



Genome-Wide Association Study of Susceptibility Loci for TCF3-PBX1 Acute Lymphoblastic Leukemia in Children

Shawn H. R. Lee, MD,^{1,2,3,†} Maoxiang Qian, PhD,^{4,*†} Wentao Yang, PhD,¹ Jonathan D. Diedrich, PhD,¹ Elizabeth Raetz, MD,⁵ Wenjian Yang , PhD,¹ Qian Dong, PhD,¹ Meenakshi Devidas, PhD,^{6,7} Deqing Pei, MS,⁸ Allen Yeoh, MD,^{2,3} Cheng Cheng, PhD,⁸ Ching-Hon Pui, MD,⁹ William E. Evans, PharmD,¹ Charles G. Mullighan, MD,¹⁰ Stephen P. Hunger, MD,¹¹ Daniel Savic, PhD,¹ Mary V. Relling, PharmD,¹ Mignon L. Loh, MD,¹² Jun J. Yang , PhD^{1,*}

¹Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN, USA; ²Khoo Teck Puat–National University Children's Medical Institute, National University Health System, Singapore; ³Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; ⁴Institute of Pediatrics and Department of Hematology and Oncology, Children's Hospital of Fudan University, Shanghai Key Laboratory of Medical Epigenetics, Institutes of Biomedical Sciences, Fudan University, Shanghai, China; ⁵Division of Pediatric Hematology and Oncology, New York University Langone Health, New York, NY, USA; ⁶Department of Global Pediatric Medicine, St. Jude Children's Research Hospital, Memphis, TN, USA; ⁷Department of Biostatistics, University of Florida, Gainesville, FL, USA; ⁸Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA; ⁹Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA; ¹⁰Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN, USA; ¹¹Department of Pediatrics and Center for Childhood Cancer Research, Children's Hospital of Philadelphia and the Perelman School of Medicine at The University of Pennsylvania, Philadelphia, PA, USA; and ¹²Department of Pediatrics, Benioff Children's Hospital and the Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA

[†]Authors contributed equally to this work.

*Correspondence to: Jun J. Yang, Department of Pharmaceutical Sciences, Hematologic Malignancies Program, Comprehensive Cancer Center, St. Jude Children's Research Hospital, 262 Danny Thomas Pl, MS313, Memphis, TN 38105 (e-mail: jun.yang@stjude.org) and Maoxiang Qian, Department of Hematology & Oncology, Institute of Pediatrics, Children's Hospital of Fudan University, and the Shanghai Key Laboratory of Medical Epigenetics, Institutes of Biomedical Sciences, Fudan University, 399 Wanyuan Rd, Shanghai, China (e-mail: mxqian@fudan.edu.cn).

Abstract

Acute lymphoblastic leukemia (ALL) is the most common cancer in children. TCF3-PBX1 fusion defines a common molecular subtype of ALL with unique clinical features, but the molecular basis of its inherited susceptibility is unknown. In a genome-wide association study of 1494 ALL cases and 2057 non-ALL controls, we identified a germline risk locus located in an intergenic region between BCL11A and PAPOLG: rs2665658, $P = 1.88 \times 10^{-8}$ for TCF3-PBX1 ALL vs non-ALL, and $P = 1.70 \times 10^{-8}$ for TCF3-PBX1 ALL vs other-ALL. The lead variant was validated in a replication cohort, and conditional analyses pointed to a single causal variant with subtype-specific effect. The risk variant is located in a regulatory DNA element uniquely activated in ALL cells with the TCF3-PBX1 fusion and may distally modulate the transcription of the adjacent gene REL. Our results expand the understanding of subtype-specific ALL susceptibility and highlight plausible interplay between germline variants and somatic genomic abnormalities in ALL pathogenesis.

Acute lymphoblastic leukemia (ALL) is the most common malignancy in children with biological subtypes defined by recurrent somatic genomic abnormalities (1,2). Germline genetic variations also contribute to ALL pathogenesis, with particular importance in inherited susceptibility. Genome-wide association studies (GWAS) have identified single nucleotide polymorphisms (SNPs) in IKZF1, CEBPE, ARID5B, PIP4K2A, and CDKN2A/CDKN2B that strongly influence ALL risk (3–8). More recent studies by us and others have also described susceptibility variants for specific ALL subtypes, for example, Philadelphia

chromosome-like (Ph-like) (5), hyperdiploid (3,6,7), ETV6-RUNX1 ALL (6), or T-ALL (9). However, the etiology of susceptibility to other ALL subtypes is largely unclear.

