

RESEARCH ARTICLE

Associations between AT-rich Interactive Domain 5B gene Polymorphisms and Risk of Childhood Acute Lymphoblastic Leukemia: a Meta-analysis

Hui Zeng^{1&}, Xue-Bin Wang^{1&}, Ning-Hua Cui², Seungyoon Nam³, Tuo Zeng⁴, Xinghua Long^{1*}

Abstract

Previous genome-wide association studies (GWAS) have implicated several single nucleotide polymorphisms (SNPs) in the AT-rich interactive domain 5B (ARID5B) gene with childhood acute lymphoblastic leukemia (ALL). However, replicated studies reported some inconsistent results in different populations. Using meta-analysis, we here aimed to clarify the nature of the genetic risks contributed by the two polymorphisms (rs10994982, rs7089424) for developing childhood ALL. Through searches of PubMed, EMBASE, and manually searching relevant references, a total of 14 articles with 16 independent studies were included. Odds ratios (ORs) with 95% confidence intervals (95% CI) were calculated to assess the associations. Both SNPs rs10994982 and rs7089424 showed significant associations with childhood ALL risk in all genetic models after Bonferroni correction. Furthermore, subtype analyses of B-lineage ALL provided strong evidence that SNP rs10994982 is highly associated with the risk of developing B-hyperdiploid ALL. These results indicate that SNPs rs10994982 and rs7089424 are indeed significantly associated with increased risk of childhood ALL.

Keywords: ARID5B - SNP - polymorphism - childhood ALL - meta-analysis

Asian Pac J Cancer Prev, 15 (15), 6211-6217

Introduction

Acute lymphoblastic leukemia (ALL), the most common malignancy in children (Linabery and Ross, 2008; Jiang et al., 2013), is a heterogeneous disease, with respect to its underlying cellular and molecular biology (Papaemmanuil et al., 2009), acquired genetic abnormalities, and social clinical responses to combination chemotherapy (Pui et al., 2004). The mechanisms involved in the etiology of childhood ALL remain obscure. It is hypothesized that both environmental and genetic factors contribute to the initiation of leukemogenesis (Greaves, 2006). Previous genome-wide association studies (GWAS) have implicated the associations of several single nucleotide polymorphisms (SNPs) in the AT-rich interactive domain 5B (ARID5B) gene with childhood ALL (Papaemmanuil et al., 2009; Trevino et al., 2009). Importantly, the ARID5B gene showed specificity for B lineage ALL and B-hyperdiploid subtype (Papaemmanuil et al., 2009; Trevino et al., 2009)

ARID5B, the gene encoding the ARID5B protein, which located on chromosome 10q21.2, is one of the common susceptibility loci associated with childhood ALL risk (Papaemmanuil et al., 2009; Trevino et

al., 2009). ARID5B protein plays a vital role in the regulation of embryonic development, cell growth and differentiation through tissue-specific repression of specific gene expression (Huang et al., 1996; Wilsker et al., 2002). Furthermore, some studies have reported that aberrant ARID5B expression in the developing fetus could halt lymphocyte maturation and contribute to leukemogenesis (Healy et al., 2010). Therefore, many studies have evaluated the associations between ARID5B gene polymorphisms and childhood ALL risk; however, the results were often irreproducible, possibly due to the ethnicity-specific genetic profiles, inadequate sample size, etc. Hence, to shed some light on these inconsistencies, we sought to assess the associations of two SNPs (rs10994982, rs7089424) from ARID5B gene with risk of childhood ALL by conducting a comprehensive meta-analysis of individual participant data from all qualified studies.

Materials and Methods

Search strategy and selection criteria

The literature retrieval was performed by two independent reviewers (Hui Zeng, Ning-hua Cui). All

¹Center for Gene Diagnosis, Zhongnan Hospital of Wuhan University, Wuhan, ²Department of Clinical Laboratory, Children's Hospital of Zhengzhou, Zhengzhou, ⁴School of Life Sciences, Guizhou Normal University, Guiyang, China, ³Cancer Genomics Branch, National Cancer Center, Goyang-si, Korea ⁴Equal contributors *For correspondence: xlong888@yahoo.com

studies included in the meta-analysis were selected by searching the PubMed, EMBASE databases up to May 2014 using the following keywords: “ (childhood acute lymphoblastic leukemia or childhood ALL) and (AT-rich interactive domain 5B gene or ARID5B) and (polymorphism or genotype or variant)”. All references in these studies were examined to identify additional research that was not indexed by the databases. Articles that reported results from more than 1 population were considered as separate studies. We selected only published articles written in English.

Selection criteria for articles

The selected studies were required to meet all the following criteria: (1) evaluating the associations between ARID5B SNPs rs10994982, rs7089424 and childhood ALL risk; (2) case-control design; (3) studies with cases younger than 15 years; (4) results with sufficient published data to estimate odds ratios (ORs), confidence interval (CI) and P value. The exclusion criteria were as follows: (1) review articles, abstracts, case reports, reports with incomplete data; (2) studies about adult ALL; (3) studies based on acute myeloid leukemia (AML) and/or Burkitt lymphoma. For duplicate articles, only the most recent or largest data set was selected.

