haematologica
Journal of the Ferrata Storti Foundation

A first-born twin has a higher risk of acute leukemia in a population-based assessment of cancer in twins in California, and a lower than anticipated rate of twin concordance

by Eric M. Nickels, Naying Zhou, and Joseph L. Wiemels

Received: August 6, 2024.
Accepted: December 31, 2024.

Citation: Eric M. Nickels, Naying Zhou, and Joseph L. Wiemels. A first-born twin has a higher risk of acute leukemia in a population-based assessment of cancer in twins in California, and a lower than anticipated rate of twin concordance.
Haematologica. 2025 Jan 16. doi: 10.3324/haematol.2024.286309 [Epub ahead of print]

A first-born twin has a higher risk of acute leukemia in a population-based assessment of cancer in twins in California, and a lower than anticipated rate of twin concordance

Eric M Nickels,[1,2] Naying Zhou,[2] Joseph L Wiemels[2]

[1]Cancer and Blood Disease Institute, Children's Hospital Los Angeles, Los Angeles, CA, USA
[2]Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

Correspondence: J.L. Wiemels - wiemels@usc.edu

Disclosures: No conflicts of interest to disclose.

Contributions:
EMN: Data curation, methodology, statistical analysis, data visualization and manuscript writing.
NZ: Statistical analysis, manuscript writing
JLW: Methodology, manuscript writing

Acknowledgements: The collection of cancer incidence data used in this study was supported by the California Department of Public Health as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885; the National Cancer Institute's Surveillance, Epidemiology, and End Results 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; and the Centers for Disease Control and Prevention's National Program of Cancer Registries, under agreement U58DP003862-01 awarded to the California Department of Public Health. The ideas and opinions expressed herein are those of the author(s), and no endorsement by the California Department of Public Health, the National Cancer Institute, or the Centers for Disease Control and Prevention or their contractors and subcontractors is intended or should be inferred.

Funding: Funding provided by the V Foundation (FP067172), the USC Norris Comprehensive Cancer Center SPORE development award (JLW), and the and the Shirley McKernan Courage Foundation (CHLA) (JLW).

Data-Sharing Statement: This study used biospecimens from the California Biobank Program. Per California Health and Safety Code Sections 124980(j), 124991(b), (g), (h), and 103850 (a) and (d), which protects the confidentiality of data obtained from biospecimens, we are respectfully unable to share raw, individual level genomic and genome-wide DNA methylation data reported in this study, which are property of the State of California. Should we be contacted regarding individual level data contributing to the findings reported in this study, inquiries will be directed to the California Department of Public Health Institutional Review Board to establish an approved protocol to utilize the data, which cannot otherwise be shared peer-to-peer.

MAIN TEXT

Investigation of twins has enabled estimations of the germline and somatic genetic contribution to many pediatric cancers.[1] Twin concordance for childhood acute leukemias is estimated at approximately 10%[2] for lymphoblastic (ALL) and myeloid leukemia (AML) phenotypes. While purpose-built twin cohorts allow for large enrollments of twin cases with leukemia, these studies are at risk of inaccuracies resulting from ascertainment bias.[3] In contrast, records-based population-based studies allow accrual in an unbiased fashion. Here, we identified substantially lower than expected pediatric and adolescent/young adult (AYA) twin leukemia concordance rates through a population-based registry assessment in California from 1982-2022. We further demonstrated a unique association showing first-born twins are more likely to develop ALL than second-born. We show the difference in risk of ALL development by birth order may be reflected in epigenome-wide alterations in DNA methylation, in which first born twins show significant hypomethylation compared to second-born siblings. Furthermore, we show significant alterations in DNA methylation at sites with known association to birth order and cortisol responsiveness, implying a potential biological explanation for this finding.

This research was performed under review by a California State IRB which permits health research on de-identified Biobank specimens. We merged registry data from the California Birth Master Statistical File (BMSF) from 1982 to 2017 and the California Cancer Registry (CCR) from 1988 to 2022 to identify a total of 1,213 twins in which at least one sibling was diagnosed with leukemia from 0-39 years of age. Of these, there were 255 twin pairs in which at least one sibling in the pair had a diagnosis of leukemia, along with 265 distinct cancer diagnoses and 258 leukemia diagnoses within these 255 twin pairs (Table 1).

