

## LYMPHOID NEOPLASIA

## In utero cytomegalovirus infection and development of childhood acute lymphoblastic leukemia

Stephen Starko Francis,<sup>1,2</sup> Amelia D. Wallace,<sup>1</sup> George A. Wendt,<sup>1</sup> Linlin Li,<sup>3</sup> Fenyong Liu,<sup>4</sup> Lee W. Riley,<sup>1,3</sup> Scott Kogan,<sup>5</sup> Kyle M. Walsh,<sup>2</sup> Adam J. de Smith,<sup>2</sup> Gary V. Dahl,<sup>6</sup> Xiaomei Ma,<sup>7</sup> Eric Delwart,<sup>3</sup> Catherine Metayer,<sup>1</sup> and Joseph L. Wiemels<sup>2</sup>

<sup>1</sup>Department of Epidemiology, University of California, Berkeley, CA; <sup>2</sup>Helen Diller Family Comprehensive Cancer Center and Departments of Neurological Surgery and Epidemiology and Biostatistics and <sup>3</sup>Blood Systems Research Institute, Department of Laboratory Medicine, University of California, San Francisco, CA; <sup>4</sup>Department of Vaccinology and Infectious Diseases, University of California, Berkeley, CA; <sup>5</sup>Helen Diller Family Comprehensive Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA; <sup>6</sup>Pediatric Hematology/Oncology, Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA; <sup>7</sup>Department of Chronic Disease Epidemiology, School of Public Health, Yale University, New Haven, CT

## Key Points

- CMV is prevalent in pretreatment bone marrow from childhood ALL and not in acute myeloid leukemia.
- In utero infection with CMV is a risk factor for ALL (OR = 3.71,  $P = .0016$ ) and is more pronounced in Hispanics (OR = 5.90,  $P = .006$ ).

It is widely suspected, yet controversial, that infection plays an etiologic role in the development of acute lymphoblastic leukemia (ALL), the most common childhood cancer and a disease with a confirmed prenatal origin in most cases. We investigated infections at diagnosis and then assessed the timing of infection at birth in children with ALL and age, gender, and ethnicity matched controls to identify potential causal initiating infections. Comprehensive untargeted virome and bacterial analyses of pretreatment bone marrow specimens ( $n = 127$  ALL in comparison with 38 acute myeloid leukemia cases in a comparison group) revealed prevalent cytomegalovirus (CMV) infection at diagnosis in childhood ALL, demonstrating active viral transcription in leukemia blasts as well as intact virions in serum. Screening of newborn blood samples revealed a significantly higher prevalence of in utero CMV infection in ALL cases ( $n = 268$ ) than healthy controls ( $n = 270$ ) (odds ratio [OR], 3.71, confidence interval [CI], 1.56-7.92,  $P = .0016$ ). Risk was more pronounced in Hispanics (OR=5.90, CI=1.89-25.96) than in non-Hispanic whites (OR=2.10 CI= 0.69-7.13). This is the first

study to suggest that congenital CMV infection is a risk factor for childhood ALL and is more prominent in Hispanic children. Further investigation of CMV as an etiologic agent for ALL is warranted. (*Blood*. 2017;129(12):1680-1684)

## Introduction

A role for infection in the etiology of childhood acute lymphoblastic leukemia (ALL) has long been hypothesized.<sup>1-3</sup> Although many individual infectious agents and proxies have been investigated,<sup>4-8</sup> a comprehensive untargeted study of possible etiologic viruses and bacteria has not previously been attempted.

Studies of tumor genetic changes in newborn blood spots (NBS) provide unambiguous evidence of the presence of ALL initiation in utero for some subtypes of the disease.<sup>9-11</sup> Decreased interleukin-10 levels at birth in children who later develop ALL<sup>12</sup> and more medically diagnosed infections early in life<sup>7,13</sup> suggest that dysregulated immunity plays an etiologic role. Hispanics have the highest risk of ALL, which appears to be increasing.<sup>14</sup> In this study we seek to answer fundamental etiologic questions: (1) What infections are unique to childhood ALL in comparison with acute myeloid leukemia (AML, used here as an immune suppressed

control group at the time of diagnosis)? (2) Are such infections acquired in utero, and are they risk factors for ALL?