Data extraction

Data were extracted from eligible articles independently by two of the authors (Xue-bin Wang, Tuo Zeng), with any disagreement resolved by consensus. The following information was collected in a predefined data collection form: first author's name, publication year, country, ethnicity, total number of cases and controls, genotyping methods, Hardy-Weinberg equilibrium (HWE) and quality assessment of studies. Quality assessment was performed in each of the acceptable studies in duplicate by independent reviewers using the Newcastle-Ottawa Quality Assessment Scale (Stang, 2010) for all case-control studies.

Statistical analysis

The deviation of HWE for the genotype distribution in controls was assessed using the chi-squared test. The strength of association was expressed as pooled ORs along with the corresponding 95% CIs. In primary, we mainly examined the overall effects for each polymorphism. Firstly, allelic comparison for each polymorphism (A vs G for SNP rs10994982, C vs A for SNP rs7089424) was used to detect overall differences. Secondly, we compared each genotype with additive model (homozygote comparison and heterozygote comparison). Thirdly, recessive model and dominant model were assessed for each polymorphism. We also planned subgroup analyses to determine the effects of ethnicity, ALL subtypes (B-lineage ALL, B-hyperdiploid ALL) on the summary risk estimate. Bonferroni correction (Shi et al., 2012) was used to control for the multiple testing in view of ten comparisons under investigation (significance was set at 0.05/10=0.005).

Heterogeneity between studies was assessed by χ^2 -based Q-tests and I^2 tests, where I^2 (%)>50% or p <0.10

was considered significantly heterogeneous (Higgins et al., 2003). The random-effects model (DerSimonian-Laird)(DerSimonian and Kacker, 2007) was used to assess pooled ORs when significant heterogeneity was observed. Otherwise, the fixed-effects model (Mantel-Haenszel)(Higgins and Thompson, 2002) was used. Further, Galbraith plot analyses were used to visualize the impact of individual studies on the overall homogeneity, which identified the outlier as possible major sources of heterogeneity (Huy et al., 2010). Sensitivity analyses were conducted to evaluate the stability of the combined results by omitting one study at a time or excluding those studies that deviated from HWE (Yang et al., 2014). Publication bias was assessed graphically by funnel plots (Langan et al., 2012) and formally by Egger's test (Egger et al., 1997) and Begg's test (Begg and Mazumdar, 1994). The power of meta-analysis for each polymorphism to detect some effect size was estimated according to the method recommended by Hedges and Pigott (Hedges and Pigott, 2004), given a significant value of 0.05. All statistical analyses were conducted using RevMan 5.2 (The Nordic Cochrane Centre, The Cochrane Collaboration) & STATA 12.0 (Stata, College Station, TX, USA).

Results

Study characteristics

A total of 41 relevant articles were initially identified from PubMed, EMBASE. After titles and abstracts were screened, 14 articles were excluded because of irrelevant or duplicate data. The full texts of the remaining 27 records were carefully reviewed. Among these articles, five articles were about other SNPs (Han et al., 2010; Yang et al., 2010; Wang et al., 2013; Guo et al., 2014) or overlapped data (Xu et al., 2012), two articles were excluded because of adult ALL (Peyrouze et al., 2012) or AML (Rudant et al., 2013), three articles were excluded due to case reports (Paulsson et al., 2010) or review papers (Levine, 2009; Qian et al., 2012)(Figure 1). Another article (Chokkalingam et al., 2013) examined SNPs rs10994982 and rs7089424, but detailed genotype counts could not be obtained despite attempts to contact the authors. Thus, this article was excluded as well. Two of the eligible articles contained data from 4 independent studies and we treated them independently (Papaemmanuil et al., 2009; Prasad et al., 2010). Therefore, 14 articles including 16 studies were included in the present meta-analysis. Here, eight articles (including 3684 patients and 25085 controls) dealt with SNP rs10994982 (Trevino et al., 2009; Healy et al., 2010; Orsi et al., 2012; Gutierrez-Camino et al., 2013; Linabery et al., 2013; Ross et al., 2013; Xu et al., 2013; Emerenciano et al., 2014). Among the included articles, two articles only reported allele contrasts (Orsi et al., 2012; Xu et al., 2012). For SNP rs7089424, nine articles including 11 studies (totally 4378 patients and 7928 controls) were included in the analyses (Papaemmanuil et al., 2009; Healy et al., 2010; Prasad et al., 2010; Vijayakrishnan et al., 2010; Pastorzak et al., 2011; Lautner-Csorba et al., 2012; Orsi et al., 2012; Gutierrez-Camino et al., 2013; Lin et al., 2014), but one article only provided allele contrast (Orsi et al., 2012). For studies only provided allele contrasts, the

deviation of HWE for the genotype distribution in controls was not performed because of insufficient data. The study characteristics and genotype distributions included in the meta-analysis were listed in (Table 1).

For SNP rs10994982, 5 articles provided genotype information of B-lineage ALL cases and other subtype ALL cases (Trevino et al., 2009; Healy et al., 2010; Gutierrez-Camino et al., 2013; Linabery et al., 2013; Emerenciano et al., 2014); 3 articles offered the genotypes of B-hyperdiploid ALL and other B-lineage ALL (Trevino et al., 2009; Gutierrez-Camino et al., 2013; Linabery et al., 2013). So subtype analyses of B-lineage ALL and B-hyperdiploid ALL were performed based on these articles. For SNP rs7089424, 5 articles including 6 studies provided the genotypes of B-lineage ALL cases and other subtype ALL cases (Papaemmanuil et al., 2009; Healy et al., 2010; Vijaykrishnan et al., 2010; Gutierrez-Camino et al., 2013; Lin et al., 2014). So we used these articles for subtype analyses.

