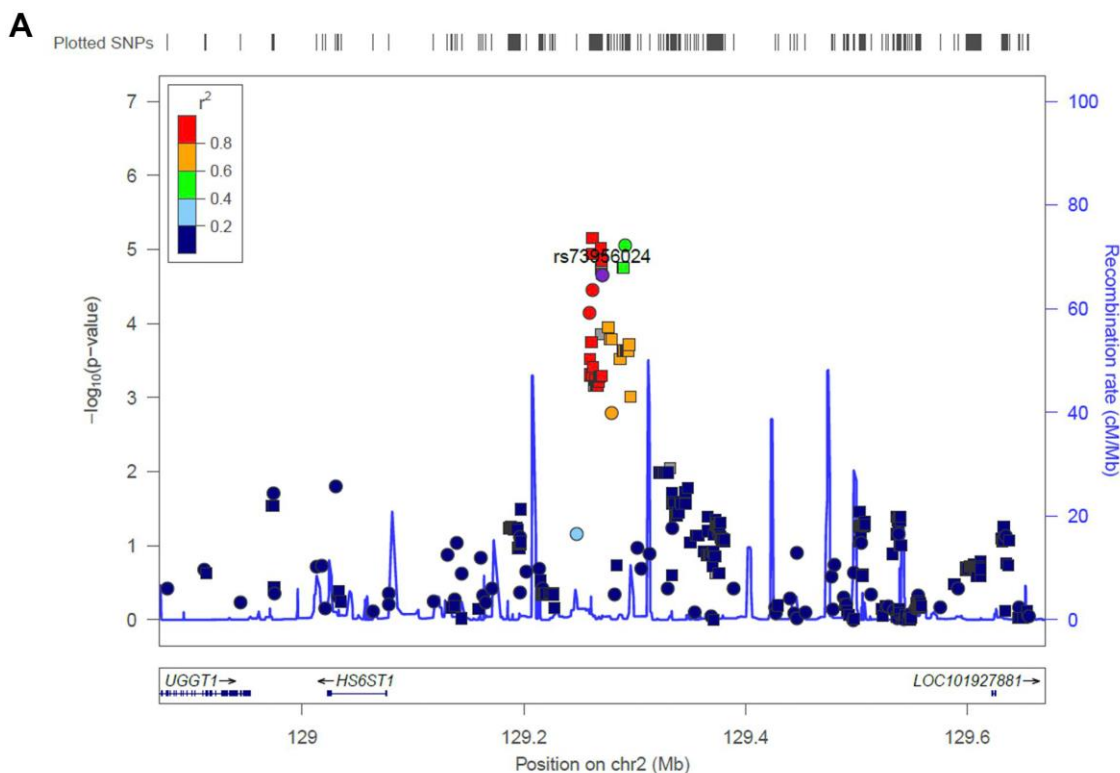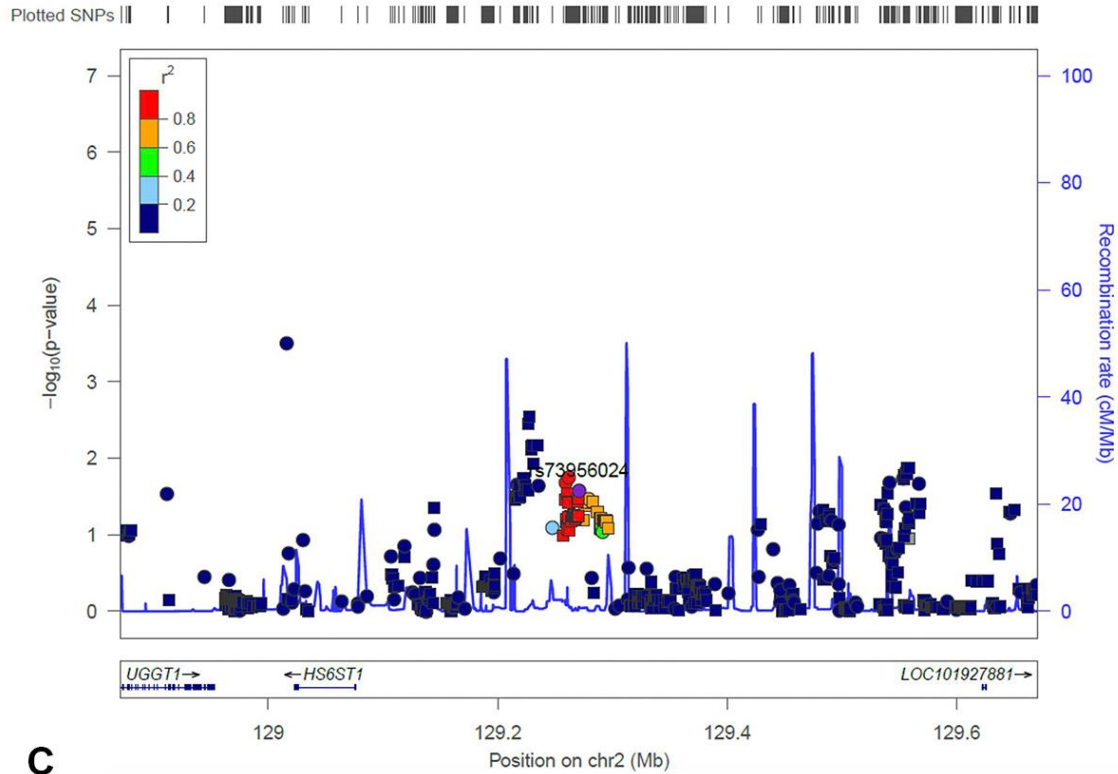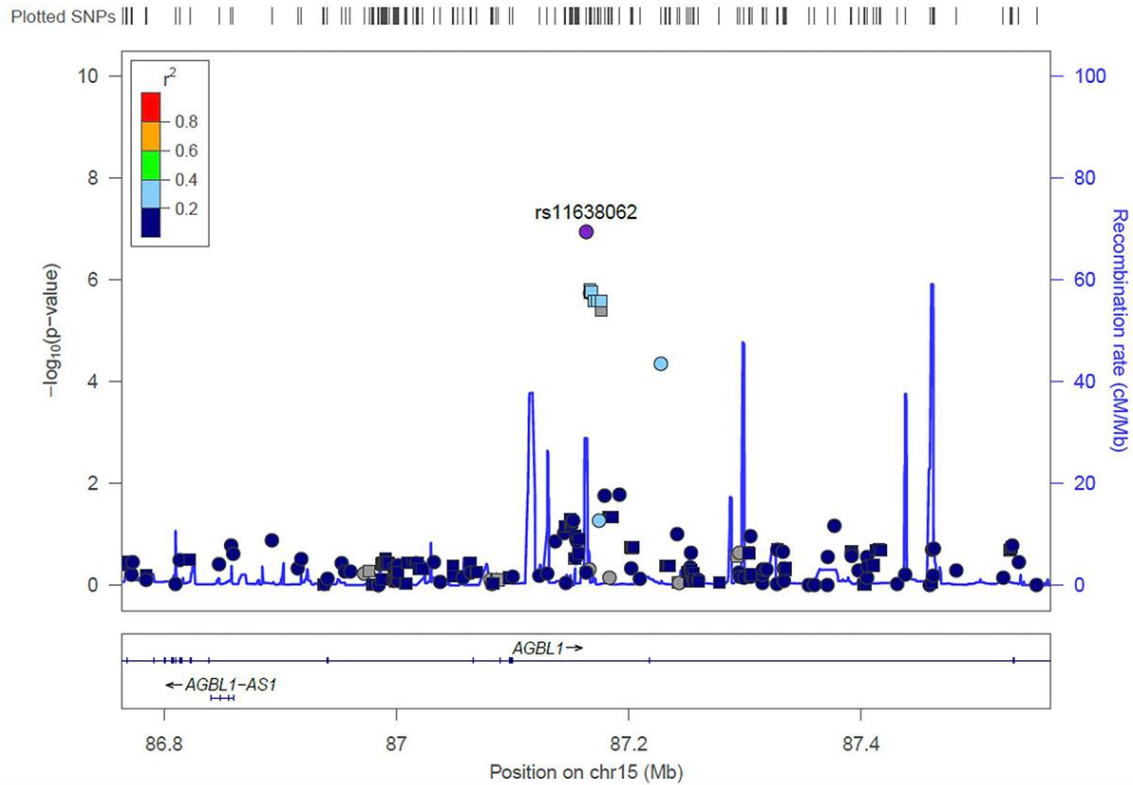
Figure 3. Regional association plot of the novel loci. Association of SNPs at 15q25.3 with ALL susceptibility in adult (**A**) and childhood (**B**). (**C**) Association of SNPs at 2q14.3 with ALL susceptibility in all-age patients. Genotyped and imputed SNPs were labeled in circles and squares, respectively.

