blood advances

TO THE EDITOR:

Epigenome-wide association study of acute lymphoblastic leukemia in children with Down syndrome

Shaobo Li,[1] Pagna Sok,[2] Keren Xu,[1] Ivo S. Muskens,[1] Natalina Elliott,[3] Swe Swe Myint,[1] Priyatama Pandey,[1] Helen M. Hansen,[4] Libby M. Morimoto,[5] Alice Y. Kang,[5] Catherine Metayer,[5] Xiaomei Ma,[6] Beth A. Mueller,[7,8] Anindita Roy,[3] Irene Roberts,[3] Karen R. Rabin,[2] Austin L. Brown,[2] Philip J. Lupo,[2] Joseph L. Wiemels,[1] and Adam J. de Smith[1]

[1]Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA; [2]Department of Pediatrics, Section of Hematology-Oncology, Baylor College of Medicine, Houston, TX; [3]Department of Paediatrics and MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford University and BRC Blood Theme, NIHR Oxford Biomedical Centre, Oxford, UK; [4]Department of Neurological Surgery, University of California San Francisco, San Francisco, CA; [5]School of Public Health, University of California, Berkeley, Berkeley, CA; [6]Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT; [7]Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; and [8]Department of Epidemiology, University of Washington, Seattle, WA

Down syndrome (DS) is caused by constitutional trisomy of chromosome 21 and is associated with an up to 30-fold increased risk of acute lymphoblastic leukemia (ALL).[1,2] While DS is associated with alterations in epigenetic markers, including DNA methylation, and gene expression,[3-6] these mechanisms have not been fully explored in relation to DS-ALL etiology.[7] Because the epigenome is sensitive to genetic and environmental influences during fetal development and can be leveraged to characterize blood cell proportions,[8] we sought to evaluate the role of the neonatal methylome in children with DS on subsequent ALL risk.

Our epigenome-wide association study (EWAS) included 126 DS-ALL cases and 198 DS control subjects from the International Study of Down Syndrome Acute Leukemia[7,9] in the Discovery dataset and 24 cases and 24 control subjects from the Michigan-based DS-ALL study[7] in the Replication group. DNA was isolated from neonatal dried bloodspots, bisulfite-converted and assayed using Illumina EPIC methylation arrays. Further details on study subjects, quality control and processing of methylation array data, and statistical analyses are included in the supplemental Methods. The Institutional Review Boards of each participating site approved the study, which was conducted according to the Declaration of Helsinki.

Demographic and birth-related data are summarized in Table 1. Unsupervised hierarchical clustering did not differentiate DS-ALL cases from DS control subjects but did demonstrate variation in blood cell proportions, determined by reference-based deconvolution using the Identifying Optimal Libraries algorithm,[10] and identified a subset of DS newborns with high nucleated red blood cell proportions, as previously shown[6] (supplemental Figure 1).

Deconvolution of blood cell proportions in the Discovery study revealed a significant increase in B-cell proportions at birth in DS-ALL cases (mean, 0.0128) compared with DS control subjects (mean, 0.00826; $P = 8.58 \times 10^{-4}$), a difference which was also observed in the Replication study ($P = .03$) and meta-analysis (effect $\text{size}_{\text{meta}} = 0.0056$; $P_{\text{meta}} = 1.69 \times 10^{-4}$; $P_{\text{het}} = .15$) (supplemental Figure 2 and Table 2). Among cell types, B cells showed the greatest proportional difference between cases and control subjects in both Discovery (+55.57% in DS-ALL) and Replication (+22.23%) studies

Submitted 20 January 2022; accepted 10 May 2022; prepublished online on *Blood Advances* First Edition 19 May 2022; final version published online 18 July 2022. DOI 10.1182/bloodadvances.2022007098.

