

COMMENTARY

***CEBPE* promoter SNPs, caught red handed? A commentary on identification of functional nucleotide and haplotype variants in the promoter of the *CEBPE* gene**

Heng Xu and Jun J Yang

Journal of Human Genetics (2013) 58, 571–572; doi:10.1038/jhg.2013.81; published online 1 August 2013

Acute lymphoblastic leukemia (ALL) is the most common cancer in children and the cure rates have improved steadily, thanks to progressive intensification and individualization of chemotherapy.¹ However, the molecular etiology of childhood ALL remains elusive. Inherited predisposition has recently been established by genome-wide association studies (GWAS) identifying *ARID5B*, *IKZF1*, *CEBPE*, *PIP4K2A* and *CDKN2A/CDKN2B* as ALL-susceptibility loci.^{2–4} For example, germline variants in *ARID5B* had the strongest association with ALL susceptibility across the genome, particularly with the risk of developing hyperdiploid ALL.^{2,3} Little is known about the physiologic functions of *ARID5B* in hematopoietic tissues, but limited phenotyping of *Arid5b* knockout mice suggests that it may be related to lymphoid cell development.⁵ Recurrent somatic mutations in *ARID5B* have also been described recently in endometrial carcinoma.⁶ *IKZF1*, an important transcription factor in all lymphoid lineages, is frequently targeted by copy-number alterations in ALL blast cells (particularly in high-risk ALL), and *IKZF1* deletion is associated with a poor prognosis.⁷ Loss of *CDKN2A/CDKN2B* occurs in up to 40% of B-precursor ALL and is likely to contribute to cell cycle deregulation in leukemia.⁷ The remarkable convergence of germline ALL-susceptibility variants and somatic aberrations on genes involved in

lymphoid cell development, cell cycle control, and tumor suppression reinforces the contribution of these key pathways to leukemogenesis and also points to the possibility that inherited and acquired genetic variations act synergistically in the development of childhood ALL.

Although a number of ALL-susceptibility loci have been identified with relatively large effect sizes and robust independent replications, the exact functional consequences of these germline variants remain largely unknown. In lymphoblastoid cells, the expression of *IKZF1* and *PIP4K2A* is associated with ALL risk variants (rs4132601 in *IKZF1*³ and rs7088318 in *PIP4K2A*,⁴ respectively), with disease risk alleles linked to higher gene expression. Similar analyses found no associations between gene transcription and susceptibility variants in *ARID5B*, *CEBPE* and *CDKN2A/2B*.^{2–4} In this issue of the journal, Lee and colleagues⁸ followed up on the *CEBPE* locus to describe the function of a number of single-nucleotide length polymorphisms (SNPs) that were previously linked to ALL susceptibility. Focusing on a 2-kb window upstream of the *CEBPE* transcription start site, they evaluated seven SNPs (including the original GWAS hit rs2239633) for their effects on transcription activity using a luciferase reporter assay. Single-nucleotide substitutions at rs2239632 and rs2239633 resulted in an increase in transcription of the reporter gene, whereas the remaining five SNPs showed no effects. In combination, multi-SNP haplotypes harboring ALL risk alleles also showed elevated

luciferase activity compared with the construct with the reference haplotype. This study is timely and its findings may provide initial clues as to how *CEBPE* is linked to ALL. However, a number of points should be raised in discussing these results: (1) did they find the causal variant explaining the GWAS signal at *CEBPE*? Probably not. The promoter activity assay suggested that both rs2239632 and rs2239633 were functional (although the latter had a somewhat weaker effect), but did not lend itself to establish causality of these variants for ALL susceptibility. Future studies are warranted to genotype both SNPs in children with ALL and in non-ALL controls, in which multivariate analyses can determine whether rs2239632 completely explains the GWAS signal at rs2239633. A large sample size or inclusion of non-European populations might be necessary because of the strong linkage disequilibrium at this locus in the Europeans. Further, it would be prudent to systematically resequence this locus to fully understand the contribution of these genetic variants to ALL risk. The results from recent GWAS follow-up studies so far suggest that many loci harbor multiple variants that independently contribute to disease susceptibility.⁹ (2) Why were the associations between *CEBPE* SNP genotypes and gene transcription not seen in previous studies? The observation that rs2239633 directly influenced *CEBPE* promoter activity is intriguing, as in the original ALL GWAS by Papaemmanuil *et al.* rs2239633 was not related to *CEBPE* expression in lymphoblastoid cell lines.³ Queries in other