Table 3. Association status of the novel loci.

SNP ID	Position*	Genes	Childhood		Adult		All-age patients		Adults *vs.* Children	
			P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)
rs11638062	15:86620033	*AGBL1*	**4.2×10^{-5}**	1.67 (1.31-2.14)	**4.7×10^{-6}**	2.53 (1.70-3.77)	**2.9×10^{-7}**	1.80 (1.44-2.25)	0.06	1.53 (0.98-2.38)
rs16977928	15:86684441	*AGBL1*	**0.0003**	1.80 (1.31-2.48)	**0.023**	1.94 (1.09-3.43)	**2.3×10^{-5}**	1.90 (1.41-2.56)	0.97	1.01 (0.55-1.85)
rs73956024	2:128513332	*HS6ST1*	0.13	0.81 (0.64-1.05)	**4.3×10^{-5}**	2.31 (1.55-3.45)	0.96	1.01 (0.78-1.30)	**3.6×10^{-6}**	3.11 (1.93-5.04)

*Chromosomal locations are based on hg19. Bold genes indicate that SNPs at this locus exhibit significance with ALL susceptibility at least in one association test. P values and ORs were estimated by the logistic regression test, and bold P values indicate *P* < 0.05; OR, odds ration; CI, confidence interval.

with a large sample size has not been evaluated, particularly those ages-, ethnicity- and subtype-specific variants. In this study, we systematically investigated the reported GWAS signals for ALL susceptibly in all-age Chinese patients, and estimated the impact of clinical features, particularly age at diagnosis on ALL susceptibility. A total of 11 SNPs located at 9 out of 17 loci can be validated in Chinese patients in the whole cohort or specific subgroup. The inconsistency is probably due to the racial difference of inherited predispositions to ALL and the small sample size (particularly for some molecular subtypes). Except for the absence of missense variants in *CDKN2A* (i.e., rs3731249), causal variants at 2q22.3 and 9q21.31 loci may be tagged by other variants rather than rs17481869 and rs76925697, risk alleles of which are absent in East Asia. Therefore, we checked the linkage equilibrium (LD) block of these two SNPs in our cohort. No statistically significant signal was identified, suggesting racial specificity of these two loci. For *ETV6-RUNX1* subtype specific locus in *IGF2BP1*, although only 77 patients carried *ETV6-RUNX1* fusion, the risk allele frequency of rs10853104 has no obvious difference between patient and non-ALL controls (12.3% *vs.* 12.4%), suggesting that insignificance of this locus is probably induced by ethnic specificity rather than the small sample size. Moreover, similar trends were also observed for the rest of the insignificant variants, while enlarging sample size may increase the statistical power for evaluating the significance of some subgroup specific loci, such as *ERG* and *LHPP* (*P* = 0.07 and 0.08 in all pediatric patients).