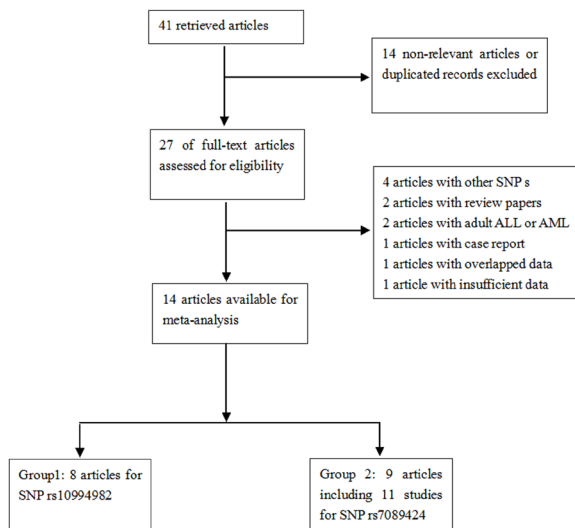


Figure 1. Flow Chart of Article Selection in our Meta-Analysis

Table 1. Characteristics of Studies Included in this Meta-Analysis

First author (year)	Country	Ethnicity	Genotyping method	Sample size (case/control)	Genotype distribution (ALL/Control)				HWE in controls	NOS
SNP rs10994982										
Emerenciano (2014)	Brazil	Mixed	Taqman	161/473	211/540	68/163	75/214	18/96	Yes	8
Ross (2013)	USA, Canada	Caucasian	Taqman	94/383	92/377	25/95	42/187	27/101	Yes	8
Linabery (2013)	USA	Caucasian	Taqman	715/383	810/377	243/95	324/187	148/101	Yes	8
Xu (2013) ^a	USA	Mixed	GWAS	1605/6661	1669/4034	-	-	-	NA	9
Gutierrez-Camino (2013)	Spain	Caucasian	Taqman	195/367	229/366	67/94	95/178	33/95	Yes	7
Orsi (2012) ^a	France	Caucasian	GWAS	441/1542	547/1480	-	-	-	NA	7
Healy (2010)	Canada	Caucasian	GWAS	275/264	325/255	100/63	125/129	50/72	Yes	8
Trevino (2009)	USA	Caucasian	GWAS	437/17916	520/17629	158/4328	204/8973	75/4615	Yes	9
SNP rs7089424										
Lin (2014)	China	Asian	HRM	79/80	71/62	23/18	25/26	31/36	No	6
Gutierrez-Camino (2013)	Spain	Caucasian	Taqman	201/370	200/259	54/52	92/155	55/163	Yes	7
Orsi (2012) ^a	France	Caucasian	GWAS	441/1542	423/1018	-	-	-	NA	7
Lautner-Csorba (2012)	Hungary	Caucasian	Sequenom MassARRAY	543/529	455/357	90/64	275/229	178/236	Yes	9
Pastorzak (2011)	Poland	Caucasian	Allele-Specific PCR	398/731	246/370	46/51	154/268	186/400	Yes	7
Vijaykrishnan (2010)	Thailand	Asian	Allele-specific PCR	190/180	180/184	40/39	100/86	50/55	Yes	7
Prasad (2010) (1)	Germany	Caucasian	Allele-specific PCR	1190/1508	1093/877	249/152	595/673	346/683	Yes	8
Prasad (2010) (2)	UK	Caucasian	Allele-specific PCR	184/356	177/241	39/38	99/165	46/153	Yes	8
Healy (2010)	Canada	Caucasian	GWAS	260/294	265/179	67/40	131/99	62/115	No	8
Papaemmanuil (2009) (1)	UK	Caucasian	GWAS	500/1430	471/965	110/167	251/631	139/632	Yes	8
Papaemmanuil (2009) (2)	UK	Caucasian	GWAS	404/960	365/663	82/123	201/417	121/420	Yes	8

*HRM high-resolution melting; GWAS genome-wide association study; HWE Hardy-Weinberg equilibrium; NA not applicable; NOS the Newcastle-Ottawa Quality Assessment Scale; ^a studies that only reported allele contrast

Meta-analysis about SNP rs10994982

For SNP rs10994982, at the effect size of their respective OR values, the power of meta-analysis for all five comparisons was greater than 0.999. Summaries of the ORs for different comparisons were provided in Table 2 and Figure 2A. In brief, the robust associations between SNP rs10994982 and children ALL were detected using all genetic models even after Bonferroni correction (for A vs G, OR=1.44, 95%CI=1.31-1.58, $p<0.00001$; for AA vs GG, OR=1.98, 95%CI=1.67-2.34, $p<0.00001$; for AG vs GG, OR=1.33, 95%CI=1.14-1.56, $P=0.0007$; for AA vs AG+GG, OR=1.61, 95%CI=1.42-1.82, $p<0.00001$; for AA+AG vs GG, OR=1.55, 95%CI=1.34-1.79, $p<0.00001$). Subgroup analyses restricted to the Caucasian populations were performed. Significant associations were also identified (Table 2).

