A common genetic variation in CEBPE and acute lymphoblastic leukemia: a meta-analysis of the available evidence

Xiao-Xia Zhang1,*
Yu-Feng Du2,*
Fan Gao3
Yu-Juan Yang4
Xian-Cang Ma3
Jun Lu3
Jie Zheng3

1Department of Pharmacy, The First Affiliated Hospital, Xi’an Jiaotong University, Xi’an, Shaanxi, People’s Republic of China; 2Department of Urology, The First Affiliated Hospital, Xi’an Jiaotong University, Xi’an, Shaanxi, People’s Republic of China; 3Clinical Research Center, The First Affiliated Hospital, Xi’an Jiaotong University, Xi’an, Shaanxi, People’s Republic of China; 4The Third Department of Cardiology, Shaanxi Provincial People’s Hospital, Xi’an, Shaanxi, People’s Republic of China

*These authors contributed equally to this work

Abstract: Acute lymphoblastic leukemia (ALL) has been studied intensively for decades, but the details of its etiology and underlying mechanisms have yet to be fully elucidated. It is now generally acknowledged that genetic factors contribute greatly to the development of this disease. The gene encoding CCAAT/enhancer-binding protein ε (CEBPE) is involved in the development of leukemia, and in particular the rs2239633 single nucleotide polymorphism (SNP) of CEBPE. The association between rs2239633 and risk of ALL has been well studied, but remains unclear. Therefore, a meta-analysis was performed in this study to establish a more precise estimation of that relationship. A comprehensive literature search of the PubMed electronic database was conducted, and relevant studies published up to February 20, 2015 were selected for analysis. The references of the retrieved articles were also screened. The extracted data were analyzed statistically, and pooled odds ratios with 95% confidence intervals were calculated using Review Manager (version 5.2) to estimate the association strength. Finally, eleven studies were included in the meta-analysis. The pooled analyses revealed that rs2239633 was associated with an increased risk of childhood ALL in Caucasians under any contrast models (P<0.01). However, this SNP did not affect the risk of ALL in adulthood among Caucasians, or in childhood among East Asians. In conclusion, these findings confirm that the CEBPE rs2239633 SNP could be considered a good marker of pediatric ALL risk in Caucasians, but not in East Asians; it is not a good marker of adult ALL risk in Caucasians.

Keywords: CEBPE, rs2239633, ALL, risk, meta-analysis

Background

Acute lymphoblastic leukemia (ALL) is the most common malignancy in childhood, accounting for 30% of all childhood cancers.1 The etiology of childhood ALL is mostly unknown, although infections in the first years of life and some environmental factors, such as ionizing radiation and parental alcohol and tobacco use, are thought to play a causative role.2

ALL is known to result from an accumulation of mutations in tumor suppressor genes and oncogenes, and genetic alterations affecting several chromosomes.2 Thus, the genetic variations may be important factors in its development. This finding has recently led to the performance of several genome-wide association studies to identify common single nucleotide polymorphisms (SNPs) affecting the susceptibility to childhood ALL.2,4 One of the most widely investigated SNPs identified in these studies is the common rs2239633 SNP of CEBPE, the gene encoding CCAAT/enhancer-binding protein ε (CEBPE), which plays an important role in the regulation of myelopoiesis.7 Papaemmanuil et al4 also found that CEBPE is a suppressor of myeloid leukemogenesis, and may be involved in B-cell precursor ALL development. The populations studied in the aforementioned genome-wide association studies were limited, and the many
subsequent association studies that have been conducted with
a view to verifying the described effects of rs2239633 on
susceptibility to ALL in childhood in different populations
have yielded varying results.8–13

A meta-analysis was performed in the present study to
review and compare all relevant studies, enabling more accu-
rate conclusions to be drawn. In addition, since the genetic
background of ALL differs between the adult and childhood
forms, some studies were also performed to make clear if
this SNP was a risk factor for adult ALL.14,15 In this study,
those two relevant studies were also compared to examine
the differential roles of rs2239633 in children and adults.