For the GWAS approach, no novel locus has been identified, arguing for a larger sample size to identify potential novel susceptibility locus in Chinese patients in the future. However, we identified a potential novel locus (i.e., 15q25.3) that may have an impact on the susceptibility of all-age patients. Since we do not have an independent replication cohort, we checked the association of this locus in GWAS in previous reports with multi-ethnic populations to validate this signal [8]. Marginally significant was observed for the top signal at this locus (i.e., rs11638062) with *P* = 0.09. Interesting,

another SNP at this locus (i.e., rs16977928 with *P* = 2.3 × 10^{-5} in our all-age GWAS of ALL susceptibility), which is in moderate LD with rs11638062 (r^2 = 0.27, D' = 0.66) in Caucasians, exhibits statistical significance in multi-ethnic population (*P* = 0.007). After considering ethnicity, despite the association trend in Caucasians and blacks, rs16977928 is only significant in Hispanics (*P* = 0.01), who are a mixture between Native American and Caucasians. Since the ancestors of Native Americans are considered to descend from East Asians [35], the causal variant for this locus may exhibit an ethnicity-specific manner in the East Asian population. Moreover, rs11638062 is located in the *AGBL1* gene, polymorphism in which was also associated with lung cancer risk in the Chinese population, suggesting its potential role on tumorigenesis [36]. On the other hand, rs73956024 is located in an enhancer region upstream of *HS6ST1* in B cells according to the public resource [37], and thus could be considered as a possible eQTL to possibly impact the expression level of the adjacent genes, including *HS6ST1*.

In the case of the impact of age on ALL, the difference of risk allele frequencies for the reported GWAS loci between patients and non-ALL controls decreased in adults compared with that in childhood except signals at *GATA3* locus, suggesting the majority of the known GWAS signals are age-specific for pediatric patients. Therefore, we conducted the first GWAS approach to screen ALL susceptibility locus in adult patients. rs3824662 at *GATA3* locus exhibits association at candidate level rather than genome-wide significance. On the other hand, a novel locus at 2q14.3 not only exhibits the most significant association with susceptibility in adult patients but also has the strongest impact on age at diagnosis, suggesting its potential role on age-specific leukemogenesis. Although no known gene located in the LD region of these SNPs, multiple ENCODE candidate cis-regulatory elements were identified, indicating the possible epigenetic effect of the causal variant in this region. Moreover, due to the limited samples and previous research on adult patients, validation for this locus is needed in independent cohorts with a large sample size in the future.

Data availability

The datasets generated for this study can be found in Array Express in the Genome Variation Map (GVM) database (http://bigd.big.ac.cn/gvm) with the accession number of GVM000060.

Ethics statement

This study was approved by the Ethics Committee of West China Second Hospital, Sichuan University, and informed consent was obtained from patients or their guardians, as appropriate.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

REFERENCES

1. Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. Lancet. 2013; 381:1943–55.
https://doi.org/10.1016/S0140-6736(12)62187-4
PMID:23523389

2. Pui CH, Evans WE. Treatment of acute lymphoblastic leukemia. N Engl J Med. 2006; 354:166–78.
https://doi.org/10.1056/NEJMra052603
PMID:16407512

3. Thomas DA, O'Brien S, Faderl S, Garcia-Manero G, Ferrajoli A, Wierda W, Ravandi F, Verstovsek S, Jorgensen JL, Bueso-Ramos C, Andreeff M, Pierce S, Garris R, et al. Chemoimmunotherapy with a modified hyper-CVAD and rituximab regimen improves outcome in de novo Philadelphia chromosome-negative precursor B-lineage acute lymphoblastic leukemia. J Clin Oncol. 2010; 28:3880–89.
https://doi.org/10.1200/JCO.2009.26.9456
PMID:20660823