Concordant cases were defined as both siblings in a twin pair having diagnoses in the same subgroup, and cross-cancer concordance as both having diagnoses in separate subgroups. Two concordant twin pairs were identified for both lymphoid leukemia and AML (Table 2). Concordance rates were defined as the total number of concordant twin pairs divided by the sum of concordant and discordant twin pairs. The twin concordance rate was 0.8% (2 of 255 twin pairs) for all leukemias, 0.5% for lymphoid leukemias (1 of 199), and 2.9% for AML (1 of 34). The percentage of monozygotic (MZ) twins was estimated assuming dizygotic twins are equally likely to be represented as same sex or opposite sex twins[3]. The estimated MZ concordance rate was 2.6% for all leukemias, 1.4% for lymphoid leukemias, and 16.7% for AML. By diagnosis age, the rate of lymphoid leukemia concordance for 1-10 years was 0.7% (MZ 2.1%). The rate of AML development for ages <1 year was 25% (MZ 100%). Standardized incidence ratios were calculated based on previous published methods[4]. The SIR for developing a leukemia of the same subgroup as the proband in twin siblings was 15.6 (95% CI 1.9-56.4), and for development of any pediatric/AYA cancer 2.9 (0.4-10.5, *Online Supplementary Table S1*).

We next assessed associations between twin plurality order (being born first or second within the twin pair) and development of leukemia. Across the full twin cohort, there was a significant increased risk of leukemia development associated with being the first-born twin (Fisher's exact test OR 1.80, P=0.001, Figure 1A). Lymphoblastic leukemias were the most highly associated with being first-born (OR 2.26, P<0.001) and did not vary by age at cancer diagnosis or mode of birth (data not shown), while AML showed a non-significant *decreased* risk (OR 0.55, P=0.127). To evaluate how plurality order may be reflected by DNA methylation at birth, we investigated 41 pediatric ALL-discordant twin pairs who previously underwent genome-wide DNA methylation profiling at birth[5] for differences in DNA methylation based on twin plurality order. Deconvolution analysis showed no significant differences in nucleated cell proportions by plurality order. We used linear regression controlling for ALL cases status, birthweight, sex, twin pair number, nucleated cell proportions and array batch effect to identify 1,394 significant differentially methylated probes (FDR<0.05, 710,010 total CpGs assessed), with a moderately elevated genomic inflation value of 2.21 (Figure 1B-C, *Online Supplementary Table S2*). There was a total of 534 significant differentially methylated regions assessed through Comb-P (Šidák P<0.05, Figure 1D, *Online Supplementary Table S3*).

We then evaluated CpGs with known association to birth order[6] across subsequent gestations. Of 341 birth-order CpGs, 305 overlapped with the 710,010 probes from the EPIC array the ALL-discordant twin cohort. Mean delta-beta values (the first minus second-born absolute DNA methylation value) were significantly lower (hypomethylated) across the 24 twin pairs in which the ALL case was first-born compared to the 17 in which the case was second born (Wilcoxon rank-sum P=0.037, Figure 1E). We further investigated 22,100 CpG sites nominally associated with cortisol levels,[7] of which 19,193 CpGs overlapped with the ALL twin cohort EPIC

array. Mean delta beta values are significantly different in instances in which the ALL case sibling was born first compared to the case sibling being born second (P=0.011).