TCF3-PBX1 fusion, which arises from t(1; 19)(q23; p13.3), occurs in approximately 5% of children with B-cell ALL (1,10). This translocation gives rise to a chimeric protein that is directly related to malignant transformation of hematopoietic cells (11). TCF3-PBX1 fusion has a higher incidence in African Americans (12) and was historically associated with poorer prognosis (13) plausibly because of central nervous system

Received: May 15, 2020; Revised: August 14, 2020; Accepted: August 20, 2020

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com

relapse (10). However, recent studies on contemporary clinical trials report improved survival of this subtype with adequate risk-adapted therapy (13).

The inherited basis of TCF3-PBX1 ALL is unknown, and the exact molecular mechanism of how TCF3-PBX1 promotes leukemogenesis is incompletely understood. To this end, we conducted a GWAS of TCF3-PBX1 ALL to identify germline genetic variants related to susceptibility to this ALL subtype and explore their functional effects.

The ALL cases investigated in this GWAS comprised children treated on 5 frontline ALL studies: Children's Oncology Group P9904/P9905/P9906 (14), and St Jude Total Therapy XIIIIB (15) and XV (16). Patient demographics and their presenting features are as previously published (14–16). There were 5518 unrelated non-ALL participants from the Multi-Ethnic Study of Atherosclerosis study used as controls. Patients of genetically defined European ancestry ($n=1494$) formed the GWAS discovery cohort, with patients of non-European ancestry ($n=1035$) as the replication cohort. Detailed methods are described in the [Supplementary Materials and Methods](#) (available online).

In the discovery GWAS, we systematically evaluated the association of SNP genotype with the occurrence of childhood ALL with TCF3-PBX1 fusion in 40 cases compared with 2057 nonrelated ALL controls ([Supplementary Figure 1](#), available online). We imputed SNP genotypes genome-wide and examined 7528340 variants using an additive logistic regression model with population substructure included as covariates. A single locus in the intergenic region between the *BCL11A* and *PAPOLG* genes at 2p16.1 was identified above the genome-wide significance threshold ($P < 5 \times 10^{-8}$; [Figure 1, A](#)), with the strongest association signal at rs2665658 (odds ratio [OR] = 4.00, 95% confidence interval [CI] = 2.47 to 6.49, $P = 1.88 \times 10^{-8}$; [Table 1](#)). We then tested the above SNP in the replication cohort (82 cases with TCF3-PBX1 ALL and 3461 non-ALL controls). Overall, SNP rs2665658 again showed a statistically significant association (OR = 1.82, 95% CI = 1.31 to 2.53, $P = 3.85 \times 10^{-4}$), particularly in African Americans (OR = 2.21, 95% CI = 1.27 to 3.86, $P = .005$). The SNP was not associated with TCF3-PBX1 ALL in Hispanics, although the prevalence of this subtype varied by race, pointing to plausible ancestry-specific effects on ALL susceptibility ([Supplementary Table 1](#), available online). Quantile-quantile plots of logistic regression test for the GWAS indicated inflation only at the tail of the distribution ($\lambda = 0.95$), suggesting that population stratification was adequately controlled ([Supplementary Figure 2](#), available online).

In parallel, we also sought to evaluate the association of SNP genotype with TCF3-PBX1 fusion positivity within ALL cases of European ancestry by comparing allele frequency in TCF3-PBX1 ALL ($n=40$) vs ALL without TCF3-PBX1 fusion ($n=1454$). The *BCL11A*-*PAPOLG* locus was once again the only genome-wide significant hit (rs2665658, OR = 3.98, 95% CI = 2.46 to 6.44, $P = 1.70 \times 10^{-8}$; [Table 1](#), [Supplementary Figure 3](#), available online). This association was confirmed in the replication cohort of ALL cases of non-European ancestry ([Table 1](#)). Compared with non-ALL controls, the risk allele at this SNP was only over-represented in TCF3-PBX1 ALL but not in any other ALL subtypes ([Figure 1, B](#)), and its frequency did not differ by ALL clinical features ([Supplementary Figure 4](#), available online). There was also a notable increased risk allele frequency in *KMT2A*-rearranged B-ALL, but it did not reach statistical significance, plausibly because of the small numbers evaluated.