Materials and methods

Literature search strategy

Two of the authors (Zhang and Du) each performed the lit-
erature search of the PubMed database independently. The
process was supervised by the third author (Zhai), and any
disagreement was resolved by consensus. The search strategy
was to identify all possible studies that involved the use of the
following test words: (14q11.2 “or” CEBPE “or” rs2239633)
“and” (leukemia). The language of publication was restricted
to English. The reference lists of retrieved articles were curated
manually. The literature search was last updated on February
20, 2015. This study was judged exempt regarding ethics.

Selection criteria

All studies with the aim of investigating the association
between presence of the rs2239633 SNP and ALL risk were
potentially included. The following inclusion criteria were
applied:
1. Published articles on human genetics (full texts or
abstracts) without racial restriction.
2. If articles contained more than one geographic or other
clinical characteristic subgroup, each subgroup was
considered separately.
3. If multiple studies were derived from the same population,
only the study with the largest sample was analyzed to
avoid overlapping data.
4. Presentation of sufficient data on the distribution of
rs2239633 in cases and in control groups or the data
necessary to calculate those distributions. If the data were
deficient or presented in an inappropriate form, the origi-
nal authors were contacted and asked for the raw data.

Extracted information

A standard data-collection procedure was used, in line with
the aforementioned inclusion criteria. Two of the authors
(Zhang and Du) independently extracted the variables
from the individual eligible studies in duplicate and made
the characteristics compatible in a pooled database. Any
encountered disagreements were resolved by discussion in
order to reach a consensus. The following information was
extracted from all eligible studies: name of the first author,
year of publication, ethnicities of the study population, age
and number of cases and controls, and main finding. The dis-
tribution of the genotypes among the controls was also tested
for conformity to Hardy–Weinberg equilibrium (HWE).

Statistical analysis

The meta-analysis examined the overall association for allele
contrasts, homozygote contrasts, and recessive and dominant
models. The presence of heterogeneity was assessed using
Cochran’s Q-statistic (with a significance cutoff of P<0.1)
and quantified using the I² statistic, which is proportional to
the degree of heterogeneity; the value of I² statistic lies in
the range 0%–100%, with a value of 0% indicating homogeneity,
and a value above 50% indicating the presence of a very high
degree of heterogeneity.16

The raw data for genotype distribution were used to calcu-
late the study-specific estimates of odds ratio (OR) and 95%
confidence interval (CI). The overall pooled OR and 95% CI
were estimated using the Mantel–Haenszel method, with a
fixed-effects model when no significant clinical or statistical
heterogeneity was present; when substantial heterogeneity
was present, the random-effects model was used. The signifi-
cance of the polled OR was determined by a test. The cutoff
for statistical significance was set at P<0.05.

Potential publication bias was estimated by construct-
ing funnel plots. If most of the data appear at the top of a
funnel plot and are distributed roughly symmetrically, this
suggests the absence of obvious publication bias, and vice
versa.13 There was no need to construct funnel plots when
there were too few analyzed studies (ie, n<5). All of these
statistical analyses were performed using Review Manager
(version 5.2) (Cochrane, London, UK).

The conformity of the rs2239633 SNP to HWE was tested
among the controls in each individual study via the chi-square
test or Fisher’s exact test using SPSS software (SPSS Inc.,
Chicago, IL, USA). Sensitivity analyses were performed by
excluding those studies for which the data deviated from HWE
(P<0.05), omitting the highest weight or changing the model.

Results

Literature search and characteristics of
the retrieved studies

Figure 1 presents a flow diagram summarizing the process
of study selection. The final search generated 65 potentially
relevant publications. After reading the titles and abstracts, 44 articles were found to be irrelevant to this study in that the research did not cover the association between rs2239633 and susceptibility to ALL, and were thus excluded from the analysis. Of the remaining 21 publications, 3 were case reports, 2 were reviews, and 1 was a conference proceedings; these were therefore also excluded. Finally, 15 publications were identified, in which the effect of rs2239633 on the risk of both adult and childhood ALL was evaluated. Of these, four did not supply enough data, two were about adults,14,15 and nine were about children.4,8–13,17,18 These studies, providing enough data, were conducted in various populations of different ethnicities: eight involved Caucasians,4,10,11,13–15,17,18 two involved East Asians,9,12 and one involved multiple races.8 The basic data for every eligible study were extracted and are listed in Table 1. Although the authors of some of

