

## ORIGINAL ARTICLE

# Identification of functional nucleotide and haplotype variants in the promoter of the *CEBPE* gene

Hyunju Ryoo, Minyoung Kong, Younyoung Kim and Chaeyoung Lee

The current study examined the promoter activity of an association signal in a 5'-upstream region of the gene encoding CCAAT/enhancer binding protein epsilon (*CEBPE*) identified from a recent genome-wide association study (GWAS) for complex acute lymphoblastic leukemia (ALL). This follow-up study first compared the activity of reporter constructs with three haplotypes estimated with the rs2239633 and its proximity nucleotide variants in strong linkage. The most frequent haplotype was CTTTTGT (H1), and the second most frequent haplotype consisted of entirely opposite alleles to H1 (TCGCACC, H2). Their luciferase activity revealed the strongest expression with H2 and the weakest with H1. Subsequent analysis revealed that different luciferase activity was found by the single-nucleotide substitution at rs2239632 and rs2239633 ( $P < 0.05$ ). Especially, the difference in luciferase activity between two alleles of rs2239632 corresponded to the difference between H1 and H2. We concluded that rs2239632 could regulate the expression of the *CEBPE* gene. We suggest a hypothesis that its risk allele (G) might increase the gene product and lead to leukemogenesis. As a result, a person with the allele or the corresponding haplotype might have increased susceptibility to ALL. Further research is warranted to investigate this hypothesis and the underlying mechanisms.

*Journal of Human Genetics* (2013) 58, 600–603; doi:10.1038/jhg.2013.62; published online 30 May 2013

**Keywords:** acute lymphoblastic leukemia; genetic association; linkage disequilibrium; luciferase assay; transcription factor binding site

## INTRODUCTION

Acute lymphoblastic leukemia (ALL) has been known as the most common childhood malignancy.<sup>1</sup> Its genetic factors have been characterized mainly by chromosomal alterations.<sup>2</sup> On the other hand, the genetic factors responsible for increasing susceptibility to complex ALL have been quite limitedly known. Recently, genome-wide association studies (GWAS) have been conducted to identify such factors. They suggested a variety of nucleotide sequence variants associated with ALL.<sup>3–6</sup> Especially, a meaningful GWAS with a large sample size (907 cases and 2398 controls) revealed associations of ALL with variants in and around the genes encoding Ikaros family zinc-finger 1 (*IKZF1*), AT-rich interactive domain 5B (*ARID5B*) and CCAAT/enhancer binding protein epsilon (*CEBPE*).<sup>5</sup> There has been great concern regarding *CEBPE* because it was targeted by translocation of immunoglobulin heavy chain in ALL, resulting in lymphoid leukemogenesis.<sup>7</sup> Its promoter single-nucleotide polymorphism (SNP, rs2239633) was found to be associated with ALL in the GWAS. This association was confirmed by another GWAS<sup>8</sup> and a replication study.<sup>9</sup> Functions of the SNP were, however, not proven. It is even unclear whether the variant is functional or linked to the functional variant. The objective of this study was to identify functional nucleotide variants and/or their haplotypes located at the previous GWAS signal in a promoter region of the *CEBPE* gene.

## MATERIALS AND METHODS

### Selection of variants

We examined the promoter activity of rs2239633 and its proximity SNPs in a 5'-upstream regulatory region of the *CEBPE* gene. The SNPs were selected from a promoter region within the range of -2 to 0 kb based on linkage disequilibrium (LD) mapping using genotypes of 381 Europeans (CEU + TSI + FIN + GBR + IBS) from the 1000 Genomes Project (<http://browser.1000genomes.org/index.html>). The LD block was constructed with SNPs with Hardy-Weinberg equilibrium ( $P > 0.001$ ), minor allele frequency  $> 0.001$  and  $r^2 > 0.8$  by using the confidence interval method of Gabriel *et al.*<sup>10</sup> Their haplotypes were estimated based on pairwise measures of LD using Haploview (version 4.2),<sup>11</sup> and the corresponding frequencies were estimated by expectation-maximization algorithm.

