



# HHS Public Access

Author manuscript

*Leukemia*. Author manuscript; available in PMC 2016 May 18.

Published in final edited form as:

*Leukemia*. 2016 May ; 30(5): 1194–1197. doi:10.1038/leu.2015.251.

## A Functional Polymorphism in the *CEBPE* Gene Promoter Influences Acute Lymphoblastic Leukemia Risk through Interaction with the Hematopoietic Transcription Factor Ikaros

Joseph L. Wiemels<sup>1,2</sup>, Adam J. de Smith<sup>1</sup>, Jianqiao Xiao<sup>1</sup>, Seung-Tae Lee<sup>3</sup>, Marcus O. Muench<sup>3</sup>, Marina E. Fomin<sup>4</sup>, Mi Zhou<sup>1</sup>, Helen M. Hansen<sup>2</sup>, Amanda Termuhlen<sup>5</sup>, Catherine Metayer<sup>6</sup>, and Kyle M. Walsh<sup>2</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA

<sup>2</sup>Division of Neuroepidemiology, Department of Neurological Surgery, University of California San Francisco, San Francisco, California USA

<sup>3</sup>Department of Laboratory Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

<sup>4</sup>Blood Systems Research Institute and Department of Laboratory Medicine, University of California San Francisco, San Francisco, California, USA

<sup>5</sup>Jonathan Jaques Children's Cancer Center, Department of Pediatrics, Keck School of Medicine, University of South California, Long Beach, California USA

<sup>6</sup>School of Public Health, University of California Berkeley, Berkeley, California USA

### Keywords

Childhood acute lymphoblastic leukemia; genetic association; functional genetics; transcription factor binding; ChIP; fine-mapping; *CEBPE*; Ikaros

---

Heritable single nucleotide polymorphisms (SNPs) in six genomic loci have been associated with childhood ALL risk in genome-wide association studies (GWAS) of European-ancestry populations. Four of these heritable ALL risk loci are near genes crucial for hematopoietic

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: [http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

**CORRESPONDING AUTHOR:** Joseph L. Wiemels, PhD, UCSF Helen Diller Family Cancer Center, Box 0520, 1450 3rd Street, HD 481, San Francisco, CA 94158-0520, ; Email: Joe.Wiemels@ucsf.edu, Phone: 415-514-0577

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### AUTHOR CONTRIBUTIONS

JLW and KMW designed experiments, analyzed the data, and wrote the manuscript. AJdS and JQ designed and performed experiments and helped write the manuscript. S-TL helped analyze the data. MOM and MEF supplied key reagents and helped perform experiments. MZ, and HH performed experiments. AT and CM secured clinical and population samples. All authors read and finalized the manuscript.

### SUPPLEMENTARY INFORMATION:

Supplementary Information accompanies this paper on the *Leukemia* website (<http://www.nature.com/leu>)

differentiation, including *CEBPE*<sup>1,2</sup>. *CEBPE* encodes one of a family of six known CCAAT-targeting bZIP transcription factors which critically regulate myelopoiesis. To identify causal polymorphisms underlying the childhood ALL association peak near *CEBPE*, we performed SNP genotyping, imputation-based fine-mapping, and functional validation analyses of the chromosome 14q11.2 locus in a multi-ethnic case-control population of Hispanic children from the California Childhood Leukemia Study (CCLS) and European-ancestry children from the Children's Oncology Group (COG), totaling 1277 cases and 3078 controls.

Supplementary Figure 1 and Supplementary Table 1 summarize the study design and study populations (full methods in online supplement). Briefly, 297 Hispanic B-cell ALL (B-ALL) cases and 454 Hispanic controls underwent genotyping and targeted fine-mapping. Additional GWAS data was used to perform imputation-based fine-mapping in 980 European-ancestry COG ALL subjects and 2624 Wellcome Trust Case Control Consortium (WTCCC)<sup>3</sup>. An additional 271 cases and 436 controls from the CCLS underwent targeted Sequenom genotyping. Association tests were performed using an allelic additive model, adjusted for ancestry-informative principal components.