H Xu and JJ Yang are at the Department of Pharmaceutical Sciences, St Jude Children's Research Hospital, Memphis, TN 38105, USA
E-mail: jun.yang@stjude.org

expression quantitative trait locus (eQTL) databases also showed little evidence that these two variants influence transcription activity of *CEBPE* (e.g., <http://www.ncbi.nlm.nih.gov/gtex/GTEX2/gtex.cgi> and <http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>). However, a number of DNase I hypersensitivity sites have been noted within this region in the ENCODE data (<http://genome.ucsc.edu/ENCODE/>). Further functional studies are needed to experimentally confirm that transcription factor binding affected by these alleles. (3) What does this study tell us about the mechanistic links between *CEBPE* and ALL? CEBPs (CACCT enhancer-binding proteins) have key roles in myeloid differentiation. *CEBPE* is specifically related to myeloid cell maturation and terminal differentiation, and mutations of *CEBPE* lead to neutrophil-specific granule deficiency.¹⁰ In childhood ALL, intrachromosomal translocations involving *IGH* and *CEBPE* have been described, resulting in the upregulation of *CEBPE* expression.¹⁰ Therefore, it is reasonable to postulate that increased *CEBPE* transcription secondary to germline genetic variation (rs2239632) may expose this locus to DNA damage and subsequent translocations, or that higher *CEBPE* expression could

modestly alter hematopoietic differentiation and prime the lymphoid progenitor cells for leukemogenesis.

Functional follow-up studies of GWAS hits are challenging because (1) the candidate genomic loci are often large with extended linkage disequilibrium, thus requiring extensive resequencing or high-density genotyping, (2) the genetic architecture of most diseases is unknown and testing different scenarios will require completely different research strategies (rare variants versus common variants, a single causal variant versus combined effects of multiple variants underlying a GWAS signal) and (3) the functional effects of even causal variants may be modest, requiring robust experimental assays. This and future efforts will reveal novel leukemia biology and lead to more innovative and efficacious therapies for children with ALL.

1 Pui, C. H. & Evans, W. E. Treatment of acute lymphoblastic leukemia. *N. Engl. J. Med.* **354**, 166–178 (2006).

2 Trevino, 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).

3 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).

4 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).

5 Lahoud, M. H., Risteovski, S., Venter, D. J., Jermiin, L. S., Bertonecello, I., Zavarsek, S. *et al.* Gene targeting of Desrt, a novel ARID class DNA-binding protein, causes growth retardation and abnormal development of reproductive organs. *Genome Res.* **11**, 1327–1334 (2001).

6 Cancer Genome Atlas Research NKandath, C., Schultz, N., Cherniack, A. D., Akbani, R., Liu, Y., Shen, H. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).

7 Mullighan, C. G., Su, X., Zhang, J., Radtke, I., Phillips, L. A., Miller, C. B. *et al.* Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N. Engl. J. Med.* **360**, 470–480 (2009).

8 Ryoo, H., Kong, M., Kim, Y. & Lee, C. Identification of functional nucleotide and haplotype variants in the promoter of the *CEBPE* gene. *J. Hum. Genet.* **58**, 600–603 (2013).

9 Huyghe, J. R., Jackson, A. U., Fogarty, M. P., Buchkovich, M. L., Stančáková, A., Stringham, H. M. *et al.* Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* **45**, 197–201 (2013).

10 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).