SUPPLEMENTARY INFORMATION:

Ethics statement: This study was approved and reviewed by all collaborating institutions, including the institutional review committees at the California Department of Public Health (CDPH), the University of California, Berkeley, and the University of California, San Francisco. All parent respondents provided written informed consent.

CCLS study population: Hispanic case and control subjects for this analysis were derived from the CCLS, a leukemia case-control study with near population-based ascertainment and control selection from the California Birth Cohort (1995-2008). DNA was isolated from dried bloodspots collected at birth and archived by the Genetic Diseases Screening Program of the California Department of Public Health. A total of 297 Hispanic B-cell ALL cases and 454 Hispanic controls are included in the fine-mapping analyses. Another 271 cases and 436 controls from the CCLS underwent targeted genotyping of the newly identified top hit.

COG study population: Constitutive DNA from European-ancestry COG cases was extracted from peripheral blood samples obtained during clinical remission for children with ALL, recruited under COG protocols P9904 and P9905 as previously detailed [1]. Genotype data for these 980 subjects were obtained from dbGaP study accession phs000638.v1.p1 (Genome-Wide Association Study of Relapse of Childhood Acute Lymphoblastic Leukemia), and have been described in detail [1, 2]. Case-control comparisons were made with 2624 European-ancestry individuals accessed from the Wellcome Trust Case Control Consortium (WTCCC) [3].

Genome-wide genotyping: Constitutive DNA samples from CCLS cases and controls were genotyped at the UC Berkeley School of Public Health Genetic Epidemiology and Genomics Laboratory using the Illumina Human OmniExpress v1 platform, as previously described [4].

Extraction from dried bloodspots was performed using the QIAamp DNA Mini Kit (QIAGEN, USA, Valencia, CA). Genotype reproducibility was verified using ten duplicate samples with average concordance across all genotypes > 99.99%. Samples with genotyping call rates < 98% were excluded from analyses. Samples were screened for cryptic relatedness using 10,000 unlinked SNPs, and were excluded if the proportion of the genome identical-by-descent was > 0.15. Samples with discordant sex information (reported vs. genotyped sex) were excluded from further analyses. SNPs with genotyping call rates <98% were excluded from analyses. To exclude poorly genotyped SNPs, any SNP with a Hardy-Weinberg Equilibrium (HWE) P-value <1x10-5 in controls was removed from analyses.

Constitutive DNA from European-ancestry COG cases were genotyped on the Affymetrix 6.0 array and genotype data were downloaded from dbGaP accession phs000638.v1.p1. Control genotype data for European-ancestry control samples genotyped on the Affymetrix 6.0 array were downloaded from the Wellcome Trust Case-Control Consortium [3]. Genotyping quality-control procedures were conducted independently in cases and in controls and were identical to those performed in CCLS samples. Genome-wide SNP data were used to ensure there was no overlap between the 297 Hispanic CCLS cases and the 980 European-ancestry COG cases.

CCLS Targeted SNP genotyping: Targeted genotyping was performed using a custom Sequenom panel containing top novel hits from our previous GWAS of Hispanic ALL [4, 5], known tag SNPs from previous GWAS [6-8], and the top newly identified SNPs from imputation analysis of CEBPE. A total of 1092 CCLS samples (properly randomized) were successfully genotyped using the Sequenom panel (call rates > 95%). This included 385 samples from the original GWAS, 271 non-overlapping B-cell ALL cases, and 436 non-overlapping controls.

Targeted genotyping was used to validate genotypes inferred from imputation analysis and to improve power in SNPxSNP interaction tests by increasing the number of samples with genotype data for top hits [*e.g.* rs2239635 (*CEBPE*) and rs4132601 (*IKZF1*)].

Additional controls for population stratification: All CCLS children with GWAS data had at least one parent who self-identified as Hispanic. Children had >5% Native American ancestry as determined by Structure v2.3.1, and a greater level of Native American ancestry than sub-Saharan African ancestry [9]. Founder population allele frequencies were defined using data from 63,303 overlapping unlinked SNPs genotyped in Human Genome Diversity Project subjects (111 sub-Saharan Africans, 107 Native Americans, 154 Europeans) [10]. To adjust for ancestry in SNP association analyses, a principal component approach was implemented using Eigenstrat [11]. Principal components were calculated separately in the Hispanic CCLS case-control data and the European-ancestry COG data. The first five principal components from Eigenstrat analyses were included as covariates in case-control analyses.

SNP Imputation: Imputation was performed using IMPUTE v2.3.1 software and its standard Markov chain Monte Carlo algorithm and default settings for targeted imputation [12]. All 1,000 Genomes Phase I integrated haplotypes were provided as the imputation reference panel [13], including 1274 SNPs. SNPs with imputation quality (info) scores less than 0.70 (N=41) or posterior probabilities less than 0.90 (N=13) were excluded to remove poorly imputed SNPs. Any SNP with a minor allele frequency <1% in case subjects was excluded from association tests (N=894), leaving 326 common variants. Association results from CCLS and COG were combined using fixed effects meta-analysis in the META software package to generate a

summary odds ratio and P-value [14]. All reported SNP associations are for an allelic additive model where odds ratios (OR) are for each additional copy of the risk allele. Inclusion of an allelic additive interaction term in the model (*i.e.* SNP_1, SNP_2, SNP_1*SNP_2) was performed to test for the presence of significant modification of effect.