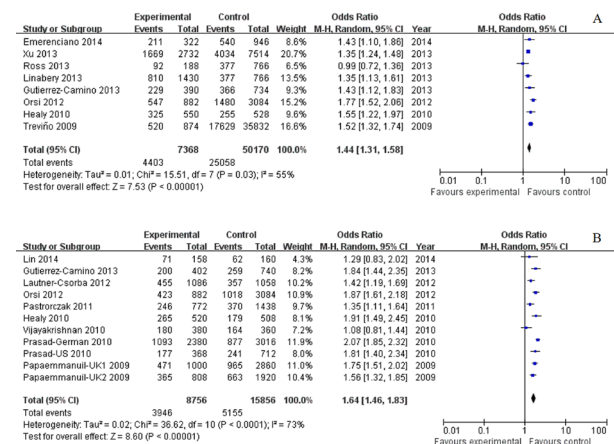


Figure 2. Forest Plots of Pooled OR with 95% CI for Association between Two ARID5B Gene Polymorphisms and Childhood ALL Risk. (A) Forest plots of pooled OR with 95%CI for association between SNP rs10994982 and childhood ALL risk (A vs G, random-effects model); (B) Forest plots of pooled OR with 95%CI for association between SNP rs7089424 and childhood ALL risk (C vs A; random-effects model)

Table 2. Total and Subgroup Analyses of ARID5B gene two Polymorphisms on Childhood ALL

Comparisons	Data	N Studies	Heterogeneity		Model ^a	Effect size OR (95%)	P (Z)	Publication bias	
			I ² (%)	P (Q)				P (Egger)	P (Begg)
SNP rs10994982									
A vs G	Overall	8	55	0.03	R	1.44 (1.31, 1.58)	<0.00001	0.887	0.386
	Caucasian	6	61	0.03	R	1.46 (1.28, 1.66)	<0.00001		
AA vs GG	Overall	6	27	0.23	F	1.98 (1.67, 2.34)	<0.00001	0.386	0.26
	Caucasian	5	40	0.16	F	1.95 (1.64, 2.33)	<0.00001		
AG vs GG	Overall	6	5	0.39	F	1.33 (1.14, 1.56)	0.0003	0.96	0.707
	Caucasian	5	0	0.44	F	1.29 (1.10, 1.52)	0.002		
AA vs AG+GG	Overall	6	0	0.51	F	1.61 (1.42, 1.82)	<0.00001	0.087	0.133
	Caucasian	5	0	0.47	F	1.64 (1.43, 1.87)	<0.00001		
AA+AG vs GG	Overall	6	29	0.22	F	1.55 (1.34, 1.79)	<0.00001	0.788	1
	Caucasian	5	33	0.2	F	1.51 (1.30, 1.76)	<0.00001		
SNP rs7089424									
C vs A	Overall	11	73	<0.0001	R	1.64 (1.46, 1.83)	<0.00001	0.074	0.119
	Caucasian	9	67	0.002	R	1.72 (1.55, 1.90)	<0.00001		
	Asian	2	0	0.5	F	1.13 (0.89, 1.45)	0.36		
CC vs AA	Overall	10	57	0.01	R	2.43 (1.99, 2.97)	<0.00001	0.115	0.21
	Caucasian	8	34	0.16	F	2.72 (2.39, 3.09)	<0.00001		
	Asian	2	0	0.58	F	1.24 (0.78, 1.99)	0.36		
CA vs AA	Overall	10	28	0.19	F	1.67 (1.52, 1.83)	<0.00001	0.733	0.721
	Caucasian	8	30	0.19	F	1.70 (1.55, 1.86)	<0.00001		
	Asian	2	0	0.76	F	1.23 (0.82, 1.83)	0.32		
CC vs CA+AA	Overall	10	45	0.06	R	1.84 (1.57, 2.16)	<0.00001	0.088	0.152
	Caucasian	8	8	0.37	F	2.00 (1.79, 2.25)	<0.00001		
	Asian	2	0	0.39	F	1.09 (0.73, 1.64)	0.67		
CC+CA vs AA	Overall	10	51	0.03	R	1.83 (1.60, 2.09)	<0.00001	0.57	0.592
	Caucasian	8	49	0.06	R	1.90 (1.67, 2.17)	<0.00001		
	Asian	2	0	0.94	F	1.24 (0.86, 1.80)	0.25		

*N number of studies; P (Q) P value of Q test for heterogeneity; P (Z) P value of Z test for significant test; P (Egger) P value of Egger's test for publication bias; P (Begg) P value of Begg's test for publication bias; ^aR: random-effect model; F: fixed-effect model

Table 3. Comparison of SNP rs10994982 and SNP rs7089424 Genotypes between B-lineage ALL Cases and Controls, B-Hyperdiploid ALL Cases and Controls, B-Hyperdiploid ALL Cases and Other B-Lineage ALL Cases in This Meta-analysis