This study used biospecimens from the California Biobank Program. Any uploading of genomic data (including genome-wide DNA methylation data) and/or sharing of these biospecimens or individual data derived from these biospecimens have been determined to violate the statutory scheme of the California Health and Safety Code Sections 124980(j), 124991(b), (g), (h), and 103850 (a) and (d), which protect the confidential nature of biospecimens and individual data derived from biospecimens. The individual-level data derived from these biospecimens that support the findings of this study are available from the corresponding author upon request (adam.desmith@med.usc.edu) and with permission from the California Biobank Program and Michigan Newborn Screening Program. Data for deconvoluted blood cell proportions and available covariates in the Discovery Study subjects are included in the supplemental Dataset.

The full-text version of this article contains a data supplement.

Table 1. Demographic and birth characteristics of DS-ALL cases and DS control subjects

	DS-ALL Discovery Study			DS-ALL Replication Study		
	DS control subjects (n = 198), n (%)	DS-ALL (n = 126), n (%)	*P* value	DS control subjects (n = 24), n (%)	DS-ALL (n = 24), n (%)	*P* value
Sex						
Male	91 (46.0)	84 (66.7)	–	14 (58.3)	13 (54.2)	–
Female	107 (54.0)	42 (33.3)	.00037*	10 (41.7)	11 (45.8)	.771*
Race/ethnicity						
Asian	10 (5.1)	2 (1.6)	–	1 (4.2)	1 (4.2)	–
Latino	96 (48.5)	86 (68.3)	–	3 (12.5)	2 (8.3)	–
Non-Latino White	54 (27.3)	32 (25.4)	–	15 (62.5)	20 (83.3)	–
Non-Latino Black	10 (5.1)	2 (1.6)	–	5 (20.8)	1 (4.2)	–
Other	28 (14.1)	4 (3.2)	.00037*	0	0	.287*
Missing	0	0		0	0	
Age at DS-ALL diagnosis (y)						
Median (range)	–	4.0 (0-14.6)	–	–	<4.0, n = 13;≥4.0, n = 11†	–
Blood collection age (d)						
Mean (SD)	2.47 (2.03)	2.03 (2.14)	.068	N/A	N/A	–
Median (range)	1.71 (0.17-15.25)	1.46 (0-18.96)	–	N/A	N/A	–
Missing	3 (1.5)	9 (7.1)	–	24 (100.0)	24 (100.0)	–
Gestational age (wk)						
Mean (SD)	38.10 (2.33)	38.22 (2.83)	.67‡	N/A	N/A	–
Median (range)	38.29 (26.42-44.71)	38.43 (25.57-44.43)	–	N/A	N/A	–
Preterm (<37)	41 (22.7)	29 (24.0)	.78*	N/A	N/A	–
Missing	17 (8.6)	5 (4.0)	–	24 (100.0)	24 (100.0)	–
Birthweight (kg)						
Mean (SD)	3.00 (0.74)	3.08 (0.60)	.31‡	N/A	N/A	–
Median (range)	3.02 (0.81-8.65)	3.12 (0.94-4.58)	–	N/A	N/A	–
Missing	4 (2.0)	1 (0.8)	–	24 (100.0)	24 (100.0)	–

**P* values calculated using a 2-tailed Fisher's exact test.
†Age-at-diagnosis only available in categories for DS-ALL cases in the Replication Study.
‡*P* values were calculated using a 2-tailed *t* test.

(supplemental Table 1). An independent deconvolution method, Epigenetic Dissection of Intra-Sample-Heterogeneity (EpiDISH),[11] confirmed the increased B-cell proportions in DS-ALL cases in both studies ($P_{meta} = 1.67 \times 10^{-4}$) (supplemental Table 2).

In analyses stratified by self-reported race and ethnicity in the Discovery study, increased neonatal B-cell proportions showed a stronger effect in Latinos (effect size = 0.0058; $P = 6.15 \times 10^{-3}$) than in non-Latino Whites (effect size = 0.0046; $P = .098$), although this difference was not statistically significant ($P_{het} = .74$) (supplemental Table 3).