We identified an overall concordance rate of 0.8% for twins with leukemia, which is notably lower than prior studies which range from 5-25%.[2] Our estimation of concordance in MZ twins for lymphoid leukemias remains low at 1.4%, while that of AML is within the previously reported range at 16.7%. Accounting for approximately 70% of MZ being monochorionic,[8] concordance increases to just 2.4% for ALL twins, and 3.4% specifically for the 1 to 10 age group. Remarkably, none of the 6 instances of lymphoid leukemia in twins <1 year demonstrated sibling concordance, despite prior estimated rates nearing 100%.[2] Twin studies previously helped to identify that the initiating somatic genetic translocation event in childhood leukemia development can occur in the intrauterine period and subsequently pass between siblings through shared placental circulation[9,10] in instances of MZ, monochorionic twins. The low rate of twin concordance in this study indicates that *acquired* genetic, epigenetic and discordant responses to environmental exposures events may contribute more heavily to pediatric and AYA cancer etiology than previously appreciated in pediatric acute leukemias,[11] or at least among those in recent decades in California. We identified a twin pair concordant for lymphoid leukemias with specific diagnoses of B-cell lymphoblastic leukemia/lymphoma and Burkitt's cell leukemia which do not share known driver mutations. This combination suggests an alternative disease mechanism of concordance than that of shared fetal circulation and may point toward shared genetic predisposition to hematologic malignancies, which we were not able to evaluate.

These results argue that concordant cases of twin acute leukemia are rarer than previously reported, which may result from biases in active twin recruitment leading to oversampling of rare concordant twins relative to the true population background rate.[2] It is also notable that this study examined twins in a multiethnic birth population in which Latino births represent the majority population; most prior population-based studies being performed in non-Latino whites. While these results do not directly alter clinical recommendations,[12] additional nuance should be considered in how concordance risk is discussed with families after an initial diagnosis of ALL in a twin sibling.

This study was limited by the lack of zygosity information available in the evaluated registry data; however, our estimation of MZ twin pairs allows for broad comparison to previously identified concordance rates. In a subset of twins with ALL from this cohort investigated with a DNA single nucleotide polymorphism array, approximately 50% of same sex pairs were MZ,[5] which is consistent with the estimated results displayed here. Furthermore, we are limited by the right censoring of registry data (to the year 2022), raising potential for lead-time bias in twin siblings born in the latest birth years assessed, though the inclusion of over three decades of cancer and birth registry data limits this impact.

These results indicate first-born twins are more likely to develop ALL compared to second born twins, a finding which has not previously been described. We conducted linear regression modeling in twins discordant for ALL to investigate how this relationship might be reflected in DNA methylation at birth in which we found evidence of genomic inflation indicative of higher-than-expected P-values across most CpG probes on the EPIC array. This finding suggests the underlying biological mechanism responsible for the difference in leukemia risk between first and second born twins is reflected in a broad alteration (generally in the hypo-direction) of DNA methylation across the epigenome.

Interestingly, first-born singleton-births (compared to subsequent gestations) are at increased risk of ALL,[13] while the inverse is seen for pediatric AML,[14] which mirrors our results here with twins. The relationship of birth order with DNA methylation[6] also mirrors our results, suggesting the biological mechanism leading to an increase in ALL risk in first born singletons may have similarities to that of first-born twins. We further found a significant association between the development of ALL and DNA methylation at cortisol-associated sites, indicating lower exposure to cortisol during birth may contribute to the significantly increased risk of ALL in first-born twins. Increased, early exposure to cortisol is hypothesized to reduce risk of ALL.[15] Further investigation in a broader cohort of twins with acute leukemia is warranted.