Comparison	B-lineage ALL vs Controls			B-hyperdiploid ALL vs Controls			B-hyperdiploid vs other B-ALL		
	N	OR (95%CI)	P (Z)	N	OR (95%CI)	P (Z)	N	OR (95%CI)	P (Z)
SNP rs10994982									
A vs G	5	1.49 (1.35, 1.64)	<0.00001	3	2.09 (1.73, 2.52)	<0.00001	3	1.34 (1.09, 1.65)	0.006
AA vs GG	5	2.16 (1.78, 2.62)	<0.00001	3	4.14 (2.75, 6.23)	<0.00001	3	1.75 (1.12, 2.71)	0.01
AG vs GG	5	1.38 (1.15, 1.66)	0.0005	3	1.89 (1.26, 2.85)	0.0002	3	1.31 (0.84, 2.04)	0.23
AA vs AG+GG	5	1.71 (1.48, 1.97)	<0.00001	3	2.40 (1.58, 3.66)	<0.00001	3	1.42 (1.07, 1.89)	0.02
AA+AG vs GG	5	1.64 (1.39, 1.95)	<0.00001	3	2.64 (1.79, 3.88)	<0.00001	3	1.50 (0.99, 2.27)	0.05
SNP rs7089424									
C vs A	6	1.64 (1.30, 2.07)	<0.00001	-	-	-	-	-	-
CC vs AA	6	2.31 (1.66, 3.23)	<0.00001	-	-	-	-	-	-
CA vs AA	6	1.78 (1.54, 2.06)	<0.00001	-	-	-	-	-	-
CC vs CA+AA	6	1.71 (1.31, 2.24)	<0.00001	-	-	-	-	-	-
CC+CA vs AA	6	1.97 (1.72, 2.25)	<0.00001	-	-	-	-	-	-

*N number of studies; P (Z) P value of Z test for significant test

In subtype analyses, the differences in all genetic models were highly significant when analyses were performed between B-lineage ALL cases and non-ALL controls (Table 3). So the primary impact of variation defined by SNP rs10994982 was for B-lineage ALL. In subtype analyses of B-lineage ALL, there were higher differences in the genotype distributions of SNP rs10994982 between B-hyperdiploid ALL cases and non-ALL controls in all genetic models (Table 3). When B-hyperdiploid ALL cases and cases with all other B-lineage ALL were compared, statistical differences were observed in allelic comparison, homozygous comparison and the recessive model, although they were all not sufficient robust to withstand the Bonferroni correction (Table 3). In general, the frequency of the risk allele A at

SNP rs10994982 was significantly different in patients with B-hyperdiploid ALL than in all other patients with B-lineage ALL and than in non-ALL controls.

Moderate heterogeneity was observed in the allelic comparison, and the Galbraith plot analyses indicated that the study from Orsi et al. (Orsi et al., 2012) was mainly responsible for the observed heterogeneity (Figure 3A). When we excluded this study, the moderate heterogeneity was significantly decreased and the association was still significantly different (for A vs G, OR=1.39, 95%CI=1.31-1.48, $p<0.00001$, P for heterogeneity=0.30, $I^2=17%$). Sensitivity analyses indicated that the pooled ORs for SNP rs10994982 were consistently significant in childhood ALL risk by omitting one study at a time. We found no evidence of publication bias through both funnel plots

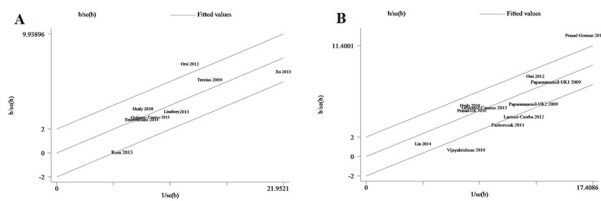


Figure 3. Galbraith plots of association between two ARID5B gene polymorphisms and childhood ALL risk. (A) Galbraith plots of association between SNP rs10994982 and childhood ALL risk (A vs G); (B) Galbraith plots of association between SNP rs7089424 and childhood ALL risk (C vs A)

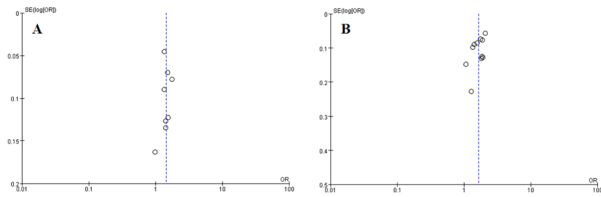


Figure 4. Funnel plots of two ARID5B gene polymorphisms and childhood ALL risk. (A) SNP rs10994982 allelic comparison, A vs G; (B) SNP rs7089424 allelic comparison, C vs A.

(Figure 4A) and Egger's test and Begg's test (Table 2).