The 2p16.1 association peak was confined to the intergenic region between *BCL11A* and *PAPOLG* genes ([Figure 1, C](#)). Association test conditioning on rs2665658 did not reveal other

variants that independently contributed to the GWAS signal ([Supplementary Figure 5](#), available online), pointing to a single causal variant at this locus. To gain insight into the biological basis underlying the association signals at the risk locus, we explored potential regulatory elements within this genomic region. We first examined chromatin accessibility inferred from the ATAC-seq of different human hematopoietic cells and of a panel of human ALL cell lines. Across 8 variants in strong linkage disequilibrium with rs2665658, none overlapped with accessible chromatin regions identified in normal hematopoietic cells ([Supplementary Figure 6](#), available online). By contrast, multiple open chromatin regions were uniquely observed in ALL leukemic cells, among which we identified a prominent ATAC-seq peak specific to TCF3-PBX1 ALL and absent in other ALL subtypes ([Figure 1, D](#); [Supplementary Figure 7](#), available online). In fact, this open chromatin region is also marked by H3K27ac and juxtaposes with the ALL risk variants rs356994 and rs356995, suggesting possible effects of these genetic variations on enhancer activity ([Figure 1, D](#)). To explore this hypothesis, we scanned expression quantitative trait loci (eQTL) signals at this locus in the GTEx dataset (17), and the risk allele at rs2665658 was linked to lower expression of *REL* in whole blood cells ($P < .001$; [Supplementary Figure 8](#), available online). Examining the expression pattern of *REL* across B-ALL molecular subtypes using various datasets (18,19), we observed that *REL* expression was consistently downregulated in TCF3-PBX1 leukemia cells compared with other ALL subtypes ([Supplementary Figure 9](#), available online).

BCL11A is a transcriptional repressor that is essential in B-cell development and maintenance of stemness in stem cells (20,21). In acute myeloid leukemia, *BCL11A* was identified to accelerate leukemogenesis by cooperating with *MLL-AF9* (22). The effects of ALL risk variants on *REL* expression are also of interest because *REL* is a critical component of NF κ B signaling and implicated in B-cell development (23). A study of diffuse large B-cell lymphomas has suggested that *REL* and *BCL11A* are frequently coamplified together, causing tumorigenesis (24). Also, regulatory variants of *REL* are linked to susceptibility to Hodgkin lymphoma (25). It is unclear why (and how) *BCL11A* and *REL* are specifically associated with TCF3-PBX1 ALL but no other subtypes. We postulate that modest perturbation of *BCL11A* and/or *REL* transcription may corroborate with TCF3-PBX1 to initiate ALL. One limitation of our study is that although this is the largest GWAS of TCF3-PBX1 ALL in children thus far, it is still not yet sufficiently powered to draw definitive conclusions on ancestry-specific effects on this subtype. Future larger cohorts, further mechanistic studies, and functional investigations are warranted to identify causal variants and fully elucidate the process by which these germline variants influence the development of TCF3-PBX1 ALL.

Funding

This work is supported by US National Institutes of Health Grants No. CA21765, CA98543, CA114766, CA98413, CA180886, CA180899, GM92666, GM115279, CA234490, CA197695, and GM097119; and the American Lebanese Syrian Associated Charities. SHRL is supported by the National Medical Research Council Singapore Research Training Fellowship (003/008–258). MQ is supported by the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning and the National Natural Science Foundation of China (81973997).