REFERENCES

1. Kadan-Lottick NS, Kawashima T, Tomlinson G, et al. The risk of cancer in twins: a report from the childhood cancer survivor study. Pediatr Blood Cancer. 2006;46(4):476-481.
2. Greaves MF, Maia AT, Wiemels JL, Ford AM. Leukemia in twins: lessons in natural history. Blood. 2003;102(7):2321-2333.
3. Buckley JD, Buckley CM, Breslow NE, Draper GJ, Roberson PK, Mack TM. Concordance for childhood cancer in twins. Med Pediatr Oncol. 1996;26(4):223-229.
4. Feng Q, Nickels E, Muskens IS, et al. Increased burden of familial-associated early-onset cancer risk among minority Americans compared to non-Latino Whites. Elife. 2021;10:e64793.
5. Nickels EM, Li S, Myint SS, et al. DNA methylation at birth in monozygotic twins discordant for pediatric acute lymphoblastic leukemia. Nat Commun. 2022;13(1):6077.
6. Li S, Spitz N, Ghantous A, et al. A Pregnancy and Childhood Epigenetics Consortium (PACE) meta-analysis highlights potential relationships between birth order and neonatal blood DNA methylation. Commun Biol. 2024;7(1):66.
7. Houtepen LC, Vinkers CH, Carrillo-Roa T, et al. Genome-wide DNA methylation levels and altered cortisol stress reactivity following childhood trauma in humans. Nat Commun. 2016;7:10967.
8. Hall JG. Twinning. Lancet. 2003;362(9385):735-743.
9. Greaves M. In utero origins of childhood leukaemia. Early Hum Dev. 2005;81(1):123-129.
10. Wiemels JL, Xiao Z, Buffler PA, et al. In utero origin of t(8;21) AML1-ETO translocations in childhood acute myeloid leukemia. Blood. 2002;99(10):3801-3805.
11. Hong D, Gupta R, Ancliff P, et al. Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. Science. 2008;319(5861):336-339.
12. Ford AM, Colman S, Greaves M. Covert pre-leukaemic clones in healthy co-twins of patients with childhood acute lymphoblastic leukaemia. Leukemia. 2023;37(1):47-52.
13. Paltiel O, Lemeshow S, Phillips GS, et al. The association between birth order and childhood leukemia may be modified by paternal age and birth weight. Pooled results from the International Childhood Cancer Cohort Consortium (I4C). Int J Cancer. 2019;144(1):26-33.
14. Von Behren J, Spector LG, Mueller BA, et al. Birth order and risk of childhood cancer: a pooled analysis from five US States. Int J Cancer. 2011;128(11):2709-2716.
15. Schmiegelow K, Vestergaard T, Nielsen SM, Hjalgrim H. Etiology of common childhood acute lymphoblastic leukemia: the adrenal hypothesis. Leukemia. 2008;22(12):2137-2141.

TABLES

Table 1: Twin leukemia cohort characteristics. *Diagnoses were categorized by ICD-O-3 histology codes following the International Classification of Childhood Cancer, Third Edition, into five leukemia subgroups.

Characteristic	Value
Total number of twin pairs	255*
Opposite sex pairs, N (%)	89 (34.9%)
Same sex pairs	166 (65.1%)
Same sex, Male	101 (60.8%)
Same sex, Female	65 (39.2%)
Birth Characteristics	
Gestational age, days (median, IQR, range)	260 days, 21, 183-309
Age of mother, years	31 years, 10, 16-49
Age of father, years	33 years, 10, 18-60
Cesarian delivery, N (%)	183 (71.8%)
Vaginal delivery	72 (28.2%)
Use of assisted reproductive technology	8 (3.1%)
Birthweight, grams (SD, P-value)	
Cases	2501 (514, P = 1)
Controls	2473 (524)
Cancer Diagnosis Characteristics	
Total cancer diagnoses, N	265
Total leukemia diagnoses	258
Leukemia diagnoses by category*	
Lymphoid leukemias	200
Acute myeloid leukemia	35
Chronic myeloproliferative diseases	11
Myelodysplastic syndrome and other myeloproliferative diseases	8
Unspecified and other specified leukemias	4
Age of leukemia diagnosis (median, IQR, range)	6 years, 10, 0-30
Number of diagnoses by age category, N	
<1	13
1-9 years	160
10-19 years	52
>20 years	33

Table 2: Twin pairs with concordant and cross-cancer concordant leukemia diagnoses. Concordant cases were defined as twin siblings sharing leukemia diagnoses in the same ICCC-3 Recode subgroup. Cross-cancer concordant cases were defined as twin siblings with cancer diagnoses in separate recode subgroups, with at least one having a leukemia diagnosis. Ages are given in Bottom panel shows rates of concordance in twin cancer cases across all leukemia diagnoses, by leukemia subgroup and by age of diagnosis. The number of monozygotic (MZ) twin pairs was estimated based on the distribution of same-sex and opposite sex twin pairs, assuming all concordant cases within the same-sex group are from MZ twins. Conc = Concordant, Disc = Discordant, MZ = Monozygotic.