4. Boissel N, Baruchel A. Acute lymphoblastic leukemia in adolescent and young adults: treat as adults or as children? Blood. 2018; 132:351–61.
https://doi.org/10.1182/blood-2018-02-778530
PMID:29895662

5. Iacobucci I, Mullighan CG. Genetic Basis of Acute Lymphoblastic Leukemia. J Clin Oncol. 2017; 35:975–83.
https://doi.org/10.1200/JCO.2016.70.7836
PMID:28297628

6. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, Kinsey SE, Lightfoot T, Roman E, Irving JA, Allan JM, Tomlinson IP, Taylor M, 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.
https://doi.org/10.1038/ng.430 PMID:19684604

7. Treviño LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, Willman C, Neale G, Downing J, Raimondi SC, Pui CH, Evans WE, Relling MV. Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nat Genet. 2009; 41:1001–05.
https://doi.org/10.1038/ng.432 PMID:19684603

8. Xu H, Yang W, Perez-Andreu V, Devidas M, Fan Y, Cheng C, Pei D, Scheet P, Burchard EG, Eng C, Huntsman S, Torgerson DG, Dean M, 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.
https://doi.org/10.1093/jnci/djt042 PMID:23512250

9. Xu H, Zhang H, Yang W, Yadav R, Morrison AC, Qian M, Devidas M, Liu Y, Perez-Andreu V, Zhao X, Gastier-Foster JM, Lupo PJ, Neale G, et al. Inherited coding variants at the CDKN2A locus influence susceptibility to acute lymphoblastic leukaemia in children. Nat Commun. 2015; 6:7553.
https://doi.org/10.1038/ncomms8553 PMID:26104880

10. Perez-Andreu V, Roberts KG, Xu H, Smith C, Zhang H, Yang W, Harvey RC, Payne-Turner D, Devidas M, Cheng IM, Carroll WL, Heerema NA, Carroll AJ, et al. A genome-wide association study of susceptibility to acute lymphoblastic leukemia in adolescents and young adults. Blood. 2015; 125:680–86.

https://doi.org/10.1182/blood-2014-09-595744
PMID:25468567

11. Vijayakrishnan J, Studd J, Broderick P, Kinnersley B, Holroyd A, Law PJ, Kumar R, Allan JM, Harrison CJ, Moorman AV, Vora A, Roman E, Rachakonda S, et al, and PRACTICAL Consortium. Genome-wide association study identifies susceptibility loci for B-cell childhood acute lymphoblastic leukemia. Nat Commun. 2018; 9:1340.
https://doi.org/10.1038/s41467-018-03178-z
PMID:29632299

12. Wiemels JL, Walsh KM, de Smith AJ, Metayer C, Gonseth S, Hansen HM, Francis SS, Ojha J, Smirnov I, Barcellos L, Xiao X, Morimoto L, McKean-Cowdin R, et al. GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosomes 17q12 and 8q24.21. Nat Commun. 2018; 9:286.
https://doi.org/10.1038/s41467-017-02596-9
PMID:29348612

13. Qian M, Xu H, Perez-Andreu V, Roberts KG, Zhang H, Yang W, Zhang S, Zhao X, Smith C, Devidas M, Gastier-Foster JM, Raetz E, Larsen E, et al. Novel susceptibility variants at the ERG locus for childhood acute lymphoblastic leukemia in Hispanics. Blood. 2019; 133:724–29.
https://doi.org/10.1182/blood-2018-07-862946
PMID:30510082

14. Vijayakrishnan J, Kumar R, Henrion MY, Moorman AV, Rachakonda PS, Hosen I, da Silva Filho MI, Holroyd A, Dobbins SE, Koehler R, Thomsen H, Irving JA, Allan JM, et al. A genome-wide association study identifies risk loci for childhood acute lymphoblastic leukemia at 10q26.13 and 12q23.1. Leukemia. 2017; 31:573–79.
https://doi.org/10.1038/leu.2016.271 PMID:27694927