Assessment of imputation accuracy: rs2239635 was directly genotyped using a Sequenom assay in 385 CCLS samples that had undergone genome-wide genotyping and imputation of the *CEBPE* region. Imputed genotype and directly assayed genotype at rs2239635 were 97.9% concordant among these 385 samples, validating imputation accuracy.

Haplotype analyses: Haplotype analyses were performed using Haploview v4.2 to identify haplotype blocks and calculate R^2 values [15]. Haplotype blocks were generated using the "Confidence Intervals" algorithm of Gabriels, et al [16].

Chromatin Immunoprecipitation Assay: Chromatin immunoprecipitation (ChIP) was done as previously described [17] with modifications. First of all, the assay was only performed on independent samples that were heterozygous at rs2239635 (n = 3, which were identified from 9 random bone marrow samples obtained from terminated fetuses). Whole cells from normal bone marrow were crosslinked for 15min at room temperature with 1% (wt/vol) formaldehyde (sigma, F8775), then lysed and sheared to 100–1000 bp in size with ChIP-IT Express Enzymatic Shearing Kit (Active Motif. Cat# 53035) according to the manufacturer's instructions. Sheared chromatin was immunoprecipitated with 10mg anti-IKZF1 antibody (Santa Cruz H-100, Cat # sc-13039) or non-specific IgG (Santa Cruz Biotechnology, Cat# sc-2025), performed as a negative control) using EZ-ChIP™ Chromatin Immunoprecipitation Kit (Millipore, Cat# 17-371) following the kit instruction manual. The agarose beads were washed and the bound

chromatin was eluted. The ChIPed and input chromatin were incubated overnight at 65°C for reversal of crosslinking. Samples were then treated with RNase A and proteinase K and purified with a Spin Filter from the Chromatin Immunoprecipitation Kit (Millipore, Cat# 17-371). Chromatin was end repaired, ligated to a linker and amplified by linker-mediated PCR (LMPCR) using GeneAmp High Fidelity PCR System (ABI, Cat# 4328212). Amplified ChIP and input DNA were purified with MinElute Reaction Cleanup Kit (Qiagen, Cat# 28204). DNA was quantified using a real-time PCR assay specifically designed for this purpose (C-LESS primers C1-TTGTATGTATGTGAGTGTGGGAGAGA; C2-TTTCTTCCACCCCTTCTCTTCC; probe 6FAM-CTCCCCCTCTAACTCTAT-MGBNFQb) [18]. Thirty nanograms DNA pull-down and input DNA.was subjected to droplet digital PCR using Taqman assay (ABI: C__16175328_10). The experiment was performed in quadruplicate from the three separate bone marrow isolates.

Luciferase assays: The cell line HEK-293 (ATCC, Manassas, Virginia) was maintained in Dulbecco's Modified Eagle Medium with 10% fetal bovine serum (FBS) (Hyclone, Logan, Utah). We confirmed the authenticity of the cell line using an STR panel (Genetica DNA Laboratories, July 2, 2013). The pGL4.23[luc2/minP] plasmid containing the luciferase reporter gene was purchased from Promega (Madison, Wisconsin). As a positive control, the *PTPRG* promoter cloned into the plasmid was used, as described previously [19]. A 206bp *CEBPE* segment from chr14:23,588,581-23,588,787 (hg19) was cloned by amplifying genomic DNA of individuals with either genotype of rs2239635 and subcloning into the vector. Since this SNP-containing element was determined to be a transcriptional repressor, we cloned it between the *PTPRG* promoter and the minP promoter of the pGL4.23 to measure its repression of *PTPRG* promotion of luciferase activity.

For Luciferase experiments, plasmid DNA was transfected into cell line HEK 293 (ATCC, Manassas, VA) using LipofectAmine 2000 according to the manufacturer's protocol. After transfection of plasmids, total cellular proteins were harvested 24 h after transfection. The Steady-Glo® Luciferase Assay System (Promega) was used to detect luciferase activity of cells transfected with plasmids encoding the luciferase gene. Luciferase activity was detected using the Bio-Tek luminometer (Bio-Tek Instruments, Winooski, VT).

Detection of somatic IKZF1 gene deletions: Multiplex ligation-dependent probe amplification (MLPA) was carried out using the SALSA MLPA probemix P335-B1 ALL-IKZF1 (MRC Holland, The Netherlands) to identify somatic loss of *IKZF1*. MLPA was performed as previously described for CCLS B-ALL samples with sufficient bone marrow DNA available [20]. Analysis of MLPA data was carried out using the Coffalyser.NET fragment analysis software (MRC Holland). A total of 171 samples had both MLPA data and rs2239635 genotyping data available for analysis.