Meta-analysis about SNP rs7089424

For SNP rs7089424, at the effect size of their respective OR values, the power of meta-analysis for all five comparisons was greater than 0.999. Details of each comparison were provided in (Table 2 and Figure 2B). In brief, associations with SNP rs7089424 were revealed in all genetic comparisons (for C vs A, OR=1.64, 95%CI=1.46-1.83, $p<0.00001$; for CC vs AA, OR=2.43, 95%CI=1.99-2.97, $p<0.00001$; for CA vs AA, OR=1.67, 95%CI=1.52-1.83, $p<0.00001$; for CC vs CA+AA, OR=1.84, 95%CI=1.57-2.16, $p<0.00001$; for CC+CA vs AA, OR=1.83, 95%CI=1.60-2.09, $p<0.00001$). In the subgroup analyses by ethnicity, significant associations were found among Caucasians but not among Asians in all genetic models (Table 2). In further analyses, the SNP rs7089424 CC genotype frequencies in non-ALL controls showed a statistically significant decreasing trend ($p<0.0001$, Chi-square test, data not shown) from Caucasian subgroup to Asian subgroup, while there was a concomitant increase in SNP rs7089424 AA genotype frequencies. These differences may contribute to racial differences in childhood ALL incidences.

In subtype analyses, all the comparisons revealed significant differences between B-lineage ALL cases and non-ALL controls (Table 3). Subtype analyses of B-lineage ALL were not performed because of limited data.

Moderate heterogeneity was observed in the allelic comparison, homozygote comparison, the recessive and dominant model. Galbraith plot analyses indicated that 3 studies (Prasad et al., 2010; Vijayakrishnan et al., 2010; Pastorczak et al., 2011) were the potential origin of heterogeneity (Figure 3B); the heterogeneity was effectively removed and the associations were still

significant after exclusion of these 3 studies (for C vs A, OR=1.69, 95%CI= 1.56-1.84, P for heterogeneity=0.20, $I^2=28%$; for CC vs AA, OR=2.57, 95%CI= 2.12-3.11, P for heterogeneity=0.23, $I^2=26%$; for CC vs CA+AA, OR=1.87, 95%CI= 1.63-2.16, P for heterogeneity=0.52, $I^2=0%$; for CC+CA vs AA, OR=1.95, 95%CI= 1.74, 2.18, P for heterogeneity=0.30, $I^2=17%$). Sensitivity analyses indicated that no single study qualitatively affected the pooled ORs (including two studies (Healy et al., 2010; Lin et al., 2014) that deviated from HWE), suggesting that the results of our meta-analysis were highly stable. No publication bias was identified for SNP rs7089424 (Table 2 and Figure 4B).

Discussion

According to our meta-analysis, there was a significant association between SNP rs10994982 and childhood ALL risk. In addition, subtype analyses of B-lineage ALL provided evidence that SNP rs10994982 was highly associated with the risk of developing B-hyperdiploid ALL. Although, only 3 studies were performed to assess the association between SNP rs10994982 and B-hyperdiploid ALL risk. However, power analyses in our pooled B-hyperdiploid ALL patients and non-ALL controls indicated there was 91-100% power to detect the observed associations. The mechanism of ARID5B on disease susceptibility remains abstruse. A pivotal piece of evidence is that ARID5B homozygous null mice show transient immune abnormalities including a reduction in early B-cell progenitors (Lahoud et al., 2001; Sherborne and Houlston, 2010). Besides, expression of ARID5B have associated with recombination activating gene 1 (RAG1) expression in bone marrow (Jensen et al., 2010), which supports a role for this gene in early B-cell development.

We have shown that a significant effect of ARID5B SNP rs7089424 on childhood ALL susceptibility. When stratified by ethnicity, SNP rs7089424 presented as a risk factor for childhood ALL in Caucasian populations, but not in Asians. It was notable that the frequency of the CC genotype was higher in the control populations of Asian subgroup than in controls of Caucasian subgroup ($p<0.0001$, Chi-square test, data not shown), while there was a concomitant decrease in SNP rs7089424 AA genotype frequency. An explanation was that the racial differences in childhood ALL incidence could partly be attributed to differences in genotype frequencies between populations of Asian and Caucasian ancestries at ARID5B loci. However, due to the limited number of relevant studies among Asian populations included in this meta-analysis, the pooled results in Asian subgroup may have insufficient statistical power to detect a slight effect or have generated a fluctuated risk estimate. Besides, in Asian subgroup, the genotype distributions in the control population of Lin's study (Lin et al., 2014) was deviated from HWE. Because of the reasons of disequilibrium, the results of genetic association studies might be spurious (Trikalinos et al., 2006). So more well designed studies about racial difference were required in the future.

Moderate heterogeneity was observed in allelic comparison for SNP rs10994982 and in the most

comparisons for SNP rs7089424. Heterogeneity is a potential problem when interpreting the results of a meta-analysis, and finding the sources of heterogeneity is one of the most important goals of meta-analysis (Ioannidis et al., 2007). So first, we attempted to use a random-effects model to adequately capture the trade-off between the association estimates in comparisons with significant heterogeneity. Second, Galbraith plot analyses were performed to identify the outliers which might be the sources of the heterogeneity and we found that all the heterogeneity was effectively eliminated or decreased after excluding the outliers. Third, sensitivity analyses were carried out and removal of each study did not alter the associations with childhood ALL risk. All these procedures suggested the reliability of these results.

Despite our efforts in performing a deeper analysis, some limitations also existed in our meta-analysis. First, this meta-analysis was conducted based on case-control studies, which might encounter recall and selection bias. Second, in subgroup analyses by ethnicity (Asian subgroup), subtype analyses, the number of studies were relatively small, although there was sufficient power ($\geq 80\%$) to detect the observed associations. Third, lack of the original data of available studies limited our further evaluation of potential associations, such as gender-specific associations. Finally, given the fact that ALL subtypes show considerable differences in terms of clinical features, treatment response, etc (Iqbal, 2014), the associations between these two ARID5B polymorphisms and other ALL subtypes (such as T-lineage ALL) should be concerned in future studies.