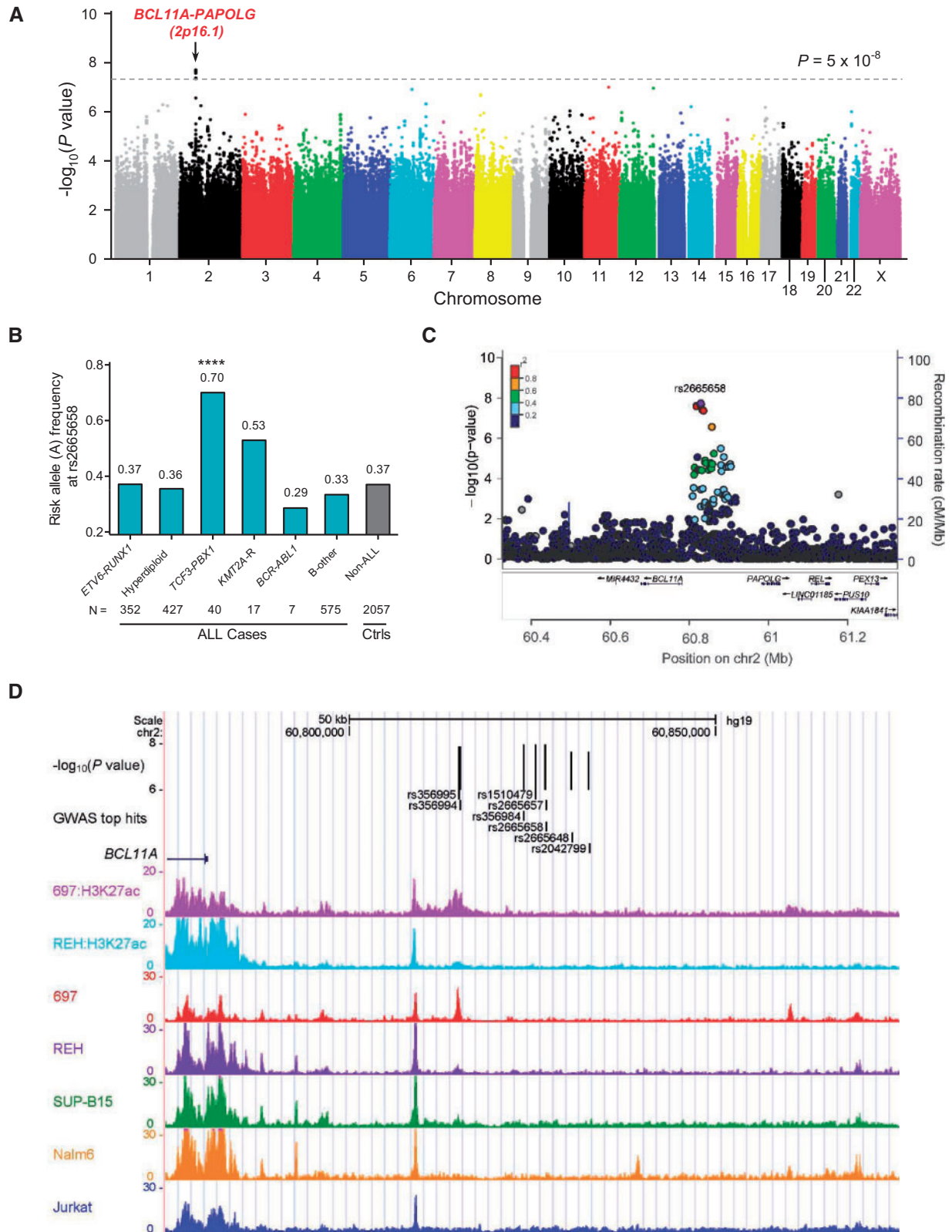


Figure 1. Genome-wide association studies (GWAS) of TCF3-PBX1 acute lymphoblastic leukemia (ALL) susceptibility and characteristics and functional annotation of genomic variants at the *BCL11A*-*PAPOLG* locus. **(A)** The association between genotype and ALL was evaluated by using a logistic regression model for 7 528 340 genotyped/imputed single nucleotide polymorphisms (SNPs) in 40 TCF3-PBX1 ALL cases and 2057 unrelated non-ALL controls of European ancestry. Association P values ($-\log_{10}(P$ value), y axis) were plotted against the respective chromosomal position of each SNP (x axis). The dashed horizontal line indicates the genome-wide significance threshold ($P < 5 \times 10^{-8}$). The *BCL11A*-*PAPOLG* locus is indicated at 2p16.1. **(B)** Risk allele frequency of SNP rs2665658 among ALL subtypes of patients of European ancestry in the discovery GWAS. *KMT2A-R* = *KMT2A*-rearranged; CtrlS = controls. Logistic regression test with rs2665658 genotype (eg, ALL with vs without TCF3-PBX1),

Table 1. Genome-wide significant association and replication of a novel susceptibility variant in BCL11A-PAPOLG locus for TCF3-PBX1 ALL