Pair #		Cancer Subgroup (ICCC-3 Recode)	ICD-O3	Diagnosis Description	Sex	Age Range (Years)
Leukemia-concordant						
1	Twin A	Acute myeloid leukemias	9872/3	Acute myeloid leukemia, minimal differentiation	M	Age <1
	Twin B	Acute Myeloid Leukemias	9872/3	Acute myeloid leukemia, minimal differentiation	M	Age <1
2	Twin A	Lymphoid leukemias	9826/3	Burkitt's cell leukemia	M	Age 1 to 10
	Twin B	Lymphoid leukemias	9811/3	B lymphoblastic leukemia/lymphoma, NOS	M	Age 1 to 10
Cross-cancer concordant						
1	Twin A	Lymphoid Leukemias	9811/3	B lymphoblastic leukemia/lymphoma, NOS	F	Age 1 to 10
	Twin B	Rhabdomyosarcomas	8910/3	Embryonal rhabdomyosarcoma, NOS	F	Age 1 to 10

Category	All			Same Sex			Estimated MZ
	Conc	*Disc*	*Rate (%)*	*Conc*	*Disc*	*Rate (%)*	*Rate (%)*
All Leukemias	**2**	**253**	**0.8**	**2**	**164**	**1.2**	**2.6**
Lymphoid Leukemias							
All Ages	**1**	**198**	**0.5**	**1**	**133**	**0.7**	**1.4**
Less than 1 year of age at diagnosis	0	6	0	0	5	0	0
1 to 10 years at diagnosis	1	148	0.7	1	97	1.0	2.0
Over 10 years at diagnosis	0	44	0	0	31	0	0
Acute Myeloid Leukemias							
All Ages	**1**	**33**	**2.9**	**1**	**19**	**5.0**	**16.7**
Less than 1 year of age at diagnosis	1	3	25	1	2	33	100
1 to 10 years at diagnosis	0	10	0	0	7	0	0
Over 10 years at diagnosis	0	20	0	0	10	0	0
Chronic myeloproliferative diseases	**0**	**11**	**0**	**0**	**8**	**0**	**0**
Myelodysplastic syndrome and myeloproliferative diseases	**0**	**8**	**0**	**0**	**3**	**0**	**0**
Unspecified and other specified leukemias	**0**	**4**	**0**	**0**	**1**	**0**	**0**

FIGURE LEGENDS

Figure 1: Birth plurality order is significantly associated with development of leukemia in leukemia-discordant twin pairs. (A) Forest plot demonstrating odds ratio from Fisher's exact test for the association between twin plurality order (first or second born of a set of twin siblings) and pediatric leukemia in leukemia-discordant twin pairs. A significantly higher odds of developing leukemia of any type is seen in first born twins. By disease group, twins discordant for the development of lymphoid leukemia had significantly higher odds of being first born. Twins discordant for development of acute myeloid leukemia, however, had lower odds of being first born, however this difference was not statistically significant. **(B)** Quantile-quantile plot of linear regression model for the relationship between DNA methylation status and twin plurality order. Lambda value represents genomic inflation. **(C)** Volcano plot from the linear regression model showing significance ($-\log_{10}P$) by regression coefficient. The 1,394 significant differentially methylated probes (FDR<0.05) are shown in red. (D) Volcano plot showing distribution of $-\log_{10}P$ values by genomic position. Probes found within the 534 significant differentially methylated regions are highlighted in red. **(E)** Distribution of mean delta beta (first-born sibling DNA methylation beta value minus second-born within a twin pair) values across 305 CpGs with known association to birth order found within the discordant acute lymphoblastic leukemia (ALL) twin DNA methylation array dataset. Mean delta beta values are significantly different in instances in which the ALL case sibling was born first (N = 24) compared to the case sibling being born second (N = 17, Wilcoxon rank-sum P = 0.037). **(F)** Distribution of mean delta beta values across 19,193 CpGs with known association to post-partum cortisol levels found within the discordant ALL twin DNA methylation array dataset. Mean delta beta values are significantly different in instances in which the ALL case sibling was born first (N = 24) compared to the case sibling being born second (P = 0.011).