In conclusion, this meta-analysis confirmed the associations between two ARID5B hot spot polymorphisms and the risk of childhood ALL. In addition, subtype analyses of B-lineage ALL provided evidence that SNP rs10994982 was highly associated with the risk of developing hyperdiploid. We believe that our findings will be useful for future studies in childhood ALL.

Acknowledgements

The Project-sponsored by SRF for ROCS, SEM. This work is supported by National Natural Science Foundation of China (Grants nos. 30873044 and 81272372 to XL), Korean National Cancer Center (NCC) Grant (NCC-1210460 to SN). The authors declare that they have no conflict of interest.

References

- Begg CB, Mazumdar M (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, **50**, 1088-101.
- Chokkalingam AP, Hsu LI, Metayer C, et al (2013). Genetic variants in ARID5B and CEBPE are childhood ALL susceptibility loci in Hispanics. *Cancer Causes Control*, **24**, 1789-95.
- DerSimonian R, Kacker R (2007). Random-effects model for meta-analysis of clinical trials: an update. *Contemp Clin Trials*, **28**, 105-14.
- Egger M, Davey Smith G, Schneider M, Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, **315**, 629-34.
- Emerenciano M, Barbosa TC, Lopes BA, et al (2014). ARID5B polymorphism confers an increased risk to acquire specific MLL rearrangements in early childhood leukemia. *BMC Cancer*, **14**, 127.
- Greaves M (2006). Infection, immune responses and the aetiology of childhood leukaemia. *Nat Rev Cancer*, **6**, 193-203.
- Guo LM, Xi JS, Ma Y, et al (2014). ARID5B gene rs10821936 polymorphism is associated with childhood acute lymphoblastic leukemia: a meta-analysis based on 39, 116 subjects. *Tumour Biol*, **35**, 709-13.
- Gutierrez-Camino A, Lopez-Lopez E, Martin-Guerrero I, et al (2013). Intron 3 of the ARID5B gene: a hot spot for acute lymphoblastic leukemia susceptibility. *J Cancer Res Clin Oncol*, **139**, 1879-86.
- Han S, Lee KM, Park SK, et al (2010). Genome-wide association study of childhood acute lymphoblastic leukemia in Korea. *Leuk Res*, **34**, 1271-4.
- Healy J, Richer C, Bourgey M, Kritikou EA, Sinnett D (2010). Replication analysis confirms the association of ARID5B with childhood B-cell acute lymphoblastic leukemia. *Haematologica*, **95**, 1608-11.
- Hedges LV, Pigott TD (2004). The power of statistical tests for moderators in meta-analysis. *Psychol Methods*, **9**, 426-45.
- Higgins JP, Thompson SG (2002). Quantifying heterogeneity in a meta-analysis. *Stat Med*, **21**, 1539-58.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003). Measuring inconsistency in meta-analyses. *BMJ*, **327**, 557-60.
- Huang TH, Oka T, Asai T, et al (1996). Repression by a differentiation-specific factor of the human cytomegalovirus enhancer. *Nucleic Acids Res*, **24**, 1695-701.
- Huy NT, Thao NT, Diep DT, et al (2010). Cerebrospinal fluid lactate concentration to distinguish bacterial from aseptic meningitis: a systemic review and meta-analysis. *Crit Care*, **14**, 240.
- Ioannidis JP, Patsopoulos NA, Evangelou E (2007). Uncertainty in heterogeneity estimates in meta-analyses. *Bmj*, **335**, 914-6.
- Iqbal Z (2014). Molecular genetic studies on 167 pediatric ALL patients from different areas of Pakistan confirm a low frequency of the favorable prognosis fusion oncogene TEL-AML1 (t 12; 21) in underdeveloped countries of the region. *Asian Pac J Cancer Prev*, **15**, 3541-6.
- Jensen K, Schaffer L, Olstad OK, et al (2010). Striking decrease in the total precursor B-cell compartment during early childhood as evidenced by flow cytometry and gene expression changes. *Pediatr Hematol Oncol*, **27**, 31-45.
- Jiang Y, Hou J, Zhang Q, et al (2013). The MTHFR C677T polymorphism and risk of acute lymphoblastic leukemia: an updated meta-analysis based on 37 case-control studies. *Asian Pac J Cancer Prev*, **14**, 6357-62.
- Lahoud MH, Risteovski S, Venter DJ, et al (2001). Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. *Genome Res*, **11**, 1327-34.
- Langan D, Higgins JP, Gregory W, Sutton AJ (2012). Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis. *J Clin Epidemiol*, **65**, 511-9.
- Lautner-Csorba O, Gezsi A, Semsei AF, et al (2012). Candidate gene association study in pediatric acute lymphoblastic leukemia evaluated by Bayesian network based Bayesian multilevel analysis of relevance. *BMC Med Genomics*, **5**, 42.
- Levine RL (2009). Inherited susceptibility to pediatric acute lymphoblastic leukemia. *Nat Genet*, **41**, 957-8.