### Promoter reporter constructs

We made a construct of the *CEBPE* promoter region encompassing all the selected polymorphisms linked to the rs2239633. The promoter region was isolated from human genomic DNA using PCR, and the PCR products were ligated into the T-blunt vector (Solgent, Daejeon, Korea) using T4 DNA ligase (Promega, Madison, WI, USA). The fragment containing the *CEBPE* promoter was released from the T-blunt vector by cutting with *KpnI* and *HindIII* restriction enzymes (NEB, Ipswich, MA, UK), gel-purified, and subcloned into promoterless pGL3-basic vector (Promega).

### Site-directed mutagenesis

Each haplotype construct was introduced by mutations of the target SNP sites using QuikChange XL Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA,

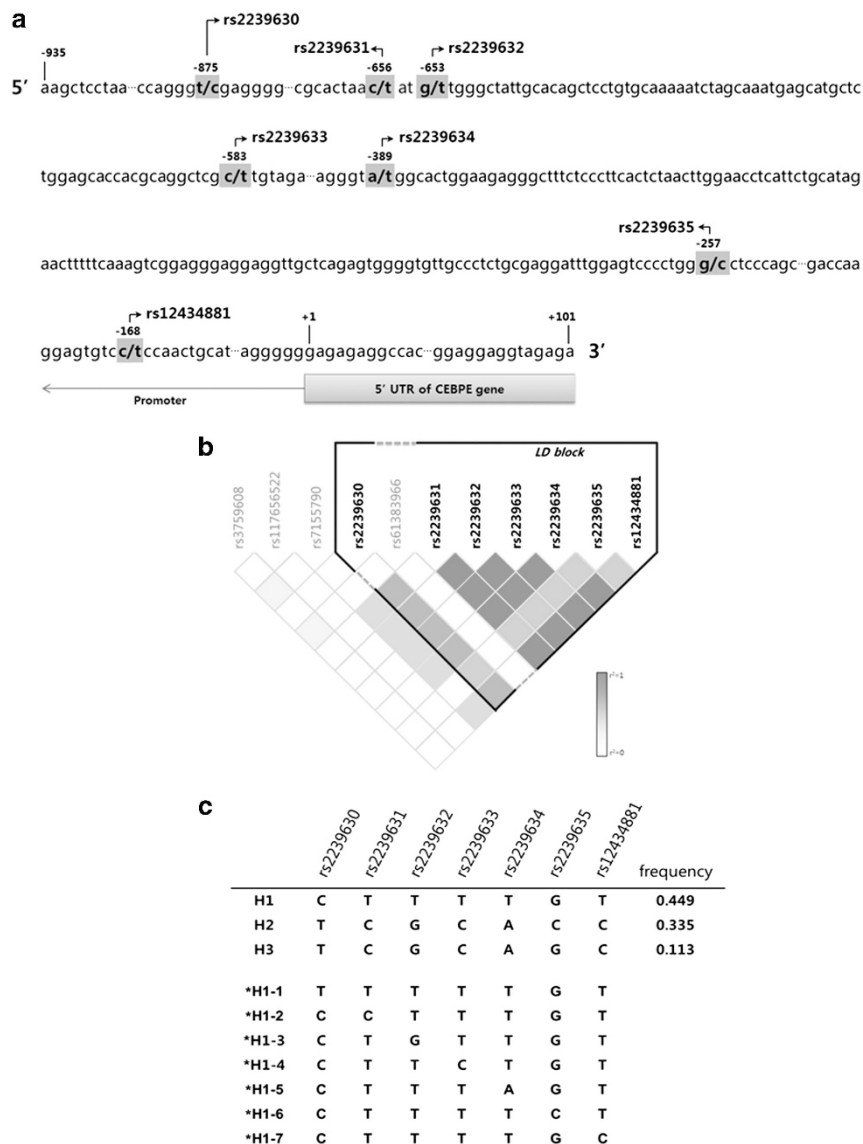
USA). The following forward primers were utilized for point mutations: 5'-TCAGGCCAGGGCGAGGGGCGAAGC-3' (rs2239630), 5'-GGCTGGA AACGCACTAATATTGGGCTATTGCACA-3' (rs2239631), 5'-CTGGAAACG CACTAACATTTGGGCTATTGCACAGCTC-3' (rs2239632), 5'-CACACG CAGGCTCGTGTGTAGAGCTTGTTC-3' (rs2239633), 5'-ACCCAAGTGGTGGTGGCACTGGAAGAGG-3' (rs2239634), 5'-TGGAGTCCCCTGGCTCC CAGCAGGGA-3' (rs2239635), 5'-CTGACCAAGGAGTGTCTCCAAGTCTGTAATAGG-3' (rs12434881). The corresponding reverse primers were as follow: 5'-GCTTCGCCCCTCGCCCTGGGCCTGA-3' (rs2239630), 5'-TGTGCAATAGCCACATATTAGTGCCTTTCCAGCC-3' (rs2239631), 5'-GCTGTGCAATAGCCCAATGTTAGTGCCTTTCCAG-3' (rs2239632), 5'-GAA CAAGCTCTACACAACGAGCCTGCGTGTG-3' (rs2239633), 5'-CCTCTTC CAGTGCCAACCCTCAGTTGGGT-3' (rs2239634), 5'-TCCCTGCTGGGA GGCCAGGGACTCCA-3' (rs2239635), 5'-CCTATTACAGCAGTTGGAGACA CTCCTGGTCAG-3' (rs12434881). All constructs were confirmed by direct sequencing.