We performed several sensitivity analyses in the Discovery study to assess potential confounders of the increased B-cell proportions in DS-ALL. First, in subjects with available birth-variable data, we adjusted the regression model for gestational age, birth weight, and bloodspot collection age, and the difference in B-cell proportions between DS-ALL cases (n = 116) and DS control subjects (n = 173) remained significant (effect size = 0.0059; $P = 3.38 \times 10^{-4}$).

Next, in Latino and non-Latino White subjects with single nucleotide polymorphism (SNP) genotype data (117 cases, 130 control subjects), we assessed whether SNPs associated with DS-ALL risk in *ARID5B* (rs7089424), *IKZF1* (rs11978267), *CDKN2A* (rs3731249), or *GATA3* (rs3824662)[7] may confound the association with B-cell proportions, as these loci were previously associated with variation in white blood cell traits.[14] We included the genotypes of these 4 SNPs in the regression model one at a time and also all together, and the significantly increased B-cell proportions in DS-ALL cases remained, with similar effect sizes in Latinos and non-Latino Whites (supplemental Table 4).

Finally, we removed *GATA1* mutation-positive control subjects (n = 30 of 184 tested, *see* supplemental Methods), and the difference in B-cell proportions remained significant (effect size = 0.0043; $P = 9.02 \times 10^{-3}$).

In the Discovery study EWAS of DS-ALL (126 cases, 198 control subjects), the genomic inflation factor was 1.11 after correction with BACON, a Bayesian method to control bias and inflation in

Table 2. Deconvoluted blood cell proportions in DS-ALL cases vs DS controls

	Discovery Study (126 cases, 198 control subjects)			Replication Study (24 cases, 24 control subjects)			Meta-analysis				
Cell type	Effect estimate*	Standard error*	*P* value*	Effect estimate*	Standard error*	*P* value*	Effect estimate†	Standard error*	P_{meta}†	P_{het}†	Direction
CD4 T cell	0.0036	0.0055	.51	−0.0147	0.0136	.29	0.0011	0.0051	.83	.21	−+
CD8 T cell	0.0071	0.0030	**.016**	0.0168	0.0102	.11	0.0079	0.0028	**.0055**	.36	++
B cell	0.0051	0.0015	**8.58×10^{-4}**	0.0152	0.0069	**.03**	0.0056	0.0015	**1.69×10^{-4}**	.15	++
NK cells	0.0028	0.0024	.24	0.0048	0.0079	.55	0.0030	0.0023	.19	.81	++
Granulocyte	0.0076	0.0178	.67	−0.0482	0.0376	.21	−0.0026	0.0161	.87	.18	−+
Monocyte	0.0010	0.0040	.81	−0.0003	0.0107	.98	0.0008	0.0038	.83	.91	−+
nRBC	−0.0301	0.0228	.19	0.0163	0.0357	.65	−0.0166	0.0192	.39	.27	+−

NK, natural killer; nRBC, nucleated red blood cells.
$P < .05$ highlighted in bold.
*P values, coefficients, and standard errors calculated using linear regression, testing each blood cell type separately as the dependent variable, with DS-ALL status as the independent variable, and including sex, batch, and ancestry-related principal components from EPISTRUCTURE[12] (n = 10 for Discovery study; n = 3 for Replication study) as covariates. P values were not adjusted for multiple comparisons.
†Meta-analysis performed using METAL.[13]

EWAS.[15] There were 38 significant differentially methylated probes (DMPs) after false discovery rate (FDR) correction and 10 epigenome-wide–significant DMPs after Bonferroni correction ($P < 7.95 \times 10^{-8}$) (supplemental Figure 3; supplemental Table 5). Pathway enrichment analysis of FDR-significant DMPs revealed significant enrichment of 21 gene ontology pathways (supplemental Table 6). The top DS-ALL–associated CpG (cg27347265; $P = 2.90 \times 10^{-12}$) was located in a putative regulatory region of the B-cell transcription factor gene *EBF1* (supplemental Figure 4; supplemental Table 5). For all 10 Bonferroni-significant DMPs, the case-control methylation β-value difference was <0.02, and none were significant in the Replication study at $P < .05$, although 6 out of 10 had consistent directions of effect.

