ASSOCIATIONS BETWEEN AT-RICH INTERACTIVE DOMAIN 5B GENE POLYMORPHISMS AND RISK OF CHILDHOOD ACUTE LYMPHOBLASTIC LEUKEMIA: A META-ANALYSIS

Hui Zeng1&, Xue-Bin Wang1&, Ning-Hua Cui2, Seungyoon Nam3, Tuo Zeng4, Xinghua Long1*

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

RESEARCH ARTICLE

ASSOCIATIONS BETWEEN AT-RICH INTERACTIVE DOMAIN 5B GENE POLYMORPHISMS AND RISK OF CHILDHOOD ACUTE LYMPHOBLASTIC LEUKEMIA: A META-ANALYSIS

Hui Zeng1&, Xue-Bin Wang1&, Ning-Hua Cui2, Seungyoon Nam3, Tuo Zeng4, Xinghua Long1*

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
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 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 $\chi^2$-based Q-tests and F tests, where F (%)>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; Pastorczak 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; Vijayakrishnan et al., 2010; Gutierrez-Camino et al., 2013; Lin et al., 2014). So we used these articles for subtype analyses.

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

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

<table>
<thead>
<tr>
<th>First author (year)</th>
<th>Country</th>
<th>Ethnicity</th>
<th>Genotyping method</th>
<th>Sample size (case/control)</th>
<th>Genotype distribution (ALL/Control)</th>
<th>HWE in controls</th>
<th>NOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP rs10994982</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A allele</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emerenciano (2014)</td>
<td>Brazil</td>
<td>Mixed</td>
<td>Taqman</td>
<td>161/473</td>
<td>211/540</td>
<td>18/96</td>
<td>8</td>
</tr>
<tr>
<td>Ross (2013)</td>
<td>USA, Canada</td>
<td>Caucasian</td>
<td>Taqman</td>
<td>94/383</td>
<td>92/377</td>
<td>25/95</td>
<td>27/101</td>
</tr>
<tr>
<td>Xu (2013)*</td>
<td>USA</td>
<td>Mixed</td>
<td>GWAS</td>
<td>1605/6661</td>
<td>1669/4034</td>
<td>NA</td>
<td>9</td>
</tr>
<tr>
<td>Orsi (2012)*</td>
<td>France</td>
<td>Caucasian</td>
<td>GWAS</td>
<td>441/1542</td>
<td>547/1480</td>
<td>19/96</td>
<td>8</td>
</tr>
<tr>
<td>Trevino (2009)</td>
<td>USA</td>
<td>Caucasian</td>
<td></td>
<td>437/17916</td>
<td>520/17629</td>
<td>50/72</td>
<td>7</td>
</tr>
<tr>
<td>SNP rs7089424</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A allele</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gutierrez-Camino (2013)</td>
<td>Spain</td>
<td>Caucasian</td>
<td>Taqman</td>
<td>201/370</td>
<td>209/259</td>
<td>54/52</td>
<td>92/155</td>
</tr>
<tr>
<td>Orsi (2012)*</td>
<td>France</td>
<td>Caucasian</td>
<td>GWAS</td>
<td>441/1542</td>
<td>423/1018</td>
<td>15/96</td>
<td>9</td>
</tr>
<tr>
<td>Pastorczak (2011)</td>
<td>Poland</td>
<td>Caucasian</td>
<td>Allele-Specific PCR</td>
<td>398/731</td>
<td>246/370</td>
<td>45/51</td>
<td>154/268</td>
</tr>
<tr>
<td>Vijayakrishnan (2010)</td>
<td>Thailand</td>
<td>Asian</td>
<td>Allele-specific PCR</td>
<td>190/180</td>
<td>180/184</td>
<td>40/39</td>
<td>100/86</td>
</tr>
<tr>
<td>Prasad (2010) (1)</td>
<td>Germany</td>
<td>Caucasian</td>
<td>Allele-specific PCR</td>
<td>1190/1508</td>
<td>1093/877</td>
<td>249/152</td>
<td>595/673</td>
</tr>
</tbody>
</table>

*HRM high-resolution melting; GWAS genome-wide association study; HWE Hardy-Weinberg equilibrium; NA not applicable; NOS the Newcastle-Ottawa Quality Assessment Scale; * studies that only reported allele contrast
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 other B-ALL cases than in non-ALL controls (Table 3). When B-hyperdiploid ALL cases and cases with 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.

ties were found among Caucasians but not among Asians (p<0.0001, Chi-square test, data not shown) from relevant studies among Asian populations included in this meta-analysis. 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 been 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 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.
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 (≥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


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.


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