## Study design

We conducted two complementary studies, briefly described below (details are available in the supplemental Materials section, available on the *Blood* Web site).

## Viral and bacterial discovery at diagnosis

We conducted a comprehensive next-generation sequencing-based virome and bacterial metagenomic analysis in pretreatment diagnostic bone marrow of both childhood ALL and AML, drawn from the California Childhood Leukemia Study (CCLS).<sup>15</sup> (All data and samples have been approved for use by the

Submitted 2 July 2016; accepted 17 September 2016. Prepublished online as *Blood* First Edition paper, 15 December 2016; DOI 10.1182/blood-2016-07-723148.

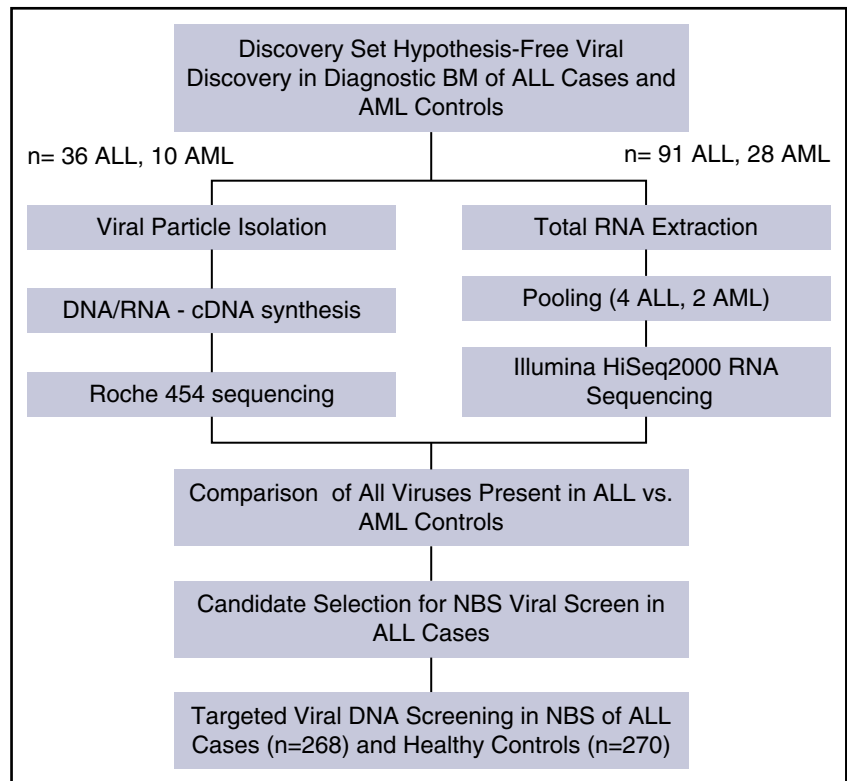
The online version of this article contains a data supplement.

There is an Inside *Blood* Commentary on this article in this issue.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2017 by The American Society of Hematology

**Figure 1. Flowchart of approach.** cDNA, complementary DNA.



University of California, Berkeley, institutional review board [California Childhood Leukemia Study/Protocol 2010-10-2438] and California State institutional review board [Childhood Cancer Record Linkage Project/Protocol 12-07-0529 and California Childhood Leukemia Study/Protocol 2010-10-2438].