We identified 31 significant differentially methylated regions (DMRs) associated with DS-ALL in the Discovery study (supplemental Table 7). Although none of the DMRs were statistically significant in the Replication study, 4 of 31 contained significant ($P < .05$) differentially methylated CpGs with the same direction of methylation changes as the Discovery study (supplemental Table 7).

In summary, an increase in the neonatal proportion of B cells was associated with DS-ALL risk, a finding that persisted after adjustment for potential confounding factors and was consistent between 2 independent case-control datasets. DS is associated with reduced fetal B-cell production[16,17] and reduced numbers of B cells in fetal life[16,18] and childhood.[19,20] We previously observed lower B-cell proportions in newborns with DS than in newborns without DS using reference-based cell-type deconvolution analysis.[6] Results from the current study support that, in the context of DS, children with greater B-cell proportions at birth have an increased risk of developing DS-ALL. A genetic predisposition to overproducing lymphocytes was recently associated with increased ALL risk in the non-DS population.[14] Further studies are required to understand the mechanisms underlying the association between increased B cells and ALL development in children with and without DS, but these may involve effects on the proliferation of preleukemic clones and generation of leukemia-forming mutations, as well as potential impacts on immune function and response to infections.[14,21]

We did not find strong evidence for differences in DNA methylation at birth that might predict subsequent DS-ALL risk, although the Replication dataset was underpowered to reproduce significance for the small differences found between cases and controls in the Discovery study. The significant *EBF1* DMP is intriguing given that this gene is frequently deleted in ALL.[22] Investigation of DNA methylation differences in sorted cell populations is required to determine cell-specific epigenetic changes associated with DS-ALL risk.

A strength of our study was the use of newborn DBS, collected before disease onset and, therefore, any case-control differences should not be confounded by the presence of leukemia cells; indeed, in the Discovery study, only 1 DS-ALL case was diagnosed <1 year of age and the B-cell case-control difference was significant both when restricted to cases with age-at-diagnosis ≤4 years (n = 64; effect size = 0.0039; $P = .034$) or >4 years of age (n = 62; effect size = 0.0061; P value = 1.01×10^{-3}).

A study limitation includes the use of a blood cell proportion deconvolution methodology developed in euploid individuals,[10] although the same approach confirmed known differences in blood cell proportions associated with DS.[6] Nonetheless, the increased B-cell proportion in DS-ALL cases requires confirmation using blood cell count measures in newborns. Another limitation was that sequencing data for somatic *GATA1* mutations, which cause transient abnormal myelopoiesis,[23] were only available for DS controls in the Discovery study; however, removal of *GATA1* mutation-positive control subjects had minimal effect on the B-cell association.

Future studies are needed to understand the role of blood cell trait variation in DS-ALL etiology and examine increased neonatal B cells as a potential risk factor for ALL in the non-DS population.

Acknowledgments: The authors thank Robin Cooley and Steve Graham (Genetic Disease Screening Program, CDPH) for their assistance and expertise in the procurement and management of DBS specimens. The authors also thank Hong Quach and Diana