CEBPE and IKZF1 expression levels during B-cell development: Cells from normal human fetal bone marrow were obtained and purified as previously described [21]. Four stages of B-cell precursors were isolated using flow cytometry using the markers CD34, CD19, and sIgM. Total RNA from each of eight specimens for all four stages was subjected to gene expression analysis using the GeneChip Human Gene ST 1.0 array (Affymetrix) [21]. Additionally, gene expression from 81 CCLS leukemia bone marrows were analyzed using isolated RNA in the same fashion [22]. Data were normalized using the Expression Console software (Affymetrix) with the Robust Multi-array Average (RMA) algorithm.

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.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the Environmental Protection Agency.

This study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113 and 085475. ALL GWAS data for COG subjects were obtained from dbGaP study accession phs000638.v1.p1 (Genome-Wide Association Study of Relapse of Childhood Acute Lymphoblastic Leukemia). The dataset was generated at St. Jude Children's Research Hospital and by the Children's Oncology Group, supported by NIH grants CA142665, CA21765, CA158568, CA156449, CA36401, CA98543, CA114766, CA140729, and U01GM92666, Jeffrey Pride Foundation, the National Childhood Cancer Foundation, and by ALSAC.

The CCLS is supported by our clinical collaborators and participating hospitals, which include: University of California Davis Medical Center (Dr. Jonathan Ducore), University of California San Francisco (Dr. Mignon Loh and Dr. Katherine Matthay), Children's Hospital of Central California (Dr. Vonda Crouse), Lucile Packard Children's Hospital (Dr. Gary Dahl), Children's Hospital Oakland (Dr. James Feusner), Kaiser Permanente Sacramento (Dr. Vincent Kiley), Kaiser Permanente Santa Clara (Dr. Carolyn Russo and Dr. Alan Wong), Kaiser Permanente San Francisco (Dr. Kenneth Leung), and Kaiser Permanente Oakland (Dr. Stacy

Month), and the families of the study participants. We thank the staff and faculty at San Francisco General Hospital Women's Options Center for assistance in the collection of human fetal tissues.

SUPPLEMENTARY REFERENCES:

1. Yang JJ, Cheng C, Devidas M, Cao X, Campana D, Yang W*, et al.* Genome-wide association study identifies germline polymorphisms associated with relapse of childhood acute lymphoblastic leukemia. *Blood* 2012 Nov 15; **120**(20)**:** 4197-4204.

2. Yang JJ, Cheng C, Devidas M, Cao X, Fan Y, Campana D*, et al.* Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nat Genet* 2011 Mar; **43**(3)**:** 237-241.

3. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007 Jun 7; **447**(7145)**:** 661-678.

4. Walsh KM, Chokkalingam AP, Hsu LI, Metayer C, de Smith AJ, Jacobs DI*, et al.* Associations between genome-wide Native American ancestry, known risk alleles and B-cell ALL risk in Hispanic children. *Leukemia* 2013 Dec; **27**(12)**:** 2416-2419.

5. Walsh KM, de Smith AJ, Chokkalingam AP, Metayer C, Roberts W, Barcellos LF*, et al.* GATA3 risk alleles are associated with ancestral components in Hispanic children with ALL. *Blood* 2013 Nov 7; **122**(19)**:** 3385-3387.

6. Sherborne AL, Hosking FJ, Prasad RB, Kumar R, Koehler R, Vijayakrishnan J*, et al.* Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nature genetics* 2010 Jun; **42**(6)**:** 492-494.

7. 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* 2009 Sep; **41**(9)**:** 1006-1010.

8. 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* 2013 May 15; **105**(10)**:** 733-742.

9. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003 Aug; **164**(4)**:** 1567-1587.

10. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S*, et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008 Feb 22; **319**(5866)**:** 1100-1104.

11. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* 2006 Aug; **38**(8)**:** 904-909.

12. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009 Jun; **5**(6)**:** e1000529.

13. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA*, et al.* A map of human genome variation from population-scale sequencing. *Nature* 2010 Oct 28; **467**(7319)**:** 1061-1073.

14. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L*, et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 2010 May; **42**(5)**:** 436-440.

15. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005 Jan 15; **21**(2)**:** 263-265.

16. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B*, et al.* The structure of haplotype blocks in the human genome. *Science* 2002 Jun 21; **296**(5576)**:** 2225-2229.

17. Lee ST, Xiao Y, Muench MO, Xiao J, Fomin ME, Wiencke JK*, et al.* A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res* 2012 Dec; **40**(22)**:** 11339-11351.

18. Weisenberger DJ, Trinh BN, Campan M, Sharma S, Long TI, Ananthnarayan S*, et al.* DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. *Nucleic Acids Res* 2008 Aug; **36**(14)**:** 4689-4698.