### Cell lines and culture conditions

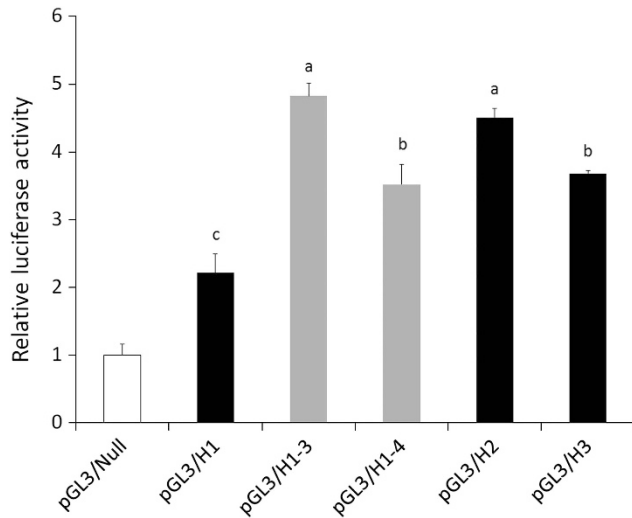
Human embryonic kidney (HEK) 293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM; Wellgene, Daegu, Korea), supplemented with 1% penicillin streptomycin (Wellgene) and 10% fetal bovine serum (FBS; Wellgene) at 37 °C in a humidified atmosphere containing 5% CO<sub>2</sub>.

### Dual-luciferase reporter assay

Twenty-four hours prior to transfection, HEK293 cells were seeded at  $2.4 \times 10^5$  cells per well on six-well plates. The cells in each well were transfected with 1200 ng of the pGL3-basic vector or the pGL3-*CEBPE* haplotype vectors containing firefly luciferase reporter gene using Lipofectamine 2000 (Invitrogen, Carlsbad, CA, USA). Ten nanograms of pRL-CMV vector (Promega) encoding Renilla luciferase reporter was used as an internal control for transient transfection efficiency. Twenty-four hours after transfection, the cells were washed with PBS, and the firefly and Renilla luciferase activities were



**Figure 1** Single-nucleotide variants and their haplotypes used in the current study. (a) Nucleotide sequence of the *CEBPE* promoter region used in this study. Shaded letters indicate the selected SNPs with major/minor alleles. (b) Linkage disequilibrium (LD) block constructed in the promoter region. Color scheme represents the estimates of squared correlation coefficient ( $r^2$ ). (c) Haplotypes and their frequencies. \* indicates nonnatural haplotype. A full color version of this figure is available at the *Journal of Human Genetics* journal online.