Table 1  Study characteristics

<table>
<thead>
<tr>
<th>Study</th>
<th>Ethnicity</th>
<th>Cases</th>
<th>Controls</th>
<th>Main results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Median age (year)</td>
<td>Number of males</td>
<td>Number of females</td>
</tr>
<tr>
<td>Emerenciano et al8</td>
<td>Brazilian</td>
<td>6.4±4.2</td>
<td>308</td>
<td>235</td>
</tr>
<tr>
<td>Lautner-Csorba et al17</td>
<td>Caucasian</td>
<td>6.5±4.2</td>
<td>322</td>
<td>255</td>
</tr>
<tr>
<td>Orsi et al18</td>
<td>European</td>
<td>&lt;15</td>
<td>252</td>
<td>176</td>
</tr>
<tr>
<td>Papaemmanuil et al4</td>
<td>UK</td>
<td>5.5±3.3</td>
<td>322</td>
<td>255</td>
</tr>
<tr>
<td>Papaemmanuil et al14</td>
<td>UK</td>
<td>5.8±3.8</td>
<td>252</td>
<td>176</td>
</tr>
<tr>
<td>Pastorczak et al11</td>
<td>European</td>
<td>4.9</td>
<td>228</td>
<td>170</td>
</tr>
<tr>
<td>Prasad et al12</td>
<td>German</td>
<td>6</td>
<td>663</td>
<td>530</td>
</tr>
<tr>
<td>Prasad et al12a</td>
<td>UK</td>
<td>6</td>
<td>101</td>
<td>90</td>
</tr>
<tr>
<td>Ross et al10</td>
<td>Caucasian</td>
<td>6</td>
<td>109</td>
<td>81</td>
</tr>
<tr>
<td>Vijayakrishnan et al15</td>
<td>Thai</td>
<td>6</td>
<td>109</td>
<td>81</td>
</tr>
<tr>
<td>Wang et al19</td>
<td>Chinese</td>
<td>1–18</td>
<td>349</td>
<td>219</td>
</tr>
<tr>
<td>Peyrouze et al15</td>
<td>European</td>
<td>16–60</td>
<td>349</td>
<td>219</td>
</tr>
<tr>
<td>Burmeister et al14</td>
<td>German</td>
<td>16–76</td>
<td>349</td>
<td>219</td>
</tr>
<tr>
<td>Enciso-Mora et al14</td>
<td>UK</td>
<td>5.4±3.6</td>
<td>464</td>
<td>360</td>
</tr>
<tr>
<td>Migliorini et al16</td>
<td>UK/German</td>
<td>5.6±3.6</td>
<td>156</td>
<td>144</td>
</tr>
<tr>
<td>Chokkalingam et al12</td>
<td>Hispanic</td>
<td>5.2±3.3</td>
<td>156</td>
<td>141</td>
</tr>
<tr>
<td>Walsh et al13</td>
<td>Hispanic</td>
<td>5.2±3.3</td>
<td>156</td>
<td>141</td>
</tr>
</tbody>
</table>

Note: ‘Indicated another population studied in the same article.
Abbreviations: HWE, Hardy–Weinberg equilibrium; N, no; Y, yes.
the publications were contacted to obtain the raw genotype data, some of the data could not be obtained.