Using all available genome-wide SNP data, we imputed all the common SNPs in the *CEBPE* region using a multi-ethnic reference panel from the 1000 Genomes Project. Three SNPs (rs2239635, rs8007237, rs56031127) were more significantly associated than either of the previously reported GWAS hits (rs2239633 and rs4982731) (Supplementary Table 2)<sup>1,4</sup>. The most statistically significant association detected in analysis of the combined CCLS and COG data was at rs2239635 ( $P_{\text{meta}}=5.1 \times 10^{-11}$ , O.R.=1.45, 95% C.I.=1.30-1.62) (Figure 1a). Imputed genotype and directly assayed genotype at rs2239635 were 97.9% concordant, validating imputation accuracy.

In addition to being more significantly associated than the previously reported GWAS hits, rs2239635 remained significantly associated with ALL risk in conditional analyses adjusted for rs2239633 and for rs4982731 ( $P=8.4 \times 10^{-5}$  and 0.011, respectively) (Figure 1b-c). Furthermore, in conditional analyses adjusted for this new top hit, the previous ALL GWAS hits at rs2239633 and rs4982731 no longer reached statistical significance ( $P=0.076$  and 0.11, respectively) (Figure 1d), suggesting that rs2239635, or a variant in strong LD, underlies the *CEBPE* association signal first detected by GWAS.

In the 100kb *CEBPE* imputation region, Haploview analysis identified 15 distinct haplotype blocks among CCLS Hispanics compared with only 8 in the COG European-ancestry population (Supplementary Figure 2). Taking advantage of this reduced LD, we performed stratified association analyses in CCLS Hispanics to further refine the *CEBPE* association signal. Among 297 Hispanic B-ALL cases and 454 Hispanic controls with GWAS data available in the CCLS, rs2239635 was again the most significantly associated SNP by nearly two orders of magnitude (Supplementary Table 2). The minor allele was associated with a 1.87-fold increase in the odds of B-ALL in Hispanics (95% CI=1.51-2.30;  $P=6.2 \times 10^{-9}$ ).

SNPs in the *CEBPE* region generally showed stronger associations with the high-hyperdiploid B-ALL subgroup (Supplementary Table 2 and Supplementary Figure 3). The minor allele of rs2239635 conferred 2.94-fold increased odds of high-hyperdiploid B-cell

ALL (95% CI = 2.12-4.06;  $P = 8.2 \times 10^{-11}$ ), but only 1.51-fold increased odds of non-hyperdiploid B-ALL (95% CI = 1.11-2.04;  $P = 8.8 \times 10^{-3}$ ). In case-case comparisons, rs2239635 was associated with a 1.66-fold increase in the odds of high-hyperdiploid B-ALL relative to non-hyperdiploid B-ALL ( $P=0.0034$ ), demonstrating a robust difference in the magnitude of effect across subtypes.

Top associated SNPs were functionally annotated using ENCODE2 data<sup>5,6</sup>, and assessed if they were cis eQTLs for *CEBPE*<sup>7</sup>. Functional annotations indicated that the top-ranked SNP, rs2239635, impacts *CEBPE* expression ( $P=0.01$ ) and is among the most functionally-relevant variants in the region (Supplementary Table 2). These annotations also suggest a potential mechanism by which rs2239635 affects expression, as the SNP is located within an Ikaros transcription factor binding site and is predicted to affect TF binding<sup>8</sup>. Ikaros is encoded by the *IKZF1* gene, a critical mediator of hematopoietic differentiation which has also been associated with ALL risk in GWAS<sup>1,2</sup>.

To test the predicted disruption of the Ikaros binding site, we identified three normal fetal bone marrow samples heterozygotes at rs2239635. Such heterozygotes contain a natural “test allele” along with a “control allele” for comparison. In ChIP experiments, the wild-type (WT) allele, which maintains the canonical Ikaros binding motif (TTTGGGAGG, Figure 2a), immunoprecipitated significantly more efficiently than the risk allele ( $P<0.05$ ) (Figure 2b). This indicates that the risk allele “C” disrupts Ikaros binding near *CEBPE* in its native physiological state. However, it is unlikely that the “C” allele completely abolishes binding due to its location at the periphery of the Ikaros binding motif and it is noted that the IgG control also displayed a slight assay readout bias towards the “C” allele.