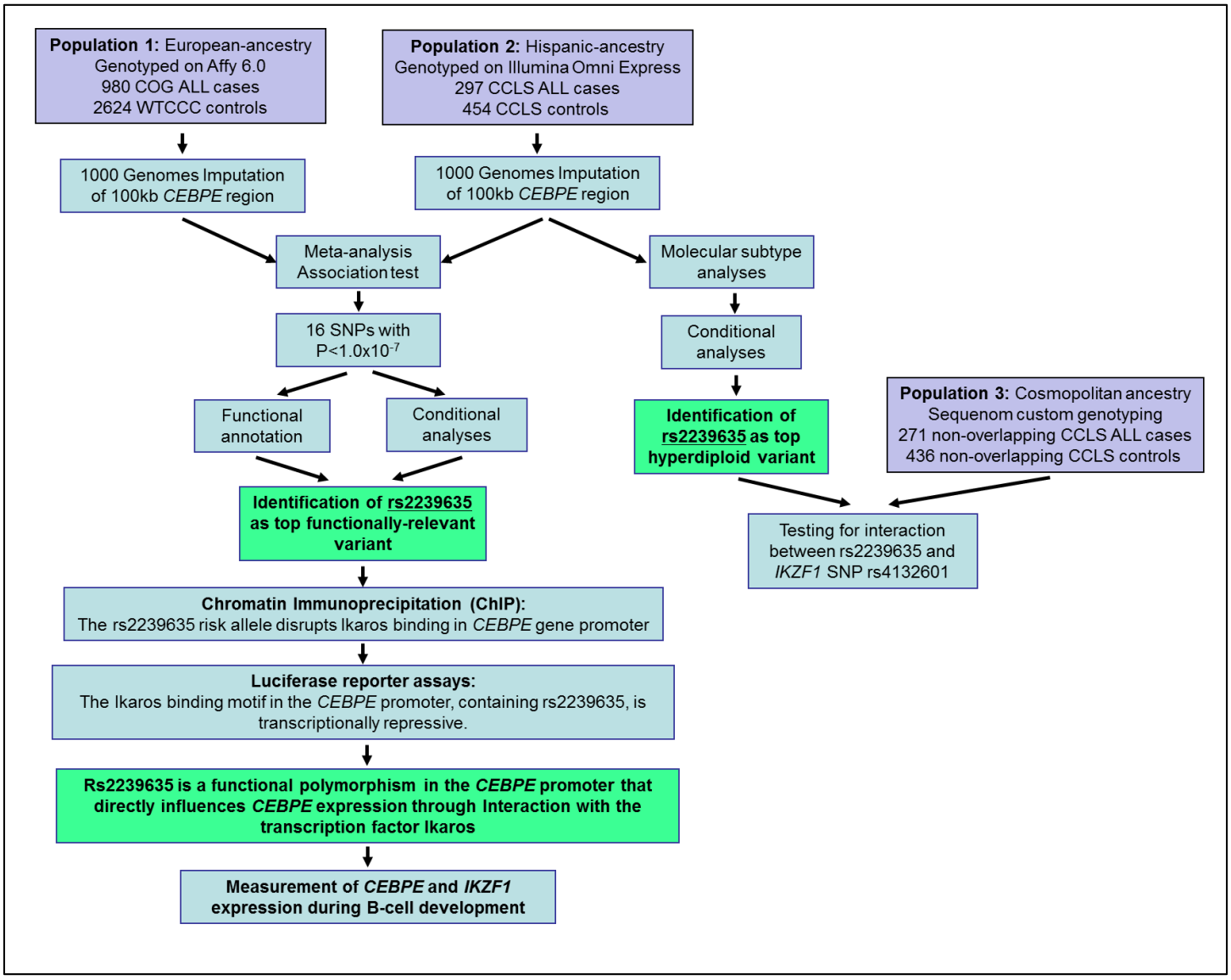
19. Xiao J, Lee ST, Xiao Y, Ma X, Andres Houseman E, Hsu LI*, et al.* PTPRG inhibition by DNA methylation and cooperation with RAS gene activation in childhood acute lymphoblastic leukemia. *Int J Cancer* 2014 Feb 4.

20. Walsh KM, de Smith AJ, Welch TC, Smirnov I, Cunningham MJ, Ma X*, et al.* Genomic ancestry and somatic alterations correlate with age at diagnosis in Hispanic children with B-cell acute lymphoblastic leukemia. *Am J Hematol* 2014 Jul; **89**(7)**:** 721-725.