15. Perez-Andreu V, Xu H, Yang JJ. The novel susceptibility variants for childhood acute lymphoblastic leukemia. J Natl Cancer Inst. 2013; 105:1512–13.
https://doi.org/10.1093/jnci/djt230
PMID:24013180

16. Qian M, Zhao X, Devidas M, Yang W, Gocho Y, Smith C, Gastier-Foster JM, Li Y, Xu H, Zhang S, Jeha S, Zhai X, Sanda T, et al. Genome-Wide Association Study of Susceptibility Loci for T-Cell Acute Lymphoblastic Leukemia in Children. J Natl Cancer Inst. 2019; 111:1350–57.
https://doi.org/10.1093/jnci/djz043 PMID:30938820

17. Vijayakrishnan J, Qian M, Studd JB, Yang W, Kinnersley B, Law PJ, Broderick P, Raetz EA, Allan J, Pui CH, Vora A, Evans WE, Moorman A, et al. Identification of four novel associations for B-cell acute lymphoblastic leukaemia risk. Nat Commun. 2019; 10:5348.
https://doi.org/10.1038/s41467-019-13069-6
PMID:31767839

18. Migliorini G, Fiege B, Hosking FJ, Ma Y, Kumar R, Sherborne AL, da Silva Filho MI, Vijayakrishnan J, Koehler R, Thomsen H, Irving JA, Allan JM, Lightfoot T, et al. Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype. Blood. 2013; 122:3298–307.
https://doi.org/10.1182/blood-2013-03-491316
PMID:23996088

19. de Smith AJ, Walsh KM, Morimoto LM, Francis SS, Hansen HM, Jeon S, Gonseth S, Chen M, Sun H, Luna-Fineman S, Antillón F, Girón V, Kang AY, et al. Heritable variation at the chromosome 21 gene ERG is associated with acute lymphoblastic leukemia risk in children with and without Down syndrome. Leukemia. 2019; 33:2746–51.
https://doi.org/10.1038/s41375-019-0514-9
PMID:31296947

20. Perez-Andreu V, Roberts KG, Harvey RC, Yang W, Cheng C, Pei D, Xu H, Gastier-Foster J, E S, Lim JY, Chen IM, Fan Y, Devidas M, et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. Nat Genet. 2013; 45:1494–98.
https://doi.org/10.1038/ng.2803
PMID:24141364

21. Lee SH, Qian M, Yang W, Diedrich JD, Raetz E, Yang W, Dong Q, Devidas M, Pei D, Yeoh A, Cheng C, Pui CH, Evans WE, et al. Genome-wide association study of susceptibility loci for TCF3-PBX1 acute lymphoblastic leukemia in children. J Natl Cancer Inst. 2020. [Epub ahead of print].
https://doi.org/10.1093/jnci/djaa133 PMID:32882024

22. Xu H, Cheng C, Devidas M, Pei D, Fan Y, Yang W, Neale G, Scheet P, Burchard EG, Torgerson DG, Eng C, Dean M, Antillon F, et al. ARID5B genetic polymorphisms contribute to racial disparities in the incidence and treatment outcome of childhood acute lymphoblastic leukemia. J Clin Oncol. 2012; 30:751–57.
https://doi.org/10.1200/JCO.2011.38.0345
PMID:22291082

23. Yang W, Treviño LR, Yang JJ, Scheet P, Pui CH, Evans WE, Relling MV. ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence. Leukemia. 2010; 24:894–96.
https://doi.org/10.1038/leu.2009.277
PMID:20054350

24. Wang P, Deng Y, Yan X, Zhu J, Yin Y, Shu Y, Bai D, Zhang S, Xu H, Lu X. The Role of ARID5B in Acute Lymphoblastic Leukemia and Beyond. Front Genet. 2020; 11:598.
https://doi.org/10.3389/fgene.2020.00598
PMID:32595701