**Figure 2** Estimates of relative luciferase activity in HEK293 cell by haplotypes of the *CEBPE* promoter region. The estimate represents the average of three values resulted from experiments in triplicates, and the error bar represents its corresponding standard error. Similar results were obtained in two other independent experiments. Relative luciferase activity estimates are presented with white, gray, and black bars for only pGL3, nonnatural haplotypes and natural haplotypes, respectively. The estimates without the same letter differ by the Tukey's multiple test ( $P < 0.05$ ).

measured with luminometer (CentroPRO LB 962; Berthold, ND, USA) using the Dual-Luciferase Report Assay (Promega). Three replicates per *CEBPE* haplotype-reporter construct were examined in each transfection experiment. The reporter assay was performed in triplicate, and thus a total of nine replicates per haplotype-reporter construct was assayed.

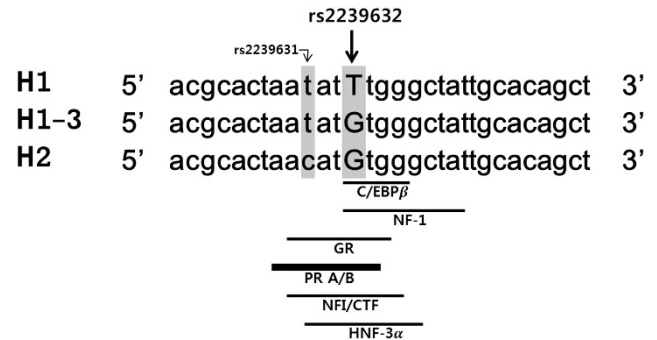
#### Statistical analysis

Differences in relative luciferase activity among haplotypes were determined with the significance threshold of  $P$ -value  $< 0.05$  by the Tukey's multiple test. The statistical analysis was conducted using SPSS software version 20 (Chicago, IL, USA).

#### RESULTS

The LD block was estimated at the previous GWAS signal in a promoter region of the *CEBPE* gene (Figure 1). The block included 7 of 16 SNPs in its upstream region of  $-2$  to  $0$  kb, and the seven SNPs were selected to examine their promoter activity. Haplotypes with the selected SNPs were estimated with their frequencies, and three haplotypes were obtained with their frequencies  $> 0.1$  (Figure 1). The most frequent haplotype (H1) was found to be CTTTTGT, with a frequency of 0.45. The second most frequent haplotype (H2, TCGCACC) had entirely opposite alleles to H1. A reporter construct of 1036 bp *CEBPE* belonging to a promoter region from  $-935$  to  $+100$  bp encompassing all the polymorphisms (rs12434881, rs2239630, rs2239631, rs2239632, rs2239633, rs2239634 and rs2239635) linked to the rs2239633 was made for each haplotype. Their luciferase activity revealed the strongest expression with H2 and the weakest with H1 (Figure 2).