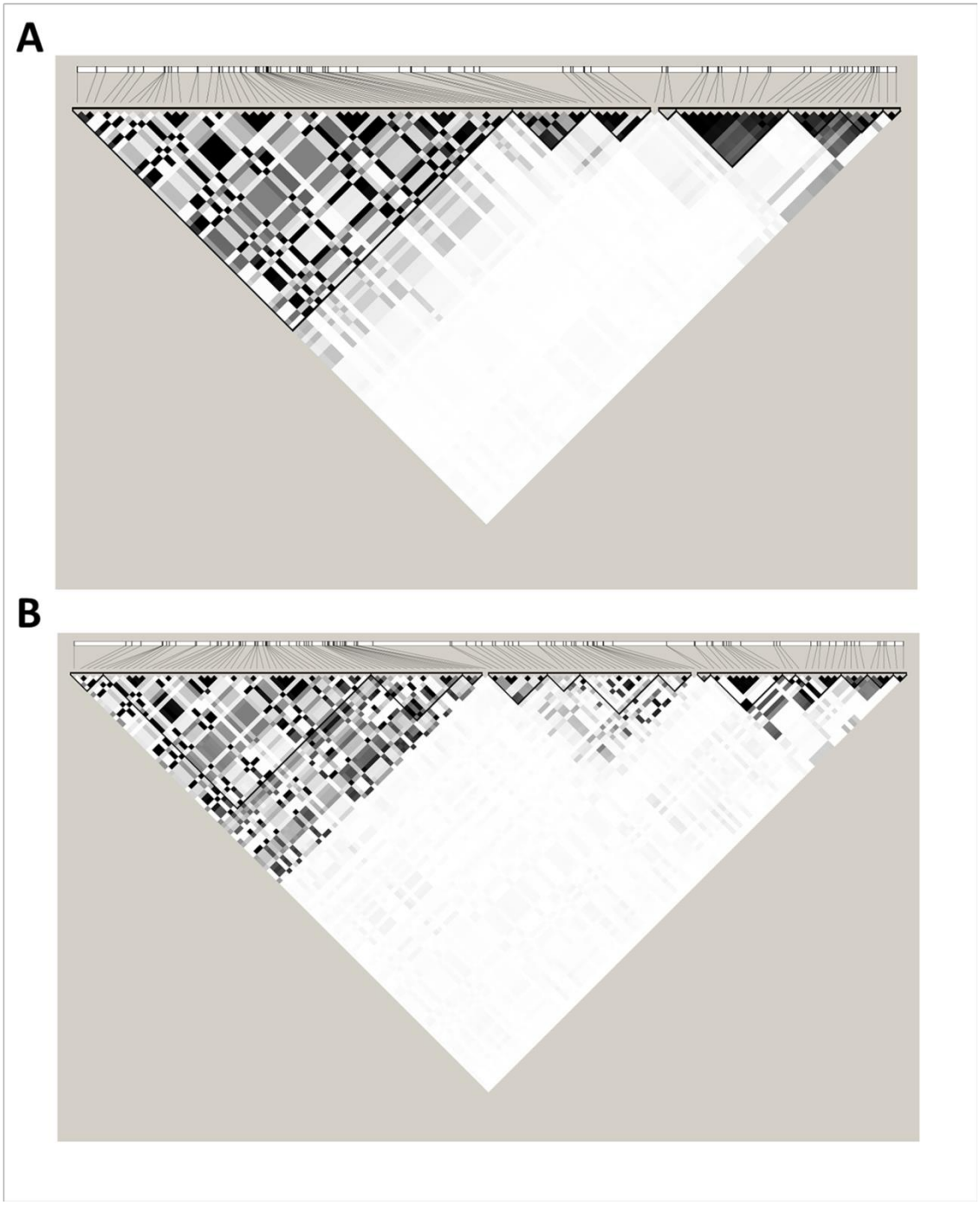
21. Lee ST, Xiao Y, Muench MO, Xiao J, Fomin ME, Wiencke JK*, et al.* A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic acids research* 2012 Dec 1; **40**(22)**:** 11339-11351.

22. Lee S-T, Muench MO, Fomin ME, Xiao J, Zhou M, de Smith AJ*, et al.* Epigenetic remodeling in B-cell acute lymphoblastic leukemia occurs in two tracks and employs embryonic stem cell-like signatures. *Nucleic Acids Res* 2015; **43**(5): 2590-602.