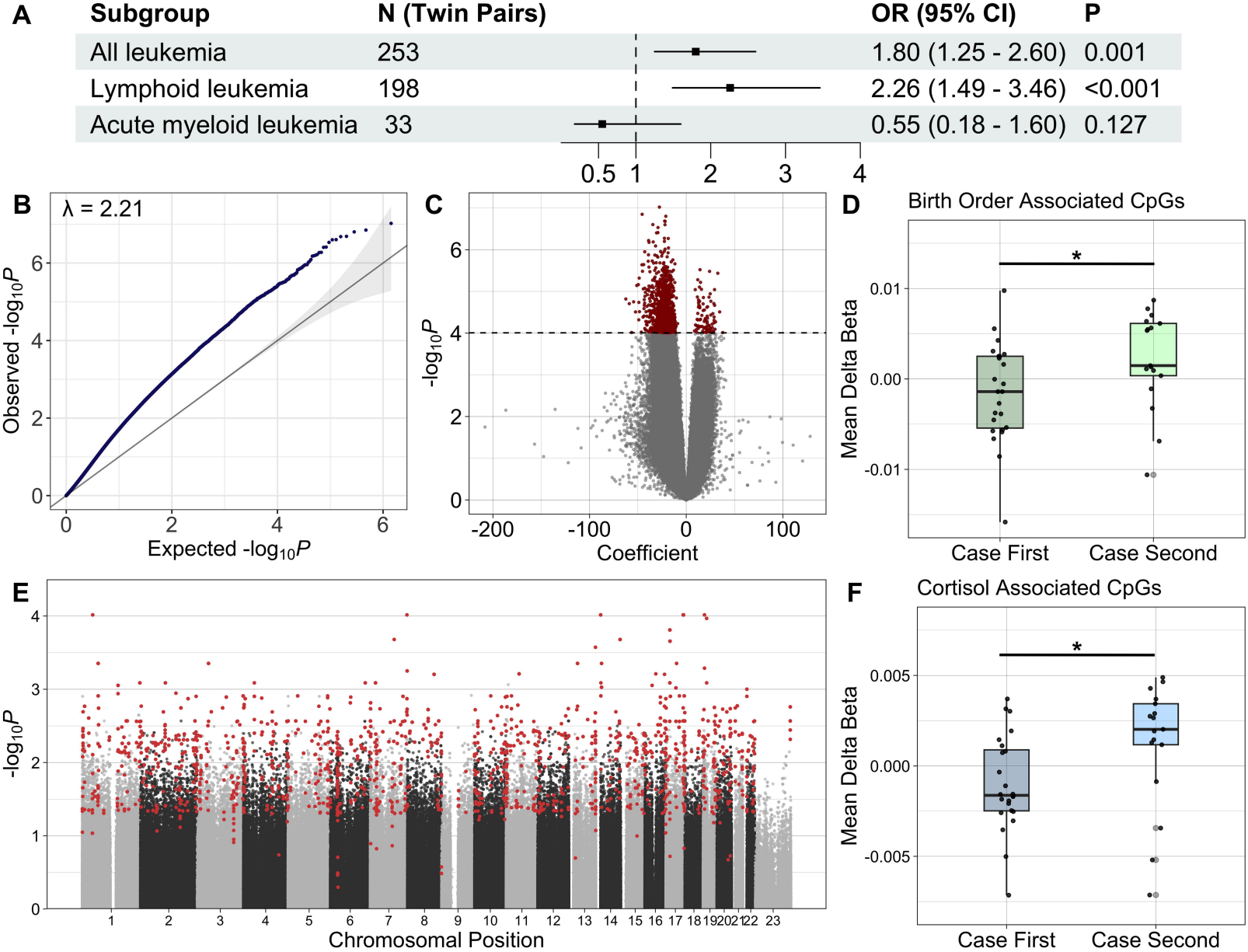
A
Subgroup
N (Twin Pairs)
OR (95% CI)
P
All leukemia
253
1.80 (1.25 - 2.60)
0.001
Lymphoid leukemia
198
2.26 (1.49 - 3.46)
<0.001
Acute myeloid leukemia
33
0.55 (0.18 - 1.60)
0.127
0.5
1
2
3
4
B
λ = 2.21
Observed -log10P
Expected -log10P
C
-log10P
Coefficient
D
Birth Order Associated CpGs
Mean Delta Beta
Case First
Case Second
E
-log10P
Chromosomal Position
F
Cortisol Associated CpGs
Mean Delta Beta
Case First
Case Second