Subsequently, luciferase activity was examined further to identify a specific functional nucleotide variant, which is a component of the haplotypes. We did not find any difference in luciferase activity by substituting the allele of rs12434881, rs2239630, rs2239631, rs2239634 or rs2239635 ( $P > 0.05$ , Supplementary Figures S1 and S2). However, the luciferase activity increased by allelic substitution at rs2239632



**Figure 3** Transcription factor binding sites predicted by haplotypes of the *CEBPE* promoter region. The thin horizontal line indicates the transcription factor binding site of H1, and the bold line indicates that of H2. They are displayed in decreasing order of similarity from top. C/EBPβ, CCAAT/enhancer binding protein β; NF-1, neurofibromatosis type 1; GR, glucocorticoid receptor; PR, progesterone receptor; NFI-CTF, nuclear factor 1—CCAAT-binding transcription factor; HNF-3α, hepatocyte nuclear factor-3α.

and at rs2239633 ( $P < 0.05$ , Figure 2). Especially, a distinctive difference in luciferase activity was observed by the nonnatural haplotype with nucleotide substitution (T to G) of rs2239632, which is 653 bp ahead of the gene. The activity level with its G allele corresponded to that of H2 ( $P > 0.05$ ).

#### DISCUSSION

The current study identified a functional sequence variant in the promoter of the *CEBPE* gene by luciferase reporter assay. This study was performed based on the results of a GWAS signal (rs2239633) in the 5'-upstream region of the *CEBPE* gene for susceptibility to ALL in English (odds ratio (OR) = 1.34,  $P = 2.88 \times 10^{-7}$ , 907 cases and 2398 controls),<sup>5</sup> Germans (OR = 1.6,  $P = 5.1 \times 10^{-5}$ , 1193 cases and 1516 controls),<sup>8</sup> and Germans and Italians (OR = 1.35,  $P = 4.0 \times 10^{-10}$ , 1438 cases and 2735 controls).<sup>9</sup> The functional examination of its sequence variants was preferred because CEBPE has a crucial role in dysregulation of CEBP, which is a suppressor of leukemogenesis.<sup>5</sup> In fact, a variety of rearrangements in the *CEBPE* gene were found in tumors derived from patients with T-cell leukemia or lymphoma.<sup>12-16</sup> Also, genetic associations of ALL with other nucleotide variants (rs4982729, OR = 1.33,  $P = 2.88 \times 10^{-10}$ ; rs12887958, OR = 1.28,  $P = 3.16 \times 10^{-8}$ ; rs4982731, OR = 1.36,  $P = 8.97 \times 10^{-12}$ ; rs8015478, OR = 1.32,  $P = 4.35 \times 10^{-10}$ ; rs17794251, OR = 1.22,  $P = 2.7 \times 10^{-5}$ ) of the gene were identified from a more recent GWAS with multiple ethnic populations (1605 cases and 6661 controls).<sup>17</sup>

We obtained a heterogeneous gene expression by haplotypes composed of the variants linked to rs2239633 located  $-653$  bases apart from its transcription start site, found to be associated with ALL in the previous study. The most frequent haplotype (CTTTTGT), with a frequency of 0.449, was found to decrease its expression compared with the haplotype (TCGCACC, frequency = 0.335) with entirely opposite alleles to the most frequent haplotype. Further investigation unveiled causal single-nucleotide variants (rs2239632 and rs2239633) for the distinctive expression by a nucleotide substitution of the most frequent haplotype. Especially, the difference in the expression between H1 and H2 haplotypes corresponded to that between the two alleles of rs2239632. Its G allele turned out to increase the expression. The increased CEBPE might be targeted by translocation of immunoglobulin heavy chain resulting in lymphoid leukemogenesis,<sup>7</sup> and accumulated lymphoid leukemogenesis would finally lead to ALL. This hypothesis concurred with the GWAS in

which the *C* allele of rs2239633 was found to be associated with increased susceptibility to ALL. The *C* allele was in complete LD with the *G* allele of rs2239632. Further studies are needed to confirm the hypothesis and to unravel detailed mechanisms. For example, this study predicted changes in transcription factors and their binding sites by nucleotide substitution of the functional variant using PROMO program (<http://alggen.lsi.upc.es>), and identified various binding sites that disappeared by the nucleotide substitution from *T* to *G* (Figure 3). Thus, we suspect that the *G* allele at rs2239632 decreases the binding affinity with proteins. A study that examines the association between susceptibility to ALL and the haplotypes analyzed in the current study is also warranted. Rigorous efforts to examine the changes in the transcription factors and further mechanisms toward phenotypic changes would contribute to our understanding of the genetic architecture for susceptibility to ALL.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

We would like to thank anonymous reviewers for their valuable comments. This study was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (Grant No. 2012002096).