Two independent patient sample sets were selected from the CCLS (see the supplemental Methods section) and interrogated using two separate metagenomic methods (Figure 1). The first method used 36 diagnostic bone marrow (BM) samples from children with ALL and 10 diagnostic BM samples from children with AML and were examined using virus isolation by ultracentrifugation and 454 sequencing. The second method used total RNA extracted from 91 ALL BM pooled into 4 groups and 28 AML BM pooled into 2 groups, then sequenced by Illumina Hi-Seq2500. Data generated from each method were analyzed using separate custom bioinformatics pipelines that categorized all nonhuman reads using BLAST (see the supplemental Methods section). Additionally, we used 16s-based bacterial metagenomics on 91 ALL BM and 28 AML BM as a control for bacterial diversity and differential bacterial contamination of the BM samples (see the supplemental Materials section). Sample size was determined on the basis of sample availability and budget constraints.

### Viral screen at birth

NBS from 268 ALL cases and 270 cancer-free controls, randomly drawn from the California Department of Public Health (see the supplemental Methods section, supplemental Table 2), were screened for 2 herpes viruses identified in the discovery set at diagnosis (ie, cytomegalovirus [CMV] and Epstein-Barr virus [EBV]). Viral screen was done by a droplet digital polymerase chain reaction PCR-based direct-detection method, specifically, ultrasensitive “third-generation” ddPCR (Figure 1).