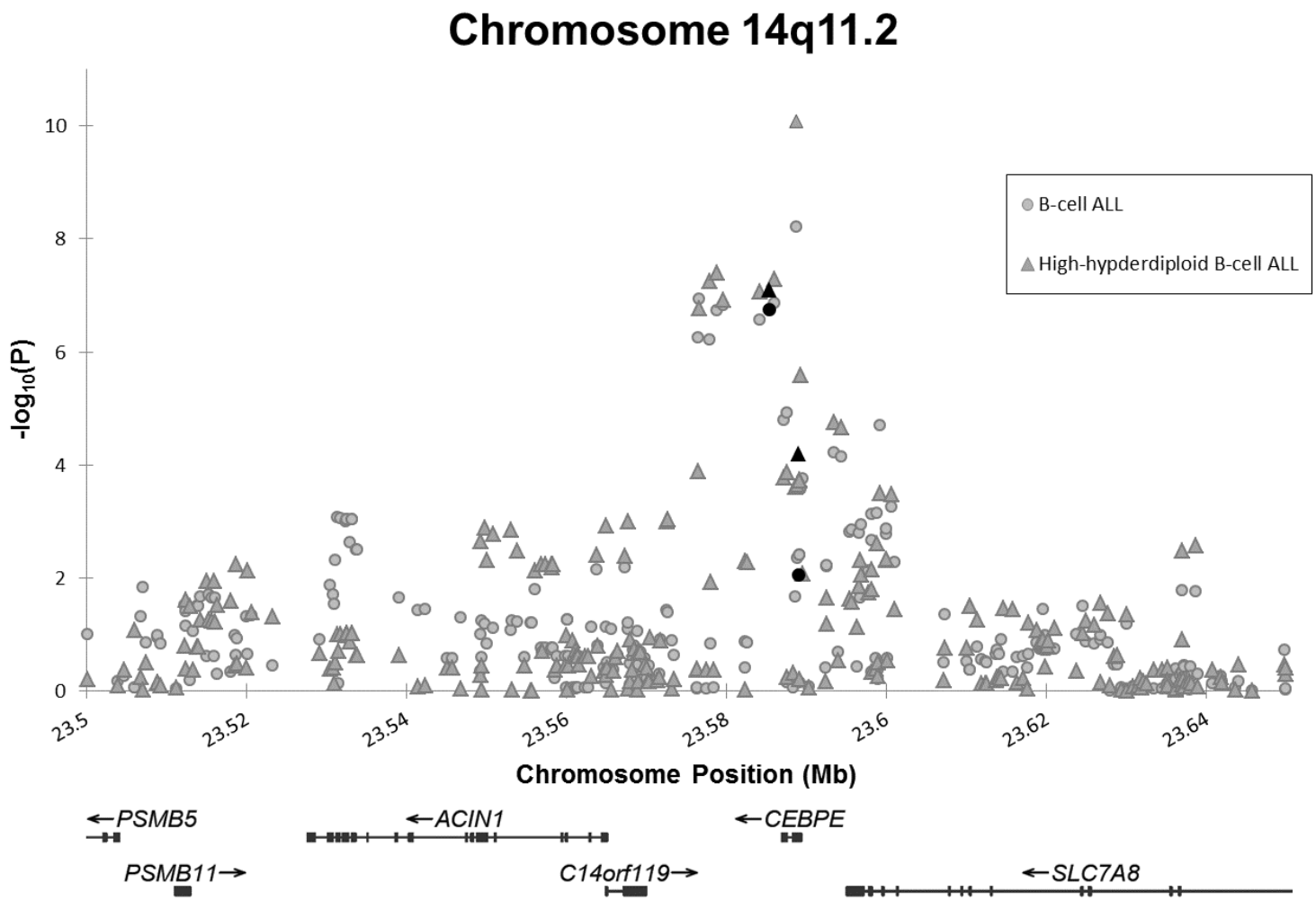
Supplementary Figure 1: Summary of Study Design and Analysis. The flowchart details flow subjects and of analyses through various stages of the study.



Supplementary Figure 2: Patterns of linkage disequilibrium in the 14q11.2 region (23.535-23.635Mb, GRCh37/hg19). (A) Haplotype structure in the European-ancestry COG and WTCCC samples (N_{cases}=980, $N_{controls}$=2624). (B) Haplotype structure in the CCLS Hispanic samples (N_{cases}=297, $N_{controls}$=454). Darker shading indicates higher R^2 values and greater correlation between SNPs.



Supplementary Figure 3: Association of SNPs in the *CEBPE* locus with B-ALL risk among Hispanic children in CCLS, by molecular subtype. Circles denote associations for B-ALL cases compared with controls (N=297 cases, 454 controls). Triangles denote associations for high-hyperdiploid B-ALL cases (97 cases) compared with controls. Two lead SNPs from previous GWAS are indicated in black.



Supplementary Figure 4: RNA expression of *CEBPE* and *IKZF1* in B-ALL cells compared to that during B-cell development, including hematopoietic progenitors (S1), pre-B-I cells (S2), pre-B-II cells (S3), immature B-cells (S4) and mature B-cells. Total RNA was isolated from diagnostic bone marrow samples of 81 B-ALL cases and from 8 fetal bone marrow specimens that were sorted by flow cytometry. RNA was measured with the Affymetrix GeneChip Human Gene 1.0 ST Array. (A) *CEBPE* expression decreases during B-cell development. (B) *IKZF1* expression increases during B-cell development.