**Meta-analysis**

**rs2239633 SNP and ALL risk in childhood**

As stated, nine papers studied the association between rs2239633 and risk of ALL in childhood and provided data about genotype distribution. The data from these studies were pooled and subjected to a meta-analysis. Overall, this SNP was found to be significantly associated with risk of ALL in childhood under the dominant model (Figure 2A, $P<0.00001$, OR $=1.27$, 95% CI $=1.15–1.39$), recessive model (Figure 2B, $P<0.00001$, OR $=1.23$, 95% CI $=1.14–1.33$), and allele contrast (Figure 2C, $P<0.00001$, OR $=1.18$, 95% CI $=1.12–1.24$). However, the presence of heterogeneity was found in the recessive model (Figure 2B, $I^2 = 56\%$) and allele contrast (Figure 2C, $I^2 = 57\%$), indicating the requirement for a subanalysis. Therefore, a submeta-analysis based on races was conducted. Among the included studies, the subjects in six were Caucasians, and those in two were East Asians, and those in the remaining study were a mixed ethnic population. The submeta-analysis revealed that under the three contrast models, rs2239633 was significantly associated with childhood ALL in Caucasians (Figure 2A, $P<0.00001$, OR $=1.31$, 95% CI $=1.18–1.46$ for dominant model; Figure 2B, $P<0.00001$, OR $=1.32$, 95% CI $=1.21–1.44$ for recessive model; Figure 2C, $P<0.00001$, OR $=1.23$, 95% CI $=1.16–1.30$ for allele contrast), but not in East Asians (Figure 2). The degree of heterogeneity in both the Caucasian and East Asian subgroups was relatively small (Figure 2, $I^2 = 28\%$, 29\%, 36\% for Caucasian under the dominant, recessive, and allele contrast models, respectively; $I^2 = 0$ for East Asian under the dominant, recessive, and allele contrast models).

**rs2239633 and ALL risk in adults**

Two of the studies investigated the relationship between the rs2239633 SNP and risk of ALL in adults, and provided the genotype data. The subjects in both of

---

**Figure 2 (Continued)**
Figure 2 Meta-analysis of the association between rs2239633 and childhood ALL risk.

Notes: (A) Under the dominant model; (B) under the recessive model; (C) under the allele contrast model. *Indicated another population studied in the same article.

Abbreviations: ALL, acute lymphoblastic leukemia; CI, confidence interval; M–H, Mantel–Haenszel method.
those studies were Caucasians. After pooling these data, further analysis revealed that this SNP was not associated with risk of ALL in adults in any model of comparison (Figure 3A, $P=0.18$, OR =1.21, 95% CI =0.91–1.61 for dominant model; Figure 3B, $P=0.79$, OR =1.03, 95% CI =0.81–1.32 for recessive model; Figure 3C, $P=0.35$, OR =1.08, 95% CI =0.92–1.26 for allele contrast). There was no heterogeneity in any of the three contrast models (Figure 3, $I^2=0$).

**Publication bias and sensitivity analysis**

The publication bias for the research with pediatric subjects was assessed. Overall, no obvious publication bias was found under the recessive model, and the funnel plot exhibited good symmetry under the dominant model if the study of Vijayakrishnan et al\(^1\) was excluded. The result did not change after excluding this article. The symmetry of the funnel plot was poor under the allele contrast model, which suggested obvious publication bias (Figure 4).

The stability of the results involving Caucasians was evaluated by performing a sensitivity analysis. Neither omitting Prasad et al’s\(^1\)\(^3\) study (highest weight) nor changing the model (random-effects model) made a significant difference to the pooled effects of the three contrast models, suggesting that the meta-analysis findings were highly stable.

**Discussion**

CEBPE is involved in terminal differentiation and functional maturation of neutrophils and macrophages, and is critical for transcription of most granule proteins localized to specific and gelatinase granules as well as for azurophil granule proteins expressed in the late promyelocyte stage.\(^1\)\(^9\)\(^2\)\(^0\)

It has also been reported that CEBPE is targeted by recurrent immunoglobulin heavy-chain (locating on chromosome band 14q32) chromosomal translocations in B-cell precursor ALL, and mutations in CEBPE lead to neutrophil-specific granule deficiency.\(^9\)\(^1\)\(^9\)\(^2\)\(^0\)

All of these findings suggested that CEBPE plays a role in the development of ALL. The studies

---

**Figure 3** Meta-analysis of the association between rs2239633 and adult ALL risk.

**Notes:** (A) Under the dominant model; (B) under the recessive model; (C) under the allele contrast model.

**Abbreviations:** ALL, acute lymphoblastic leukemia; CI, confidence interval; M–H, Mantel–Haenszel method.
conducted thus far on the relationship between rs2239633 and risk of ALL have yielded a relatively small amount of data and have produced conflicting results. Those data were evaluated in the present meta-analysis. The primary findings of this investigation are the increased OR of childhood ALL susceptibility associated with presence of the rs2239633 SNP in a fixed-effects model, and the lack of an association between this SNP and risk of ALL in adults.