Quach at the UC Berkeley QB3 Genetic Epidemiology and Genomics Laboratory for their support in preparing and processing samples for genome-wide DNA methylation arrays. The authors additionally thank the families for their participation in the California Childhood Leukemia Study (formerly known as the Northern California Childhood Leukemia Study). For recruitment of subjects enrolled in the California Childhood Leukemia Study, the authors gratefully acknowledge the clinical investigators at the following collaborating hospitals: University of California, Davis Medical Center (Jonathan Ducore), University of California, San Francisco (Mignon Loh and Katherine Matthay), Children's Hospital of Central California (Vonda Crouse), Lucile Packard Children's Hospital (Gary Dahl), Children's Hospital Oakland (James Feusner and Carla Golden), Kaiser Permanente Roseville (formerly Sacramento) (Kent Jolly and Vincent Kiley), Kaiser Permanente Santa Clara (Carolyn Russo, Alan Wong, and Denah Taggart), Kaiser Permanente San Francisco (Kenneth Leung), Kaiser Permanente Oakland (Daniel Kronish and Stacy Month), California Pacific Medical Center (Louise Lo), Cedars-Sinai Medical Center (Fataneh Majlessipour), Children's Hospital Los Angeles (Cecilia Fu), Children's Hospital Orange County (Leonard Sender), Kaiser Permanente Los Angeles (Robert Cooper), Miller Children's Hospital Long Beach (Amanda Termuhlen), University of California, San Diego Rady Children's Hospital (William Roberts), and University of California, Los Angeles Mattel Children's Hospital (Theodore Moore).

This work was supported by an Alex's Lemonade Stand Foundation "A" Award (A.J.d.S.), a National Institutes of Health (NIH) National Cancer Institute (NCI) Grant (R01CA249867 [K.R.R., P.J.L., and A.J.d.S.]), an NIH NCI Administrative Supplement grant (3R01CA175737-05S1 [J.L.W., X.M., and A.J.d.S.]), and National Institute for Environmental Health Sciences (NIEHS) grants (R01ES009137, P42ES004705, and R24ES028524). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. A subset of biospecimens and/or data used in this study were obtained from the California Biobank Program at the California Department of Public Health (CDPH), SIS request numbers 572 and 26, in accordance with Section 6555(b), 17 CCR. The CDPH is not responsible for the results or conclusions drawn by the authors of this publication.

The collection of cancer incidence data used in the CCRLP study was supported by the California Department of Public Health pursuant to California Health and Safety Code Section 103885, Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries, under cooperative agreement 5NU58DP003862-04/DP003862, the National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) Program under contract HHSN261201000140C awarded to the Cancer Prevention Institute of California, contract HHSN261201000035C awarded to the University of Southern California, and contract HHSN261201000034C awarded to the Public Health Institute. The ideas and opinions expressed herein are those of the author(s) and do not necessarily reflect the opinions of the State of California, the Department of Public Health, the National Institutes of Health, and the Centers for Disease Control and Prevention or their Contractors and Subcontractors. This study used birth data obtained from the State of California Center for Health Statistics and Informatics. The California Department of Public Health is not responsible for the analyses, interpretations, or conclusions drawn by the authors regarding the birth data used in this publication.

Contribution: A.J.d.S., J.L.W., P.J.L., and A.L.B. designed and supervised this study; S.L., P.S., K.X., I.S.M., N.E., P.P., and A.J.d.S. analyzed the data; N.E., S.S.M., and H.M.H. performed experiments; L.M.M., A.Y.K., C.M., X.M., B.A.M., A.R., I.R., and K.R.R. provided resources; S.L., P.S., K.R.R., A.L.B., P.J.L., J.L.W., and A.J.d.S. prepared the manuscript; and all authors edited and approved the paper.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profiles: S.L., 0000-0002-0544-5338; K.X., 0000-0003-4615-2459; S.S.M., 0000-0002-2430-088X; H.M.H., 0000-0001-5456-4985; L.M.M., 0000-0002-1509-739X; K.R.R., 0000-0002-4081-8195; A.L.B., 0000-0001-5802-5073; P.J.L., 0000-0003-0978-5863; A.J.d.S., 0000-0003-4880-7543.

Correspondence: Adam J. de Smith, USC Norris Comprehensive Cancer Center, 1450 Biggy St, NRT-1509H, Los Angeles, CA 90033; e-mail: adam.desmith@med.usc.edu.