- 1 Pui, C. H., Robison, L. L. & Look, A. T. Acute lymphoblastic leukaemia. *Lancet* **371**, 1030–1043 (2008).
- 2 Pui, C. H., Relling, M. V. & Downing, J. R. Acute lymphoblastic leukemia. *N. Engl. J. Med.* **350**, 1535–1548 (2004).
- 3 Healy, J., Bélanger, H., Beaulieu, P., Larivière, M., Labuda, D. & Sinnett, D. Promoter SNPs in G1/S checkpoint regulators and their impact on the susceptibility to childhood leukemia. *Blood* **109**, 683–692 (2007).
- 4 de Jonge, R., Tissing, W. J., Hooijberg, J. H., Jansen, G., Kaspers, G. J., Lindemans, J. *et al.* Polymorphisms in folate-related genes and risk of pediatric acute lymphoblastic leukemia. *Blood* **113**, 2284–2289 (2009).
- 5 Papaemmanuil, E., Hosking, F. J., Vijayakrishnan, J., Price, A., Olver, B., Sheridan, E. *et al.* Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat. Genet.* **41**, 1006–1010 (2009).
- 6 Treviño, L. R., Yang, W., French, D., Hunger, S. P., Carroll, W. L., Devidas, M. *et al.* Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat. Genet.* **41**, 1001–1005 (2009).
- 7 Akasaka, T., Balasas, T., Russell, L. J., Sugimoto, K. J., Majid, A., Walewska, R. *et al.* Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). *Blood* **109**, 3451–3461 (2007).
- 8 Prasad, R. B., Hosking, F. J., Vijayakrishnan, J., Papaemmanuil, E., Koehler, R., Greaves, M. *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* **115**, 1765–1767 (2010).
- 9 Ellinghaus, E., Stanulla, M., Richter, G., Ellinghaus, D., te Kronnie, G., Cario, G. *et al.* Identification of germline susceptibility loci in ETV6-RUNX1-rearranged childhood acute lymphoblastic leukemia. *Leukemia* **26**, 902–909 (2012).
- 10 Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
- 11 Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
- 12 Denny, C. T., Yoshikai, Y., Mak, T. W., Smith, S. D., Hollis, G. F. & Kirsch, I. R. A chromosome 14 inversion in a T-cell lymphoma is caused by site-specific recombination between immunoglobulin and T-cell receptor loci. *Nature* **320**, 549–551 (1986).
- 13 Forrester, W. C., Novak, U., Gelinas, R. & Groudine, M. Molecular analysis of the human beta-globin locus activation region. *Proc. Natl Acad. Sci. USA* **86**, 5439–5443 (1989).
- 14 Griesser, H., Tkachuk, D., Reis, M. D. & Mak, T. W. Gene rearrangements and translocations in lymphoproliferative diseases. *Blood* **73**, 1402–1415 (1989).
- 15 Hecht, F., Morgan, R., Hecht, B. K. & Smith, S. D. Common region on chromosome 14 in T-cell leukemia and lymphoma. *Science* **226**, 1445–1447 (1984).
- 16 McGuire, E. A., Hockett, R. D., Pollock, K. M., Bartholdi, M. F., O'Brien, S. J. & Korsmeyer, S. J. The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including Ttg-1, a gene encoding a potential zinc finger protein. *Mol. Cell. Biol.* **9**, 2124–2132 (1989).
- 17 Xu, H., Yang, W., Perez-Andreu, V., Devidas, M., Fan, Y., Cheng, C. *et al.* Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. *J. Natl. Cancer. Inst.* **105**, 733–742 (2013).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)