Using a model expression vector, we tested whether the genomic region including rs2239635 affects transcription of *CEBPE*. Since Ikaros is a known transcriptional repressor we cloned the segment into a vector with a known strong transcriptional activation sequence, the promoter of *PTPRG*<sup>9</sup>. Plasmids containing the *CEBPE* promoter segment were significantly repressive compared to the *PTPRG* promoter alone ( $P < 0.001$ ), confirming the repressive properties of the *CEBPE* promoter segment. Gene repression was not significantly different for the “C” allele than the “G” allele possibly due to the comparatively low expression of Ikaros in 293 cells.

Because a previous GWAS identified an ALL risk locus at rs4132601 in *IKZF1* that is associated with reduced gene expression<sup>1</sup>, and because our experiments suggest biologic interaction between *IKZF1* and *CEBPE*, we tested for statistical interaction between SNPs. In case-control analyses of high-hyperdiploid B-ALL, rs2239635 and rs4132601 displayed significant multiplicative interaction in CCLS Hispanics ( $P=0.021$ ;  $N=147$  cases, 714 controls). A weaker interaction was detected in analysis of all CCLS B-ALL samples with genotyping data for both SNPs ( $P=0.061$ ;  $N=449$  cases, 714 controls). The direction of the interaction indicates that the combined effect of rs2239635 (*CEBPE*) and rs4132601 (*IKZF1*) risk alleles is greater than would be expected if they operated independently. Although rs2239635 showed a stronger effect in high-hyperdiploid leukemia, effects were similar in cases with *IKZF1* deletions as in those without (OR=2.17 and 2.23, respectively).

*CEBPE* and *IKZF1* expression levels were compared to investigate changes in expression during B-cell development. *CEBPE* expression was highest in multipotent progenitor cells (S1) and declined sharply as cells progressed to B-cell-committed progenitors, including pre-B-I cells (S2), pre-B-II cells (S3) and immature B-cells (S4) (Supplementary Figure 4a). *CEBPE* expression dropped again when cells transitioned to mature B-cells. This pattern was tightly and inversely coordinated with expression of *IKZF1* (Supplementary Figure 4b). As expected, expression of both *CEBPE* and *IKZF1* in patient leukemic B-cells was most similar to that in early stage B-cells (Supplementary Figure 4), believed to be the B-ALL cell-of-origin.

We identify an allele in the *CEBPE* gene promoter that impacts risk of childhood ALL by disrupting binding of the Ikaros transcriptional repressor. Additionally, we note a cytogenetic subtype-specific association between the functional SNP of *CEBPE* and high-hyperdiploid B-ALL. Because high-hyperdiploid leukemias typically carry an extra chromosome 14, leading to 3 copies of *CEBPE*, overexpression of *CEBPE* may be critical for development of high-hyperdiploid B-ALL.

*CEBPE* is a critical modulator of myelopoiesis. Disruption of C/EBP proteins, including C/EBP $\epsilon$ , alters myelopoiesis<sup>10</sup>. C/EBP $\epsilon$  is not required for B-cell maturation or function, opening a question as to why a polymorphism affecting a subtype of pre-B-ALL may be located proximal to the gene. As progenitor cells choose their lineage fate, transcription factors that are required for other fates must be shut off permanently<sup>11</sup>. Also, maturation towards a pre-B cell from a common lymphoid progenitor is associated with strong expression of Ikaros, a transcription factor essential for B-cell development with both activating and repressive properties<sup>12</sup>. Our results suggest that lineage commitment for pre-B cells involves the suppression of *CEBPE* by Ikaros, and a polymorphism that disrupts this repression promotes B-ALL risk. Incomplete suppression of *CEBPE* by Ikaros may lead to lineage confusion, a common feature of leukemogenesis<sup>13</sup>.

That overexpression of a non-B-cell specific transcription factor would increase risk of a B-cell malignancy is not without precedent. A heritable polymorphism in the T-cell-specific transcription factor *GATA3* is associated with higher *GATA3* expression and also with increased B-ALL risk<sup>14</sup>. Together, these observations suggest that disruption of the precise control of lineage choice and commitment may be a more common feature of leukemogenesis than represented by the rare diagnosis of mixed-lineage leukemias (typically associated with *MLL1* or *BCR-ABL1* gene translocations). Common heritable genetic modifiers are likely to contribute much more subtle and specific alterations in hematopoietic cell development than the broad epigenetic changes induced by acquired gene fusions, but may add significantly to the burden of disease due to their high prevalence. Additional work is needed to conclusively determine the role of *CEBPE* control in lymphoid neoplasia; however, its connection with Ikaros function is supported by our ChIP experiments and a demonstrated statistical interaction between the two SNPs in conferring leukemia risk.