Table S1: Standardized incidence ratios (SIRs) of risk of any cancer development in siblings of twins with cancer for the same pediatric or adolescent/young adult (AYA) cancer by ICCC-3 subgroup (top panel) and for any pediatric/AYA cancer type (bottom panel). Results are displayed by broad cancer group diagnosed in the proband, and SIRs are based on risk of any cancer type in the proband. The SIR was obtained by dividing the observed number of siblings with cancer by the expected number of cancers among siblings, with expected values generated as the total years at risk are multiplied by the corresponding age group (5-year intervals) and sex-specific incidence rate of cancer, calculated individually and summed. Years at risk are individually determined by subtracting the age at the start of risk (the year of cancer diagnosis in the proband) from the age at the end of risk (the end year of follow up or the year when the sibling received a cancer diagnosis). Age group and sex-specific incidence rates of cancer were obtained from SEER*Stat version 8.4.3, utilizing California data from SEER 17 Registries, and 95% confidence intervals (CIs) were computed assuming a Poisson distribution.

SIR for development of the same cancer type				
Category	**Probands**	**Observed**	**Expected**	**SIR (95% CI)**
Leukemias	254	2	0.13	15.6 (1.9 - 56.4)
Lymphomas	134	1	0.06	16.3 (0.4 - 90.7)
CNS Neoplasms	233	1	0.06	17.2 (0.4 - 95.6)
Neuroblastoma	54	1	0.004	249.9 (6.3 - 1392.4)
Retinoblastoma	17	2	0.0006	3166.0 (383.4 - 11436.9)
Renal Tumors	39	0		
Hepatic Tumors	33	0		
Malignant bone tumors	54	0		
Soft Tissue Sarcomas	91	1	0.02	62.4 (1.6 - 347.9)
Germ Cell Tumors	95	3	0.04	66.7 (13.8 - 195)
Epithelial Neoplasms	214	2	0.37	5.4 (0.7 - 19.7)
Other Neoplasms	7	0		
SIR for development of the any cancer type				
Category	**Probands**	**Observed**	**Expected**	**SIR (95% CI)**
All Cancer	1213	15	3.47	4.3 (2.4 - 7.1)
Broad groups				
Leukemias	254	2	0.69	2.9 (0.4, 10.5)
Lymphomas	134	1	0.41	2.5 (0.1, 13.7)
CNS Neoplasms	233	1	0.62	1.6 (0.0, 9.0)
Neuroblastoma	54	1	0.14	7.2 (0.2, 40.0)
Retinoblastoma	17	2	0.07	29.2 (3.5, 105.4)
Renal Tumors	39	0		
Hepatic Tumors	33	0		
Malignant bone tumors	54	1	0.16	6.4 (0.2, 35.4)
Soft Tissue Sarcomas	91	2	0.29	6.8 (0.8, 24.6)
Germ Cell Tumors	95	4	0.26	15.5 (4.2, 39.7)
Epithelial Neoplasms	214	2	0.67	3.0 (0.4, 10.8)
Other Neoplasms	12	0		

Table S2: Top 20 significant differentially methylated probes at birth identified from birth plurality order linear regression analysis in 41 ALL-discordant twin pairs. Positions referenced to Hg19. Std. Error = Standard error. FDR = False discovery rate. Chr. = Chromosome.