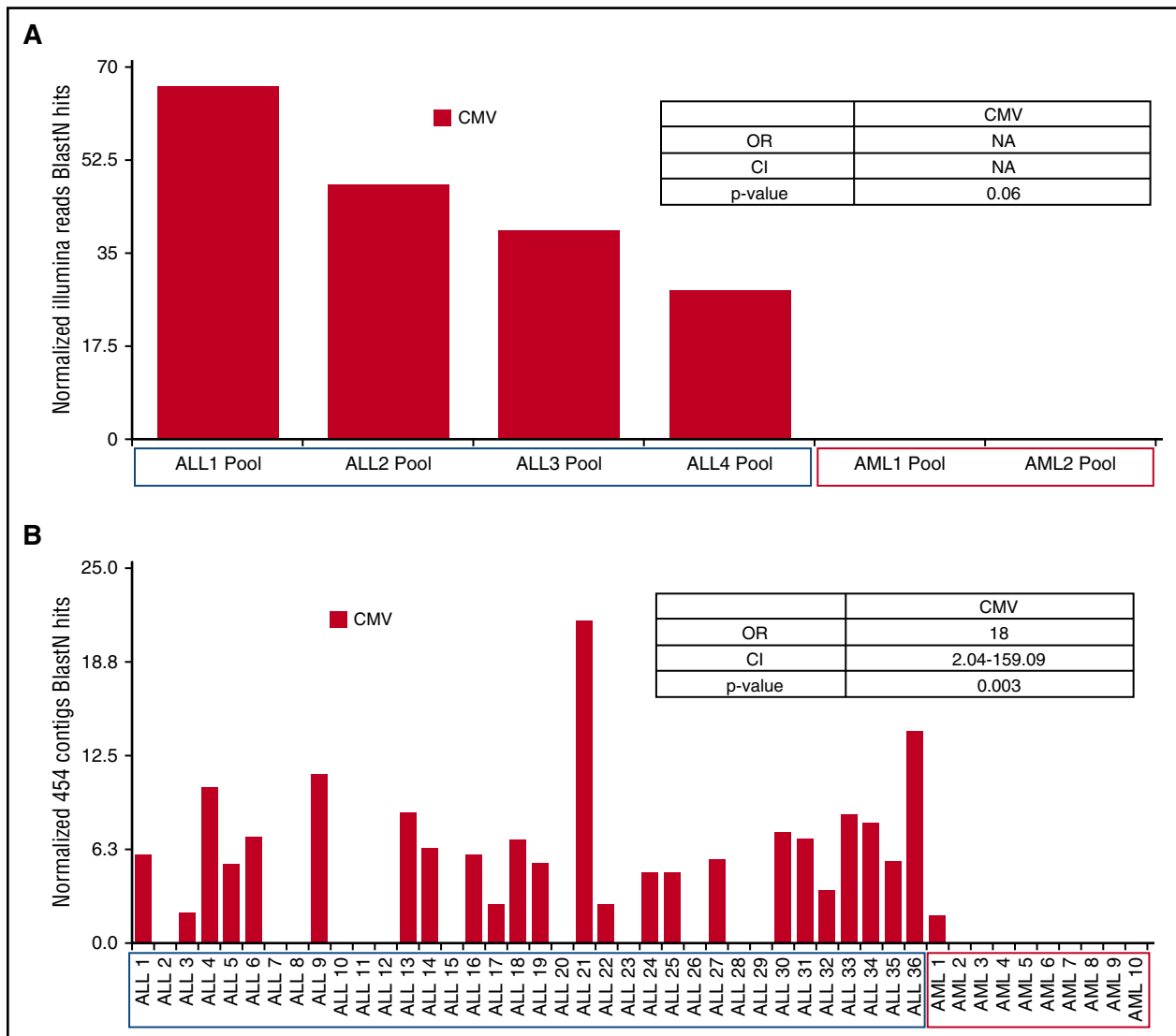
## Results and discussion

In general, ALL patients were younger than AML patients but were otherwise demographically similar (supplemental Table 1). A limitation

of our methods is the ability to primarily detect active infections, although evidence of viruses and bacteria was detected in all pretreatment bone marrow specimens from both ALL and AML groups by both Illumina sequencing (supplemental Figures 1-7) and virus enrichment isolation/454 sequencing. Only herpes viruses distinguished ALL from AML in both analysis sets; all 4 ALL Illumina pools expressed CMV transcripts, whereas no CMV transcripts were detected in AML pools (Figure 2A). Particle isolation and 454 sequencing showed a greater prevalence of CMV (OR = 18,  $P = .003$ ) in individual ALL bone marrow specimens in comparison with AML specimens (Figure 2B).

We initiated this study with no a priori candidate, following identification of CMV at diagnosis and given the in utero origins of ALL. Children who went on to develop ALL were 3.71 times more likely to be CMV positive at birth ( $P = .0016$ ) (Table 1). We found suggestive evidence for statistical interaction between CMV positivity and Hispanic ethnicity (logistic interaction term  $P = .227$ ). Stratification by Hispanic ethnicity showed 5.9-fold increased risk of ALL in Hispanics infected perinatally with CMV (CI = 1.9, 26.0). The prevalence of CMV infection was higher in whites overall, and an increased risk for ALL was also demonstrated in this group; however, the result was not statistically significant (OR = 2.10; CI = 0.69, 7.13). We observed no difference in EBV prevalence at birth between cases and controls (11 positive cases and 11 positive controls; OR = 1.01,  $P = 1.0$ ). After stratifying by ethnicity, no significant effect modification was observed (Table 1). CMV viral load among positive case samples averaged 0.214 copies/ul and was higher than the CMV viral load among positive control samples (0.071 copies/ul; supplemental Table 2;  $P = .003$ ). There was no significant difference in the age at diagnosis between CMV-positive (mean age, 4.8 years) and CMV-negative (mean age, 5.15) ALLs at birth.

Several studies of childhood ALL have examined herpes viruses at diagnosis with conflicting results.<sup>5,6</sup> Two previous studies examined herpes viruses in NBS from children who developed ALL. The first