## Supplementary Material

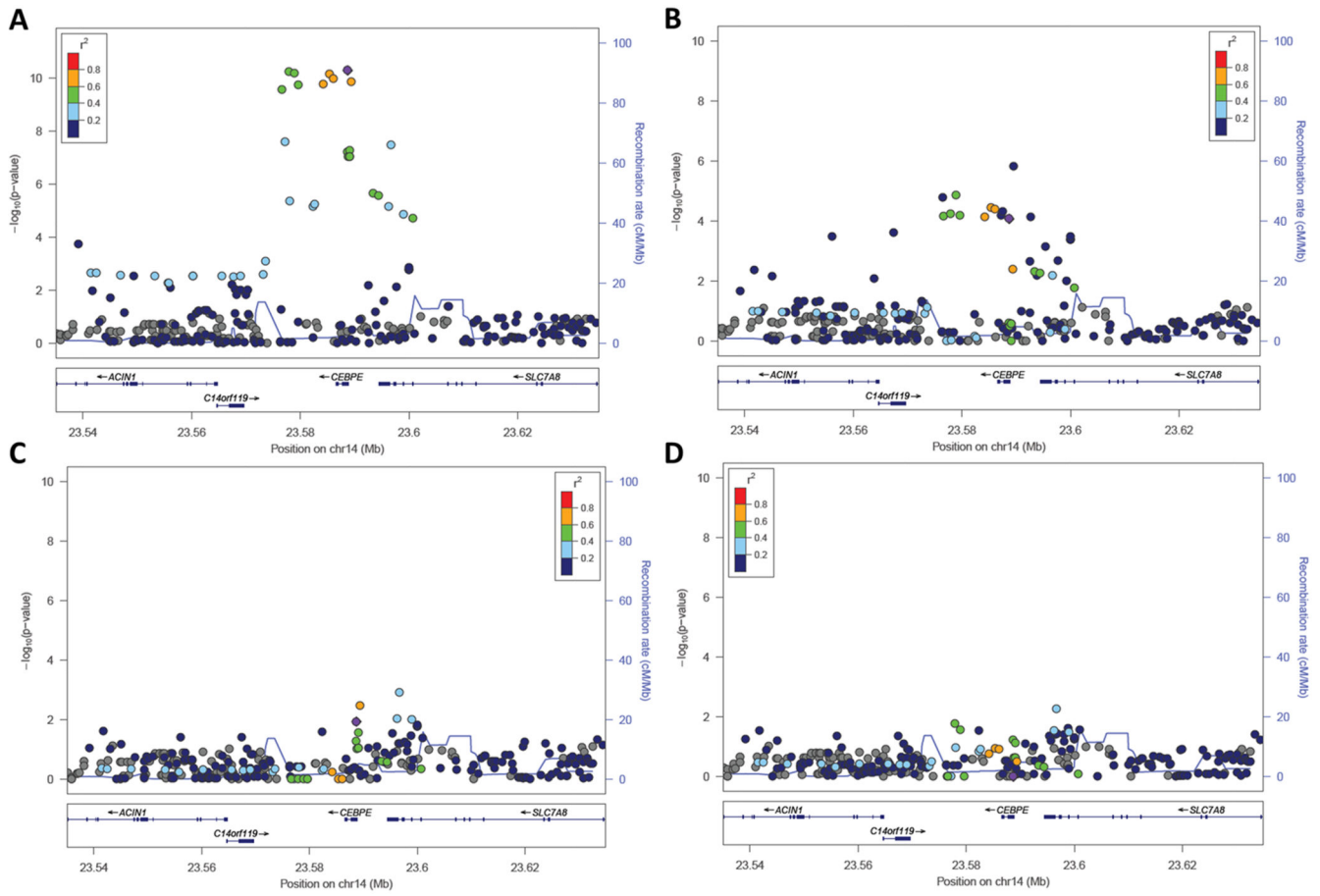
Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

This work was financially supported by NCI R01CA155461 (J.L.W., K.M.W.), NCI R25CA112355 (K.M.W.), NIEHS and EPA P01ES018172 (C.M., J.L.W.), NIEHS R01ES09137 (C.M., J.L.W.), and the National Institute of Diabetes and Digestive and Kidney Diseases at the National Institutes of Health P01DK088760 (M.O.M. and M.E.F.). A full list of acknowledgements can be found in the online supplement.