- Lin CY, Li MJ, Chang JG, et al (2014). High-resolution melting analyses for genetic variants in ARID5B and IKZF1 with childhood acute lymphoblastic leukemia susceptibility loci in Taiwan. *Blood Cells Mol Dis*, **52**, 140-5.
- Linabery AM, Ross JA (2008). Trends in childhood cancer incidence in the U.S. (1992-2004). *Cancer*, **112**, 416-32.
- Linabery AM, Blommer CN, Spector LG, et al (2013). ARID5B and IKZF1 variants, selected demographic factors, and childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Leuk Res*, **37**, 936-42.
- Orsi L, Rudant J, Bonaventure A, et al (2012). Genetic polymorphisms and childhood acute lymphoblastic leukemia: GWAS of the ESCALE study (SFCE). *Leukemia*, **26**, 2561-4.
- Papaemmanuil E, Hosking FJ, Vijaykrishnan J, et al (2009). Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet*, **41**, 1006-10.
- Pastorczyk A, Gorniak P, Sherborne A, et al (2011). Role of 657del5 NBN mutation and 7p12.2 (IKZF1), 9p21 (CDKN2A), 10q21.2 (ARID5B) and 14q11.2 (CEBPE) variation and risk of childhood ALL in the Polish population. *Leuk Res*, **35**, 1534-6.
- Paulsson K, Forestier E, Lilljebjorn H, et al (2010). Genetic landscape of high hyperdiploid childhood acute lymphoblastic leukemia. *Proc Natl Acad Sci USA*, **107**, 21719-24.
- Peyrouze P, Guihard S, Gardel N, et al (2012). Genetic polymorphisms in ARID5B, CEBPE, IKZF1 and CDKN2A in relation with risk of acute lymphoblastic leukaemia in adults: a Group for Research on Adult Acute Lymphoblastic Leukaemia (GRAALL) study. *Br J Haematol*, **159**, 599-602.
- Prasad RB, Hosking FJ, Vijaykrishnan J, et al (2010). Verification of the susceptibility loci on 7p12.2, 10q21.2, and 14q11.2 in precursor B-cell acute lymphoblastic leukemia of childhood. *Blood*, **115**, 1765-7.
- Pui CH, Relling MV, Downing JR (2004). Acute lymphoblastic leukemia. *N Engl J Med*, **350**, 1535-48.
- Qian XF, Yang GH, Yin CY, Chen X, Shen YF (2012). Research advances on correlation of ARID5B gene with childhood acute lymphoblastic leukemia - review. *Zhongguo Shi Yan Xue Ye Xue Za Zhi*, **20**, 1280-3 (in Chinese).
- Ross JA, Linabery AM, Blommer CN, et al (2013). Genetic variants modify susceptibility to leukemia in infants: a Children's Oncology Group report. *Pediatr Blood Cancer*, **60**, 31-4.
- Rudant J, Orsi L, Bonaventure A, et al (2013). Are ARID5B and IKZF1 polymorphisms also associated with childhood acute myeloblastic leukemia: the ESCALE study (SFCE)? *Leukemia*, **27**, 746-8.
- Sherborne AL, Houlston RS (2010). What are genome-wide association studies telling us about B-cell tumor development? *Oncotarget*, **1**, 367-72.
- Shi Q, Pavvey ES, Carter RE (2012). Bonferroni-based correction factor for multiple, correlated endpoints. *Pharm Stat*, **11**, 300-9.
- Stang A (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol*, **25**, 603-5.
- Trevino LR, Yang W, French D, et al (2009). Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet*, **41**, 1001-5.
- Trikalinos TA, Salanti G, Khoury MJ, Ioannidis JP (2006). Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations. *Am J Epidemiol*, **163**, 300-9.
- Vijaykrishnan J, Sherborne AL, Sawangpanich R, et al (2010). Variation at 7p12.2 and 10q21.2 influences childhood acute lymphoblastic leukemia risk in the Thai population and may contribute to racial differences in leukemia incidence. *Leuk Lymphoma*, **51**, 1870-4.
- Wang Y, Chen J, Li J, et al (2013). Association of three polymorphisms in ARID5B, IKZF1 and CEBPE with the risk of childhood acute lymphoblastic leukemia in a Chinese population. *Gene*, **524**, 203-7.
- Wilsker D, Patsialou A, Dallas PB, Moran E (2002). ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. *Cell Growth Differ*, **13**, 95-106.
- Xu H, Cheng C, Devidas M, et al (2012). ARID5B genetic polymorphisms contribute to racial disparities in the incidence and treatment outcome of childhood acute lymphoblastic leukemia. *J Clin Oncol*, **30**, 751-7.
- Xu H, Yang W, Perez-Andreu V, et al (2013). Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. *J Natl Cancer Inst*, **105**, 733-42.
- Yang W, Trevino LR, Yang JJ, et al (2010). ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence. *Leukemia*, **24**, 894-6.
- Yang YB, Shang YH, Tan YL, et al (2014). Methylenetetrahydrofolate reductase polymorphisms and susceptibility to esophageal cancer in Chinese populations: a meta-analysis. *Asian Pac J Cancer Prev*, **15**, 1345-9.