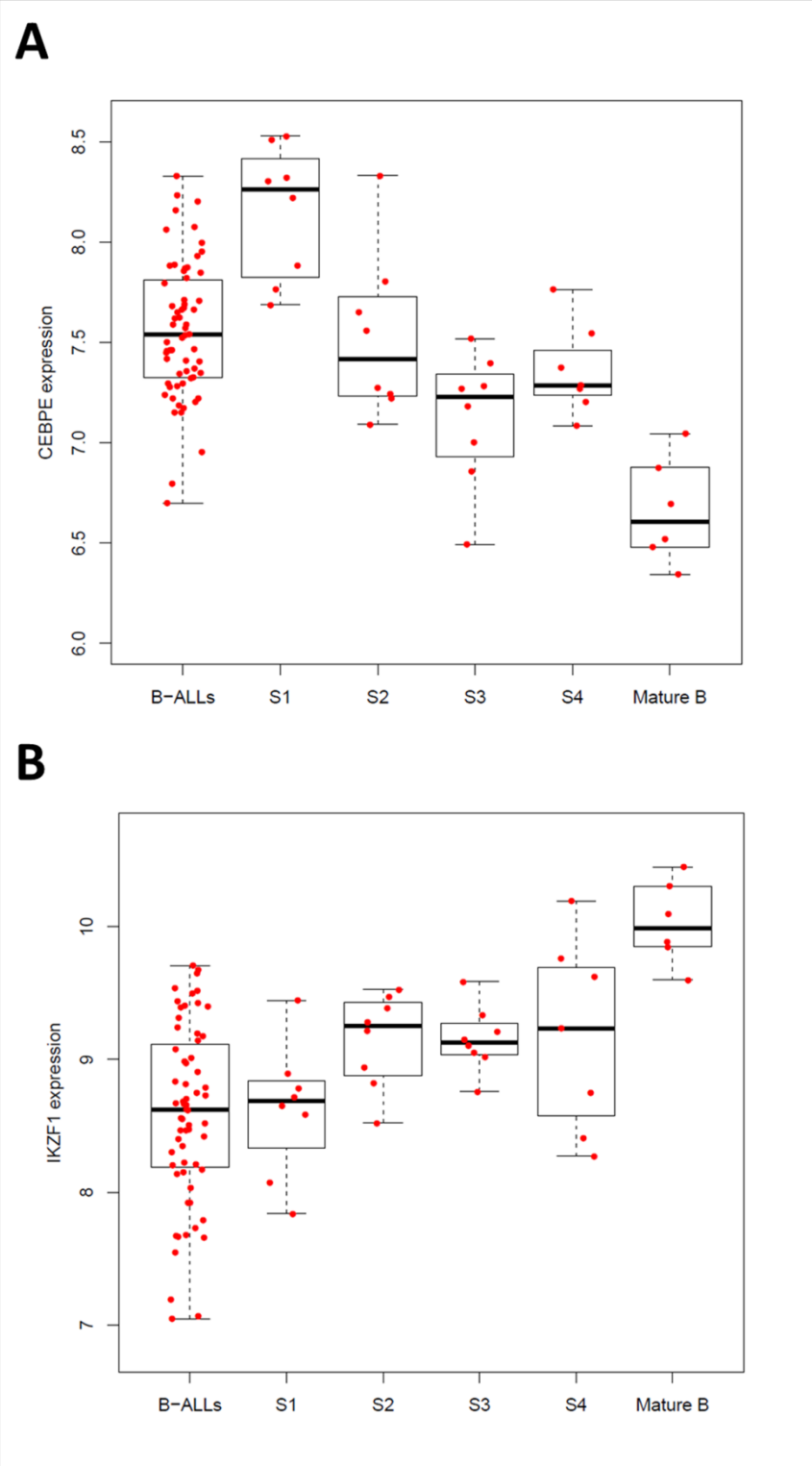


Table S1: Demographic and tumor characteristics of acute lymphoblastic leukemia cases and controls from the California Childhood Leukemia Study, the Children's Oncology Group, and the Wellcome Trust Case-Control Consortium appearing in association analyses.

	CCLS GWAS samples		COG/WTCCC GWAS samples[1]		CCLS Sequenom-only samples	
	Cases	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**
Sample Size	297	454	980	2624	271	436
Male (%)	156 (52.5)	240 (52.9)	503 (51.3)	1356 (51.7)	148 (54.6)	264 (60.1)
Hispanic (%)	297 (100.0)	454 (100.0)	0 (0)	0 (0)	32 (11.8)	35 (8.0)
European (%)	0 (0)	0 (0)	980 (100)	2624 (100)	145 (53.5)	350 (80.3)
B-cell (%)	297 (100.0)	NA	Unknown	NA	271 (100.0)	NA
High-hyperdiploid (%)	97 (32.7)	NA	Unknown	NA	84 (31.0)	NA
Tel-AML1 (%)	40 (13.5)	NA	Unknown	NA	62 (22.9)	NA
Somatic *IKZF1* deletion (%)	28 (9.4)	NA	Unknown	NA	Unknown	NA
Average Age (st. dev)	5.2 (3.3)	5.3 (3.4)	Unknown	Unknown	5.3 (3.3)	5.4 (3.4)

[1]Individual-level data, including age, tumor immunophenotype, and molecular subtype, were not available for COG cases or the WTCCC controls.