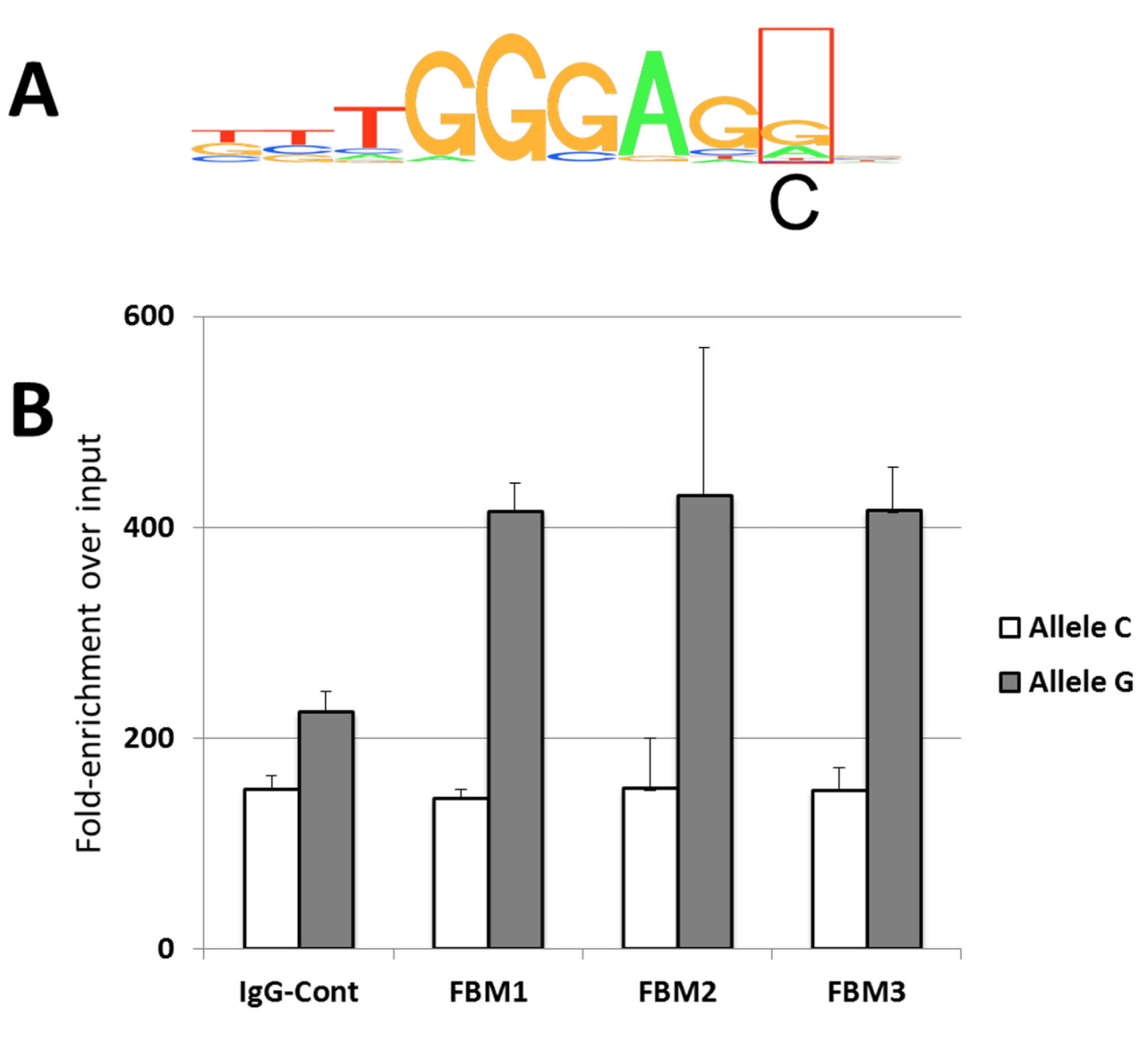
## REFERENCES

1. Papaemmanuil E, Hosking FJ, 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. *Nature genetics*. Sep; 2009 41(9):1006–1010. [PubMed: 19684604]
2. Trevino LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nature genetics*. Sep; 2009 41(9):1001–1005. [PubMed: 19684603]
3. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. Jun 7; 2007 447(7145):661–678. [PubMed: 17554300]
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. *Journal of the National Cancer Institute*. May 15; 2013 105(10):733–742. [PubMed: 23512250]
5. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res*. Jan.2012 40:D930–934. Database issue. [PubMed: 22064851]
6. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. Sep; 2012 22(9):1790–1797. [PubMed: 22955989]
7. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. Jun; 2013 45(6):580–585. [PubMed: 23715323]
8. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. *Genome Res*. Mar; 2011 21(3):447–455. [PubMed: 21106904]
9. Xiao J, Lee ST, Xiao Y, Ma X, Andres Houseman E, Hsu LI, et al. PTPRG inhibition by DNA methylation and cooperation with RAS gene activation in childhood acute lymphoblastic leukemia. *Int J Cancer*. 135(5):1101–9. [PubMed: 24496747]
10. Yamanaka R, Barlow C, Lekstrom-Himes J, Castilla LH, Liu PP, Eckhaus M, et al. Impaired granulopoiesis, myelodysplasia, and early lethality in CCAAT/enhancer binding protein epsilon-deficient mice. *Proc Natl Acad Sci U S A*. Nov 25; 1997 94(24):13187–13192. [PubMed: 9371821]
11. Starnes LM, Sorrentino A. Regulatory circuitries coordinated by transcription factors and microRNAs at the cornerstone of hematopoietic stem cell self-renewal and differentiation. *Curr Stem Cell Res Ther*. Jun; 2011 6(2):142–161. [PubMed: 21190539]
12. Schwickert TA, Tagoh H, Gultekin S, Dakic A, Axelsson E, Minnich M, et al. Stage-specific control of early B cell development by the transcription factor Ikaros. *Nat Immunol*. Mar; 2014 15(3):283–293. [PubMed: 24509509]