**Figure 2. Nonhuman sequence search identification (“blast hits”) from pretreatment bone marrow.** (A) Results of deep-sequencing RNA pools (73 ALLs in 4 pools and 28 AMLs in 2 pools) after ribosomal RNA depletion (RiboZERO). Raw Illumina reads were aggressively quality filtered, then aligned to hg19 using BowTie2; nonaligning reads were categorized using an E value less than or equal to  $1 \times 10^{-10}$  with blastn and the nt database. Cytomegalovirus hits were normalized by sequencing effort per pool (blast hits/quality reads  $\times 1.0 \times 10^8$ ). Odds ratios were calculated by comparing presence/absence of virus in ALL versus AML pools; P values were computed by using Fisher’s exact test (for additional details, see the supplemental Methods section). (B) Results of particle isolation and 454 sequencing from an independent study of 36 ALL and 10 AML patients. Virus was isolated using centrifugation and Millipore filter, and then viral nucleic acids were extracted and sequenced using Roche 454 (for additional details, see the supplemental Methods section). Recovered sequences were assembled into contigs in which alignments to hg19 were removed using Bowtie2. Contigs were then categorized with a cutoff of E value less than or equal to  $1.0 \times 10^{-10}$  using blastn and the refseq viral database. Cytomegalovirus contigs were normalized by recovered contigs per patient (blast hits/contigs assembled  $\times 1 \times 10^3$ ); odds ratios were calculated comparing presence/absence of virus in ALL versus AML patients; P values were computed using Fisher’s exact test (for additional details, see the supplemental Methods section). CMV was found to be the only virus that showed statistically significant variation between ALL and AML in the Illumina and 454 groups. CI, confidence interval; NA, not applicable; OR, odds ratio.

study examined EBV and HHV-6, in which no association was found, but did not assess CMV.<sup>16</sup> Another screened NBS for CMV from 48 cases and 46 controls yet found no infected children in either group.<sup>17</sup> With a significantly larger sample size and a more sensitive assay, we could detect extremely low quantities of CMV DNA that may otherwise have been missed. Further replication of our findings is warranted using higher input volumes of NBS DNA, protein-based detection methods, and samples from geographically diverse locations.

CMV is a common virus that infects over 90% of adults worldwide. In the United States, seroprevalence of adult CMV ranges from ~50% to 80% and is highest in Hispanics and blacks,<sup>18</sup> which is particularly interesting given the high rate of ALL in Hispanics.<sup>14</sup> Timing of infection has important implications. In utero infection with CMV is

a leading cause of birth defects, specifically hearing loss. Although primary CMV infection during pregnancy yields the greatest risk for vertical transmission, most women of childbearing age are previously infected with CMV, and therefore vertical transmission of reactivated virus accounts for the majority of congenital CMV infections.<sup>19</sup> The presence of CMV in the child prior to birth may have important immune control implications. Infection prior to the development of a robust adaptive immune response in the fetus and neonate may affect central tolerance and enable CMV to persist in an infectious course, unlike the majority of children who are infected after birth, thus leading to increased risk of developing ALL.

Two features of CMV infection support a role in oncogenesis in ALL. First, congenital CMV infection has been noted to cause

**Table 1. Results of neonatal blood spot screen: cytomegalovirus and Epstein Barr virus**

	CMV–	CMV+	OR (CI)	P value	EBV–	EBV+	OR (CI)	P value
<b>Overall*</b>								
ALL	242	26	3.71 (1.71-8.95)	0.0016	257	11	1.01 (0.42-2.42)	1
Controls	262	8	—	—	259	11	—	—
<b>By race†</b>								
Hispanic ALL	134	15	5.90 (1.89-25.96)	0.006	141	8	1.38 (0.46-4.35)	0.56
Hispanic controls	152	3	—	—	149	6	—	—
White ALL	71	10	2.10 (0.69-7.13)	0.204	79	2	0.53 (0.09-2.94)	0.68
White controls	82	5	—	—	83	4	—	—
Asian/PI ALL	25	1	N/A	1	25	1	N/A	1
Asian/PI controls	21	0	—	—	21	0	—	—
Black ALL	10	0	—	—	10	0	—	—
Black controls	4	0	—	—	4	0	—	—
Other ALL	2	0	—	—	2	0	—	1
Other controls	3	0	—	—	2	1	—	—

ALL, acute lymphoblastic leukemia; CI, confidence interval; CMV, cytomegalovirus; EBV, Epstein Barr virus; OR, odds ratio; PI, Pacific Islander.

\*Logistic regression adjusted for race, age, and sex.