SNP	CHR	Position ^a	Alleles ^b	Cohort ^c	Risk allele frequency (Participants, No.)			TCF3-PBX1 ALL vs non-ALL		TCF3-PBX1 ALL vs other ALL	
					TCF3-PBX1 ALL	Other ALL	Non-ALL	OR (95% CI) ^d	P ^e	OR (95% CI) ^d	P ^e
rs2665658	2	60826802	C/A	Discovery European American Validation African American Hispanics Others Combined	0.70 (40)	0.35 (1454)	0.37 (2057)	4.00 (2.47 to 6.49)	1.88 x 10 ⁻⁸	3.98 (2.46 to 6.44)	1.70 x 10 ⁻⁸
					0.66 (28)	0.42 (186)	0.47 (1380)	2.21 (1.27 to 3.86)	.005	2.49 (1.39 to 4.47)	.002
					0.37 (30)	0.29 (424)	0.32 (682)	1.27 (0.74 to 2.20)	.39	1.52 (0.88 to 2.63)	.13
					0.50 (24)	0.30 (343)	0.38 (1399)	2.44 (1.26 to 4.75)	.009	2.88 (1.45 to 5.73)	.003
					0.51 (82)	0.32 (953)	0.40 (3461)	1.82 (1.31 to 2.53)	3.85 x 10 ⁻⁴	2.12 (1.52 to 2.96)	1.06 x 10 ⁻⁵

^aChromosomal locations are based on GRCh37/hg19. ALL = acute lymphoblastic leukemia; CHR = chromosome; CI = confidence interval; OR = odds ratio; SNP = single nucleotide polymorphism.

^bThe A allele is the risk allele for ALL.

^cDiscovery and replication cohorts consist of individuals of European and non-European ancestry, respectively.

^dAssociation of SNP genotype with TCF3-PBX1 ALL was evaluated by comparing allele frequencies in TCF3-PBX1 ALL and non-ALL and in TCF3-PBX1 ALL and other ALL, after adjusting for genetic structure. Odds ratio values represent the increase in risk of developing TCF3-PBX1 ALL for each copy of the risk allele compared with participants who do not carry the risk allele.

^eP values were estimated by the additive logistic regression test adjusting for population structure.

Figure 1. Continued

****P < .001. (C) Univariate analysis and regional plot of genotype association at the BCL11A-PAPOLG locus. The lead SNP from the discovery GWAS (rs2665658, OR = 4.00, 95% CI = 2.47 to 6.49, P = 1.88 × 10⁻⁸) is indicated by its rsID and purple diamond shape, and other SNPs are displayed by color showing their extent of linkage disequilibrium with this lead SNP. Recombination rate, chromosomal position (hg19), and nearby genes (RefSeq) are indicated. P values were calculated by additive logistic regression test in univariate analysis. (D) Functional annotation of genomic variants at the BCL11A-PAPOLG locus. The default tracks including genomic positions and scale for the human genome assembly February 2009 (GRCh37/hg19) are shown on the top. The log-transformed P values for SNPs tested for association with TCF3-PBX1 ALL are shown in the bed graph. The gene structure, histone modification (H3K4ac), and ATAC-seq signals available in different ALL cell lines (697 [TCF3-PBX1], REH [ETV6-RUNX1], SUP-B15 [BRC-ABL1], Nalm6 [DUX4-IGH], Jurkat [T-ALL]) are also included.

SPH is the Jeffrey E. Perelman Distinguished Chair in Pediatrics at The Children's Hospital of Philadelphia.

Notes

Role of the funder: The study sponsors were not directly involved in the design of the study, the collection, analysis, and interpretation of the data, the writing of the manuscript, or the decision to submit the manuscript.

Disclosures: CGM receives research funding from Abbvie and Pfizer, and receives consulting fees for Illumina. All other authors certify that there is no actual or potential conflict of interest in relation to this article.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Acknowledgments: The authors thank the patients and parents who participated in the clinical protocols included in this study, and the clinicians and research staff at participating institutions.

Role of the authors: JJY is the principal investigator of this study, has full access to all of the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis; MQ, WTY, WJY, JDD, QD, DP, and CC performed data analysis; SHRL, MQ, and JJY wrote the manuscript; ER, MD, AY, C-HP, WEE, CGM, SPH, DS, MVR, and MLL contributed reagents, materials, and/or data; SHRL, MQ, WTY, and JJY interpreted the data and the research findings; and all of the coauthors reviewed the manuscript.

Data availability

The GWAS data are deposited in the NIH dbGAP (<https://www.ncbi.nlm.nih.gov/gap/>) under phs000638.v1.p1 and phs000637.v1.p1.