13. Schmidt CA, Przybylski GK. What can we learn from leukemia as for the process of lineage commitment in hematopoiesis? *Int Rev Immunol*. Feb; 2001 20(1):107–115. [PubMed: 11342300]
14. Perez-Andreu V, Roberts KG, Harvey RC, Yang W, Cheng C, Pei D, et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nature genetics*. Dec; 2013 45(12):1494–1498. [PubMed: 24141364]
15. Migliorini G, Fiege B, Hosking FJ, Ma Y, Kumar R, Sherborne AL, et al. Variation at 10p12.2 and 10p14 influences risk of childhood B-cell acute lymphoblastic leukemia and phenotype. *Blood*. Nov 7; 2013 122(19):3298–3307. [PubMed: 23996088]



**Figure 1. Association of SNPs in the *CEBPE* locus with ALL risk, with adjustment for previous GWAS lead SNPs and novel lead SNP rs2339635**

Lead SNP rs2339635 appears in purple, with other SNPs colored to indicate their degree of linkage with rs2339635. A) Association of SNPs in the *CEBPE* locus with ALL risk. B) Association of SNPs in the *CEBPE* locus with ALL risk, adjusted for rs2239633, the top hit from two previous GWAS<sup>1,15</sup>. C) Association of SNPs in the *CEBPE* locus with ALL risk, adjusted for rs4982731, the top hit from a previous GWAS<sup>4</sup>. D) Association of SNPs in the *CEBPE* locus with ALL risk, adjusted for rs2239635, the top hit from our fine-mapping analyses.



**Figure 2. Chromatin immunoprecipitation of *CEBPE* promoter region**

(A) The region of the promoter of *CEBPE* showing the perfect match for IKZF1 binding (*Regulome* database) and the position of rs2339635 (boxed). The identity of the risk allele is indicated below the box. (B) Bone marrow from 3 independent subjects heterozygous for rs2339635 was fixed, sonicated, and precipitated with an antibody for IKZF1. Assays for each allele of *CEBPE* pulled down only the “G” variant of rs2339635, without pulling down the “C” variant any more than a nonspecific IgG control antibody ( $P < 0.05$ , t-test) using a ddPCR assay. ddPCR analysis of the input (prior to ChIP with either IKZF1 or isotype control) is displayed on the left side of the figure.