Subgroup analysis for childhood ALL according to race revealed that this SNP impacted Caucasians, and not Asians. This finding was extremely robust, remaining significant irrespective of the different model assumptions (fixed-effects versus random-effects models), genetic models, and additional sensitivity analysis. This disparity between races may reflect the existence of a true population-specific disease variant, but may also be attributable to differences in genomic structure at these loci between populations. Wang et al.9 determined the linkage disequilibrium (LD) patterns at the locus of rs2239633 in European and Chinese populations from the HapMap database, and found that rs2239633 is located in different LD blocks in the two populations. Therefore, we speculate that the pathogenic locus might not be rs2239633, but rather other SNPs in the LD block. Other factors contributing to the differential results for the two populations include differences in allele frequency and differences in genetic and environmental backgrounds that interact with the variants.12

It has been proposed that genetics plays a role in adult cancer susceptibility, although this has been difficult to evaluate because of confounders, such as environmental factors, the immune system, and exposure to pathogens. To the best of our knowledge, no predisposing genetic marker has been proven for adult ALL. No association between rs2239633 and risk of adult ALL was found in this study.

Although the failure to detect any weak association may be affected by the relatively small sample, these data suggest that the genetic impact of polymorphisms at these loci is at least weaker for adult ALL than it is for pediatric ALL and highlights genetic differences between ALL occurring in adults versus children.15

Based on a meta-analysis, Wang et al.13 concluded that rs2239633 confers an increased risk of childhood ALL, especially among Caucasians and Hispanics. The main findings of the present meta-analysis are consistent with that conclusion.
While, the adults’ ALL risk related with rs2239633 was also evaluated, three different genetic models were used in our study, which showed more comprehensive evaluation of the impact of rs2239633 on ALL. In addition, we believe that the article of Wang et al\textsuperscript{21} contained some data errors. The research data – and in particular the OR and 95% CI values – of Chokkalingam et al\textsuperscript{22} and Walsh et al\textsuperscript{23} that were used in the meta-analysis performed by Wang et al\textsuperscript{21} were calculated using logistic regression. We speculate that since the OR and 95% CI values generally differed before and after data adjustment, Wang et al\textsuperscript{21} did not use the raw data from the studies of Chokkalingam et al\textsuperscript{22} and Walsh et al\textsuperscript{23} thus potentially causing bias. It is not possible to derive raw data using the adjusted data. Before completing our study, we attempted to contact the authors of those two articles to obtain the raw genotype data,\textsuperscript{22,23} but were unsuccessful. Therefore, these two articles were excluded from the analysis, which we believe was the most appropriate action.

It is worth noting that there were no differences in the results among the three contrast models, although the funnel figure obtained under the allele contrast model exhibited poor symmetry, and a clear outlier appeared in that obtained under the dominant model, indicating that the impact of genetic models on the results is not large.

Our study demonstrated that genotype testing is very important, although just a statistical association was found. Based on genotype testing, the susceptible population will be able to take the necessary protective measures, which is particularly important for those with long-term exposure to environmental pollutants.

This study had some limitations. First, the effect of gene–gene and gene–environment interactions was not addressed in the meta-analysis due to lack of data. Second, the meta-analysis was based on unadjusted estimates, and the confounding factors could not be controlled for because most studies did not provide these data. Third, for publication bias, the best symmetry was observed for the funnel figure obtained under the allele contrast model exhibited poor symmetry, and a clear outlier appeared in that obtained under the dominant model, indicating that the impact of genetic models on the results is not large.

Conclusion

In summary, the findings of the present meta-analysis support a positive association between \textit{CEBPE} rs2239633 and childhood risk of ALL in Caucasians, but there was no such association for risk of ALL in Caucasian adults or East Asian children. Further large-scale studies assessing gene–gene and gene–environment interactions are required.

Acknowledgment

This work was supported by the Natural Science Foundation of Shaanxi Province (number 2015JM8415) and the Fundamental Research Funds for the Central Universities of China (number 2011jdhz55).

Disclosure

The authors report no conflicts of interest in this work.

References