References

1. Marlow EC, Ducore J, Kwan ML, et al. Leukemia risk in a cohort of 3.9 million children with and without Down syndrome. *J Pediatr.* 2021;234:172-180.e3.
2. Lupo PJ, Schraw JM, Desrosiers TA, et al. Association between birth defects and cancer risk among children and adolescents in a population-based assessment of 10 million live births. *JAMA Oncol.* 2019;5(8):1150-1158.
3. Liu B, Filippi S, Roy A, Roberts I. Stem and progenitor cell dysfunction in human trisomies. *EMBO Rep.* 2015;16(1):44-62.
4. Letourneau A, Santoni FA, Bonilla X, et al. Domains of genome-wide gene expression dysregulation in Down's syndrome. *Nature.* 2014;508(7496):345-350.
5. Lane AA, Chapuy B, Lin CY, et al. Triplication of a 21q22 region contributes to B cell transformation through HMGN1 overexpression and loss of histone H3 Lys27 trimethylation. *Nat Genet.* 2014;46(6):618-623.
6. Muskens IS, Li S, Jackson T, et al. The genome-wide impact of trisomy 21 on DNA methylation and its implications for hematopoiesis. *Nat Commun.* 2021;12(1):821.
7. Brown AL, de Smith AJ, Gant VU, et al. Inherited genetic susceptibility to acute lymphoblastic leukemia in Down syndrome. *Blood.* 2019;134(15):1227-1237.
8. Felix JF, Cecil CAM. Population DNA methylation studies in the developmental origins of health and disease (DOHaD) framework. *J Dev Orig Health Dis.* 2019;10(3):306-313.
9. de Smith AJ, Walsh KM, Morimoto LM, et al. Heritable variation at the chromosome 21 gene ERG is associated with acute lymphoblastic leukemia risk in children with and without Down syndrome. *Leukemia.* 2019;33(11):2746-2751.
10. Koestler DC, Jones MJ, Usset J, et al. Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). *BMC Bioinformatics.* 2016;17(1):120.
11. Zheng SC, Breeze CE, Beck S, Teschendorff AE. Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods.* 2018;15(12):1059-1066.

12. Rahmani E, Shenhav L, Schweiger R, et al. Genome-wide methylation data mirror ancestry information. *Epigenetics Chromatin*. 2017;10:1.

13. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010; 26(17):2190-2191.

14. Kachuri L, Jeon S, DeWan AT, et al. Genetic determinants of blood-cell traits influence susceptibility to childhood acute lymphoblastic leukemia. *Am J Hum Genet*. 2021;108(10):1823-1835.

15. van Iterson M, van Zwet EW, Heijmans BT; BIOS Consortium. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* 2017;18(1):19.

16. Jardine L, Webb S, Goh I, et al. Blood and immune development in human fetal bone marrow and Down syndrome. *Nature.* 2021; 598(7880):327-331.

17. Roy A, Cowan G, Mead AJ, et al. Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy21. *Proc Natl Acad Sci USA.* 2012;109(43):17579-17584.

18. Thilaganathan B, Tsakonas D, Nicolaides K. Abnormal fetal immunological development in Down's syndrome. *Br J Obstet Gynaecol.* 1993;100(1):60-62.

19. de Hingh YCM, van der Vossen PW, Gemen EFA, et al. Intrinsic abnormalities of lymphocyte counts in children with down syndrome. *J Pediatr.* 2005;147(6):744-747.

20. Verstegen RHJ, Kusters MAA, Gemen EFA, DE Vries E. Down syndrome B-lymphocyte subpopulations, intrinsic defect or decreased T-lymphocyte help. *Pediatr Res.* 2010;67(5):563-569.

21. Greaves M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat Rev Cancer.* 2018;18(8):471-484.

22. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature.* 2007; 446(7137):758-764.

23. Roberts I, Alford K, Hall G, et al; Oxford-Imperial Down Syndrome Cohort Study Group. GATA1-mutant clones are frequent and often unsuspected in babies with Down syndrome: identification of a population at risk of leukemia. *Blood.* 2013;122(24): 3908-3917.