CpG	Estimate	Std. Error	P-Value	FDR	Chr.	Position	UCSC_RefGene_Name
cg26335251	-45.606557	5.89765682	1.41E-07	0.02548192	chr17	75539921	NA
cg10817023	-21.832225	2.89403871	2.08E-07	0.02548192	chr5	133853517	NA
cg19442702	-22.309636	2.95434736	2.05E-07	0.02548192	chr5	398425	AHRR
cg01321488	-21.106998	2.74806767	1.57E-07	0.02548192	chr9	127380298	NR6A1;NR6A1;NR6A1
cg09738193	-21.973898	2.94816072	2.51E-07	0.02548192	chr16	67926317	PSKH1
cg26618645	-34.533609	4.63063711	2.49E-07	0.02548192	chr16	30780332	RNF40
cg01813254	-27.728941	3.49704088	9.49E-08	0.02548192	chr11	9036548	NA
cg19262334	-29.220747	3.9599124	2.93E-07	0.02602671	chr18	74631861	ZNF236
cg05573381	-30.678108	4.23816372	3.94E-07	0.02796016	chr5	1653258	NA
cg05052194	-29.266438	4.03893695	3.88E-07	0.02796016	chr1	2160249	SKI
cg18690833	-34.167392	4.87896345	6.49E-07	0.02949527	chr20	13972022	SEL1L2
cg10329345	-22.583062	3.18165365	5.30E-07	0.02949527	chr1	45083079	RNF220
cg09407429	-22.558022	3.21493351	6.30E-07	0.02949527	chr3	4534383	ITPR1;ITPR1;ITPR1
cg25528260	-38.485094	5.50421806	6.65E-07	0.02949527	chr7	72777727	NA
cg07876051	-38.942799	5.53220661	6.01E-07	0.02949527	chr10	3159145	PFKP
cg07805959	-18.607996	2.62367993	5.37E-07	0.02949527	chr17	2595004	KIAA0664
cg27120649	-31.710823	4.59687307	8.12E-07	0.03393192	chr7	94286261	SGCE;PEG10;SGCE;PEG10;SGCE
cg07607077	-19.43779	3.30165743	7.63E-06	0.03408781	chr7	919584	C7orf20
cg04376617	32.6600413	5.25666516	3.66E-06	0.03408781	chr18	43685360	ATP5A1;HAUS1;HAUS1
cg12285834	-31.06016	4.7980595	2.05E-06	0.03408781	chr3	72345017	NA

Table S3: Top 20 significant differentially methylated regions at birth identified from Comb P analysis of birth plurality order linear regression analysis in 41 ALL-discordant twin pairs. Positions referenced to Hg19. Min. P-value = Minimum P-value of CpGs within region. Sidak P-value = Adjusted P-value for multiple comparisons across the epigenome.

Chromosome	Start	End	Min. P-value	#Probes	P-value	Sidak P-value	Gene
chr16	89349945	89352007	0.007374	10	1.30E-10	4.46E-08	*ANKRD11*
chr16	89437341	89439497	0.009317	12	1.46E-09	4.81E-07	*ANKRD11*
chr5	169658215	169659950	0.001232	16	1.20E-09	4.92E-07	NA
chr19	48674931	48676053	0.001771	6	1.62E-09	1.02E-06	*ZSWIM9*
chr2	168724333	168726607	0.01452	9	6.68E-09	2.09E-06	*B3GALT1*
chr17	77951408	77952463	0.002708	7	3.60E-09	2.43E-06	*TBC1D16*
chr13	113718453	113719440	2.67E-04	5	3.57E-09	2.57E-06	*MCF2L*
chr1	68516272	68517691	4.44E-04	15	7.42E-09	3.71E-06	*DIRAS3*
chr2	228497892	228499386	0.004894	6	8.25E-09	3.92E-06	*C2orf83*
chr17	74303724	74304023	9.65E-05	8	1.84E-09	4.38E-06	*QRICH2*
chr12	131271323	131272183	0.001841	3	7.46E-09	6.16E-06	NA
chr5	170883661	170884806	0.01135	6	9.94E-09	6.17E-06	*FGF18*
chr19	46917727	46918365	0.005363	3	6.03E-09	6.71E-06	NA
chr6	87804765	87806133	0.001279	8	1.37E-08	7.11E-06	*CGA*
chr15	69221572	69223018	0.007071	8	1.58E-08	7.78E-06	*MIR548H4*
chr1	192917420	192918151	0.002707	4	8.35E-09	8.11E-06	NA
chr17	17124521	17124969	1.55E-04	6	6.23E-09	9.87E-06	*PLD6; FLCN*
chr12	101752716	101754080	0.002954	5	2.17E-08	1.13E-05	*UTP20*
chr11	118501494	118502454	0.002753	6	1.90E-08	1.40E-05	*PHLDB1*
chr2	106959205	106959878	8.18E-04	7	1.35E-08	1.42E-05	NA