†Logistic regression adjusted for age and sex.

chromosomal instability, which is suspected to be related to its teratogenic properties.<sup>19</sup> This is of interest in leukemia, because specific chromosomal lesions are present at birth in children with ALL.<sup>20</sup> Second, CMV has the largest genome of any known human viral pathogen and harbors many immune evasion genes, indicating that host immune dysregulation is a critical aspect of the CMV lifecycle. The functions of many CMV genes remain poorly understood, and it is likely that their genomic complexity is the direct result of the evolution of latency. Along these lines, CMV has been hypothesized to be an “oncomodulator” in adult gliomas.<sup>21</sup> A delicate balance has evolved between CMV and the immune system, which has led to a high population prevalence of the virus with relatively low morbidity. Nevertheless, disruptions to this balance can have serious and complex pathologic consequences for the host.

This study provides a putative candidate for future investigations. The comparison of ALLs with AMLs was designed solely to generate a putative agent in ALL and was not matched to race, ethnicity, or age. Furthermore, the preponderance of lymphoid versus myeloid cells in the BM samples may bias the results in relation to viruses with a tropism for B cells. A major limitation in the dried blood spots (DBS) CMV screen is our lack of information on ALL subtype and cytogenetic characteristics. Additionally, the low input quantity of DBS DNA may have resulted in false negatives, though this may drive results toward the null. Additional studies defining in utero CMV infection in terms of ALL subtype, cytogenetics, and overall prevalence are needed.

Our findings lead us to hypothesize that in utero or perinatal CMV infection initiates immune dysregulation during the critical period of fetal immune development, allowing a greater number of and more fulminant infections later in life, as observed in epidemiologic studies.<sup>7,13</sup> Further validation of our findings is warranted to establish ALL as an additional reason to develop an effective CMV vaccine.<sup>22</sup>

## Acknowledgments

The authors thank the families that participate in the California Childhood Leukemia Study. Without their time and effort none of our studies would be possible. The authors also thank our clinical collaborators throughout California for their continued support of our research and commitment to their patients. Finally, the authors thank the California Department of Public Health Genetic Disease Screening Program.

This work was supported by grants from the National Institutes of Health (NIH), National Institute of Environmental Health Sciences (R01ES09137); the NIH, National Heart, Lung, and Blood Institute (R01HL105770); and the NIH, National Cancer Institute (1T32CA151022-01, R01CA155461, and R01CA185058).

## Authorship

Contribution: S.S.F. conceived and designed the study, performed laboratory experiments, analyzed data, and wrote the primary and subsequent versions of the manuscript; C.M. and G.V.D. enrolled participants and obtained samples; A.D.W. and L.L. helped to design and conduct laboratory experiments; G.A.W. provided bioinformatic and data analysis support; S.K., X.M., and K.M.W. aided in refining the study and helped prepare the manuscript; E.D., F.L., X.M., and L.W.R. helped to design the study; J.L.W. helped to design and conduct the study and prepare the manuscript; all coauthors have reviewed and edited the manuscript.

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

Correspondence: Stephen Starko Francis, 1450 3rd St, HD442, University of California, San Francisco, CA 94158; e-mail: ssfrancis@unr.edu or stephen.francis@ucsf.edu.

## References

- Greaves MF. Speculations on the cause of childhood acute lymphoblastic leukemia. *Leukemia*. 1988;2(2):120-125.
- Kinlen L. Evidence for an infective cause of childhood leukaemia: comparison of a Scottish new town with nuclear reprocessing sites in Britain. *Lancet*. 1988;2(8624):1323-1327.
- Curnen MGM, Varma AAO, Christine BW, Turgeon LR. Childhood leukemia and maternal infectious diseases during pregnancy. *J Natl Cancer Inst*. 1974;53(4):943-947.
- Heegaard ED, Schmiegeler K. Serologic study on parvovirus b19 infection in childhood acute lymphoblastic leukemia during chemotherapy: clinical and hematologic implications. *J Pediatr Hematol Oncol*. 2002;24(5):368-373.
- MacKenzie J, Gallagher A, Clayton RA, et al. Screening for herpesvirus genomes in common acute lymphoblastic leukemia. *Leukemia*. 2001; 15(3):415-421.