References

- Pui CH, Relling MV, Downing JR. Acute lymphoblastic leukemia. *N Engl J Med*. 2004;350(15):1535-1548.
- Mullighan CG. Genomic characterization of childhood acute lymphoblastic leukemia. *Semin Hematol*. 2013;50(4):314-324.
- Trevino LR, Yang W, French D, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet*. 2009;41(9):1001-1005.
- Xu H, Yang W, Perez-Andreu V, et al. Novel susceptibility variants at 10p12.31-12.2 for childhood acute lymphoblastic leukemia in ethnically diverse populations. *J Natl Cancer Inst*. 2013;105(10):733-742.

5. Perez-Andreu V, Roberts KG, Xu H, et al. A genome-wide association study of susceptibility to acute lymphoblastic leukemia in adolescents and young adults. *Blood*. 2015;125(4):680–686.
6. Vijayakrishnan J, Studd J, Broderick P, et al.; The PRACTICAL Consortium. Genome-wide association study identifies susceptibility loci for B-cell childhood acute lymphoblastic leukemia. *Nat Commun*. 2018;9(1):1340.
7. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nat Genet*. 2009;41(9):1006–1010.
8. Churchman ML, Qian M, Te Kronnie G, et al. Germline genetic IKZF1 variation and predisposition to childhood acute lymphoblastic leukemia. *Cancer Cell*. 2018;33(5):937–948 e8.
9. Qian M, Zhao X, Devidas M, et al. Genome-wide association study of susceptibility loci for T-cell acute lymphoblastic leukemia in children. *J Natl Cancer Inst*. 2019;111(12):1350–1357.
10. Jeha S, Pei D, Raimondi SC, et al. Increased risk for CNS relapse in pre-B cell leukemia with the t(1; 19)/TCF3-PBX1. *Leukemia*. 2009;23(8):1406–1409.
11. Aspland SE, Bendall HH, Murre C. The role of E2A-PBX1 in leukemogenesis. *Oncogene*. 2001;20(40):5708–5717.
12. Pui CH, Sandlund JT, Pei D, et al. Results of therapy for acute lymphoblastic leukemia in black and white children. *JAMA*. 2003;290(15):2001–2007.
13. Felice MS, Gallego MS, Alonso CN, et al. Prognostic impact of t(1; 19)/ TCF3-PBX1 in childhood acute lymphoblastic leukemia in the context of Berlin-Frankfurt-Munster-based protocols. *Leuk Lymphoma*. 2011;52(7):1215–1221.
14. Borowitz MJ, Devidas M, Hunger SP, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood*. 2008;111(12):5477–5485.
15. Pui CH, Sandlund JT, Pei D, et al. Improved outcome for children with acute lymphoblastic leukemia: results of Total Therapy Study XIIIIB at St Jude Children's Research Hospital. *Blood*. 2004;104(9):2690–2696.
16. Pui CH, Campana D, Pei D, et al. Treating childhood acute lymphoblastic leukemia without cranial irradiation. *N Engl J Med*. 2009;360(26):2730–2741.
17. Battle A, Brown CD, Engelhardt BE, et al. eQTL Manuscript Working Group. Genetic effects on gene expression across human tissues. *Nature*. 2017; 550(7675):204–213.
18. Qian M, Zhang H, Kham SK, et al. Whole-transcriptome sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant genomic abnormalities of EP300 and CREBBP. *Genome Res*. 2017;27(2):185–195.
19. Gu Z, Churchman ML, Roberts KG, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet*. 2019;51(2):296–307.
20. Ippolito GC, Dekker JD, Wang YH, et al. Dendritic cell fate is determined by BCL11A. *Proc Natl Acad Sci USA*. 2014;111(11):E998–E1006.
21. Liu P, Keller JR, Ortiz M, et al. Bcl11a is essential for normal lymphoid development. *Nat Immunol*. 2003;4(6):525–532.
22. Bergerson RJ, Collier LS, Sarver AL, et al. An insertional mutagenesis screen identifies genes that cooperate with Mll-AF9 in a murine leukemogenesis model. *Blood*. 2012;119(19):4512–4523.
23. Gilmore TD, Kalaitzidis D, Liang MC, et al. The c-Rel transcription factor and B-cell proliferation: a deal with the devil. *Oncogene*. 2004;23(13):2275–2286.
24. Satterwhite E, Sonoki T, Willis TG, et al. The BCL11 gene family: involvement of BCL11A in lymphoid malignancies. *Blood*. 2001;98(12):3413–3420.
25. Enciso-Mora V, Broderick P, Ma Y, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). *Nat Genet*. 2010;42(12):1126–1130.