25. Prasad RB, Hosking FJ, Vijayakrishnan J, Papaemmanuil E, Koehler R, Greaves M, Sheridan E, Gast A, Kinsey SE, Lightfoot T, Roman E, Taylor M, Pritchard-Jones K, 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; 115:1765–67.
https://doi.org/10.1182/blood-2009-09-241513
PMID:20042726

26. Liao F, Yin D, Zhang Y, Hou Q, Zheng Z, Yang L, Shu Y, Xu H, Li Y. Association Between PIP4K2A Polymorphisms and Acute Lymphoblastic Leukemia Susceptibility. Medicine (Baltimore). 2016; 95:e3542.
https://doi.org/10.1097/MD.0000000000003542
PMID:27149463

27. Dai YE, 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. 2014; 9:e113748.
https://doi.org/10.1371/journal.pone.0113748
PMID:25423013

28. Liao F, Ye Y, Yin D, Qin Y, Zhao J, Zhang W, Zhang Y, Deng Z, Wang Y, Ying B, Wang L, Gao J, Shu Y, et al. Validations of Top and Novel Susceptibility Variants in All-Age Chinese Patients With Acute Lymphoblastic Leukemia. Front Genet. 2020; 11:1004.
https://doi.org/10.3389/fgene.2020.01004
PMID:33193587

29. Walsh KM, de Smith AJ, Chokkalingam AP, Metayer C, Roberts W, Barcellos LF, Wiemels JL, Buffler PA. GATA3 risk alleles are associated with ancestral components in Hispanic children with ALL. Blood. 2013; 122:3385–87.
https://doi.org/10.1182/blood-2013-08-524124
PMID:24203929

30. Shi Y, Du M, Fang Y, Tong N, Zhai X, Sheng X, Li Z, Xue Y, Li J, Chu H, Chen J, Song Z, Shen J, et al. Identification of a novel susceptibility locus at 16q23.1 associated with childhood acute lymphoblastic leukemia in Han Chinese. Hum Mol Genet. 2016; 25:2873–80.
https://doi.org/10.1093/hmg/ddw112 PMID:27094129

31. Burmeister T, Bartels G, Gröger D, Trautmann H, Schwartz S, Lenz K, Tietze-Bürger C, Viardot A, Wäsch R, Horst HA, Reinhardt R, Gökbuget N, Hoelzer D, et al. Germline variants in IKZF1, ARID5B, and CEBPE as risk factors for adult-onset acute lymphoblastic leukemia: an analysis from the GMALL study group. Haematologica. 2014; 99:e23–25.
https://doi.org/10.3324/haematol.2013.090720
PMID:24497567

32. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016; 48:1284–87.
https://doi.org/10.1038/ng.3656 PMID:27571263

33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75.
https://doi.org/10.1086/519795 PMID:17701901

34. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, et al, and Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020; 581:434–43.
https://doi.org/10.1038/s41586-020-2308-7
PMID:32461654

35. Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, Terhorst J, Kamm JA, Albrechtsen A, Malaspinas AS, Sikora M, Reuther JD, Irish JD, Malhi RS, et al. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature. 2018; 553:203–07.
https://doi.org/10.1038/nature25173 PMID:29323294

36. Zhang L, Li X, Quan X, Tian W, Yang X, Zhou B. A Case/Control Study: AGBL1 Polymorphism Related to Lung Cancer Risk in Chinese Nonsmoking Females. DNA Cell Biol. 2019; 38:1452–59.
https://doi.org/10.1089/dna.2019.4851
PMID:31603707

37. Zhou X, Maricque B, Xie M, Li D, Sundaram V, Martin EA, Koebbe BC, Nielsen C, Hirst M, Farnham P, Kuhn RM, Zhu J, Smirnov I, et al. The Human Epigenome Browser at Washington University. Nat Methods. 2011; 8:989–90.
https://doi.org/10.1038/nmeth.1772
PMID:22127213