Supplementary Table 2. Top SNP associations in the *CEBPE* locus evaluated in European-ancestry ALL patients from COG and Hispanic ALL patients from CCLS, with functional annotation.

rsID[1]	Position	MAF[2]	P_{Meta}[3]	OR_{meta}[3]	$P_{Hispanics}$[4]	$OR_{Hispanics}$[4]	P_{Hyper}[5]	OR_{Hyper}[5]	RegulomeDB Score[6]	Conserved[7]	Promoter[8]	Enhancer[8]	DNAse Hypersensitivity[8]	Motifs altered[9]	*CEBPE* eQTL[10]
rs2239635	14:23588731	0.33	$5.10x10^{-11}$	1.45	$6.23x10^{-9}$	1.87	$8.21x10^{-11}$	2.94	2b	-	Yes	Yes	-	Ikaros, Zic1, Zic2	Yes
rs8007237	14:23577934	0.34	$5.48x10^{-11}$	1.43	$6.15x10^{-7}$	1.76	$5.58x10^{-8}$	2.58	6	-	-	-	-	LUN-1	Yes
rs56031127	14:23578903	0.24	$6.61x10^{-11}$	1.49	$1.87x10^{-7}$	1.86	$4.00x10^{-8}$	2.81	6	-	-	Yes	-	HNF1	-
rs4982731	14:23585333	0.31	$7.08x10^{-11}$	1.42	$1.74x10^{-7}$	1.75	$7.74x10^{-8}$	2.42	5	Yes	-	Yes	Yes	ETS1, MYC, POLR2A, ELF1	-
rs8015478	14:23586018	0.31	$1.03x10^{-10}$	1.41	$1.38x10^{-7}$	1.76	$5.06x10^{-8}$	2.46	5	-	-	Yes	Yes	-	-
rs2239630	14:23589349	0.48	$1.34x10^{-10}$	1.42	$2.66x10^{-4}$	1.50	$2.59x10^{-6}$	2.24	2b	-	Yes	Yes	Yes	SZF1-1, MZF1	Yes
rs10143875	14:23584265	0.31	$1.69x10^{-10}$	1.41	$2.67x10^{-7}$	1.73	$8.77x10^{-8}$	2.41	2b	-	-	-	Yes	BCL6B, STAT5A	-
rs8014370	14:23579678	0.30	$1.77x10^{-10}$	1.42	$1.51x10^{-7}$	1.81	$1.23x10^{-7}$	2.50	5	-	-	Yes	-	MEOX2, CDP, Cutl1	Yes
rs4982729	14:23576611	0.30	$2.75x10^{-10}$	1.42	$1.15x10^{-7}$	1.84	$1.67x10^{-7}$	2.53	No Data	-	-	Yes	-	-	-
rs9743723	14:23577198	0.41	$2.56x10^{-8}$	1.37	$2.89x10^{-7}$	1.86	$3.88x10^{-7}$	2.53	5	-	-	-	Yes	E2A, c-MYC	-
rs11625112	14:23596740	0.50	$3.32x10^{-8}$	1.34	$2.31x10^{-2}$	1.30	$4.86x10^{-3}$	1.65	5	-	-	-	Yes	-	-
rs2239633	14:23589057	0.46	$5.27x10^{-8}$	1.33	$8.87x10^{-3}$	1.33	$6.33x10^{-5}$	1.99	4	-	Yes	Yes	Yes	-	Yes
rs12434881	14:23588642	0.43	$6.26x10^{-8}$	1.33	$2.18x10^{-2}$	1.29	$2.46x10^{-4}$	1.89	2b	-	Yes	Yes	Yes	NF-kappaB	No Data
rs2239631	14:23589130	0.49	$8.93x10^{-8}$	1.33	$3.92x10^{-3}$	1.37	$1.88x10^{-4}$	1.88	4	-	Yes	Yes	Yes	-	Yes
rs2239632	14:23589127	0.49	$8.93x10^{-8}$	1.33	$3.92x10^{-3}$	1.37	$1.87x10^{-4}$	1.88	4	-	Yes	Yes	Yes	-	Yes
rs2239634	14:23588863	0.49	$9.33x10^{-8}$	1.33	$4.41x10^{-3}$	1.37	$2.24x10^{-4}$	1.86	4	Yes	Yes	Yes	Yes	-	Yes

[1]The 16 most significantly associated SNPs in analysis of the full case-control set appear here. There was a large fall-off in statistical significance at the 17th ranked SNP, rs17794251 ($P= 2.2x10^{-6}$). [2]MAF calculated in all cases and controls combined; [3]Totaling 1277 cases and 3078 controls; [4]Totaling 297 cases and 454 controls; [5]Totaling 97 cases and 454 controls; [6]RegulomeDB scores: 2a, TF binding+matched TF motif+matched DNase Footprint+DNase peak; 2b, TF binding+any motif+DNase Footprint+DNase peak; 3a, TF binding+any motif+DNase peak; 4, TF binding+DNase peak; 5, TF binding or DNase peak; 6, other. [7]Genomic Evolutionary Rate Profiling (GERP) scores >2 were taken to indicate evolutionary constraint/genetic conservation. [8]Promoter Histone Marks, Enhancer Histone Marks and DNase hypersensitivity sites appearing in blood and extracted from ENCODE2 data using HaploRegV3; [9]Altered motifs from footprinting and position-weight matrix analysis, extracted from RegulomeDB; [10]Genotype-Tissue Expression (GTEx) project association p-values <0.05, in whole-blood, were taken to indicate a significant eQTL.