6. Hermouet S, Sutton CA, Rose TM, et al. Qualitative and quantitative analysis of human herpesviruses in chronic and acute B cell lymphocytic leukemia and in multiple myeloma. *Leukemia*. 2003;17(1):185-195.
7. Chang JS, Tsai CR, Tsai YW, Wiemels JL. Medically diagnosed infections and risk of childhood leukaemia: a population-based case-control study. *Int J Epidemiol*. 2012;41(4):1050-1059.
8. Rudant J, Lightfoot T, Urayama KY, et al. Childhood acute lymphoblastic leukemia and indicators of early immune stimulation: a Childhood Leukemia International Consortium study. *Am J Epidemiol*. 2015;181(8):549-562.
9. Wiemels JL, Cazzaniga G, Daniotti M, et al. Prenatal origin of acute lymphoblastic leukaemia in children. *Lancet*. 1999;354(9189):1499-1503.
10. Gruhn B, Taub JW, Ge Y, et al. Prenatal origin of childhood acute lymphoblastic leukemia, association with birth weight and hyperdiploidy. *Leukemia*. 2008;22(9):1692-1697.
11. Greaves MF, Wiemels J. Origins of chromosome translocations in childhood leukaemia. *Nat Rev Cancer*. 2003;3(9):639-649.
12. Chang JS, Zhou M, Buffler PA, Chokkalingam AP, Metayer C, Wiemels JL. Profound deficit of IL10 at birth in children who develop childhood acute lymphoblastic leukemia. *Cancer Epidemiol Biomarkers Prev*. 2011;20(8):1736-1740.
13. Crouch S, Lightfoot T, Simpson J, Smith A, Ansell P, Roman E. Infectious illness in children subsequently diagnosed with acute lymphoblastic leukemia: modeling the trends from birth to diagnosis. *Am J Epidemiol*. 2012;176(5):402-408.
14. Barrington-Trimis JL, Cockburn M, Metayer C, Gauderman WJ, Wiemels J, McKean-Cowdin R. Rising rates of acute lymphoblastic leukemia in Hispanic children: trends in incidence from 1992 to 2011. *Blood*. 2015;125(19):3033-3034.
15. Francis SS, Selvin S, Metayer C, et al. Mode of delivery and risk of childhood leukemia. *Cancer Epidemiol Biomarkers Prev*. 2014;23(5):876-881.
16. Bogdanovic G, Jernberg AG, Priftakis P, Grillner L, Gustafsson B. Human herpes virus 6 or Epstein-Barr virus were not detected in Guthrie cards from children who later developed leukaemia. *Br J Cancer*. 2004;91(5):913-915.
17. Gustafsson B, Jernberg AG, Priftakis P, Bogdanovic G. No CMV DNA in Guthrie cards from children who later developed ALL. *Pediatr Hematol Oncol*. 2006;23(3):199-205.
18. Staras SA, Dollard SC, Radford KW, Flanders WD, Pass RF, Cannon MJ. Seroprevalence of cytomegalovirus infection in the United States, 1988-1994. *Clin Infect Dis*. 2006;43(9):1143-1151.
19. Cheeran MC, Lokensgard JR, Schleiss MR. Neuropathogenesis of congenital cytomegalovirus infection: disease mechanisms and prospects for intervention. *Clin Microbiol Rev*. 2009;22(1):99-126. [Table of Contents.]
20. Mori H, Colman SM, Xiao Z, et al. Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proc Natl Acad Sci USA*. 2002;99(12):8242-8247.
21. Dey M, Ahmed AU, Lesniak MS. Cytomegalovirus and glioma: putting the cart before the horse. *J Neurol Neurosurg Psychiatry*. 2015;86(2):191-199.
22. Bialas KM, Permar SR. The march towards a vaccine for congenital CMV: rationale and models. *PLoS Pathog*. 2016;12(2):e1005355.