

UC Berkeley

UC Berkeley Previously Published Works

Title

A Pregnancy and Childhood Epigenetics Consortium (PACE) meta-analysis highlights potential relationships between birth order and neonatal blood DNA methylation

Permalink

<https://escholarship.org/uc/item/8xt9556r>

Journal

Communications Biology, 7(1)

ISSN

2399-3642

Authors

Li, Shaobo

Spitz, Natalia

Ghantous, Akram

et al.

Publication Date

2024

DOI

10.1038/s42003-023-05698-x

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

A Pregnancy and Childhood Epigenetics Consortium (PACE) meta-analysis highlights potential relationships between birth order and neonatal blood DNA methylation

Shaobo Li ¹, Natalia Spitz ², Akram Ghantous², Sarina Abrishamcar ³, Brigitte Reimann⁴, Irene Marques^{5,6}, Matt J. Silver ⁷, Sofía Aguilar-Lacasaña^{8,9,10}, Negusse Kitaba ¹¹, Faisal I. Rezwan ^{11,12}, Stefan Röder ¹³, Lea Sirignano¹⁴, Johanna Tuhkanen¹⁵, Giulia Mancano¹⁶, Gemma C. Sharp^{16,17}, Catherine Metayer¹⁸, Libby Morimoto ¹⁸, Dan J. Stein ¹⁹, Heather J. Zar^{19,20}, Rossella Alfano ⁴, Tim Nawrot⁴, Congrong Wang⁴, Eero Kajantie ^{21,22,23}, Elina Keikkala^{21,24}, Sanna Mustaniemi ^{21,24}, Justiina Ronkainen ²⁵, Sylvain Sebert ²⁵, Wnurinham Silva ²⁵, Marja Vääräsmäki^{21,24}, Vincent W. V. Jaddoe ^{5,6}, Robin M. Bernstein²⁶, Andrew M. Prentice ²⁷, Marta Cosin-Tomas^{8,9,10}, Terence Dwyer ²⁸, Siri Eldevik Håberg²⁹, Zdenko Herceg², Maria C. Magnus²⁹, Monica Cheng Munthe-Kaas³⁰, Christian M. Page^{29,31}, Maja Völker¹⁴, Maria Gilles³², Tabea Send³², Stephanie Witt ¹⁴, Lea Zillich ¹⁴, Luigi Gagliardi³³, Lorenzo Richiardi³⁴, Darina Czamara ¹⁵, Katri Räikkönen¹⁵, Lida Chatzi³⁵, Marina Vafeiadi ³⁶, S. Hasan Arshad^{37,38}, Susan Ewart³⁹, Michelle Plusquin⁴, Janine F. Felix ^{5,6}, Sophie E. Moore⁴⁰, Martine Vrijheid ^{8,10}, John W. Holloway ^{11,37}, Wilfried Karmaus ⁴¹, Gunda Herberth ¹³, Ana Zenclessen ^{13,42}, Fabian Streit ¹⁴, Jari Lahti ¹⁵, Anke Hüls^{3,43}, Thanh T. Hoang ⁴⁴, Stephanie J. London ⁴⁴ & Joseph L. Wiemels ^{1✉}

Higher birth order is associated with altered risk of many disease states. Changes in placenta and exposures to in utero growth factors with successive pregnancies may impact later life disease risk via persistent DNA methylation alterations. We investigated birth order with Illumina DNA methylation array data in each of 16 birth cohorts (8164 newborns) with European, African, and Latino ancestries from the Pregnancy and Childhood Epigenetics Consortium. Meta-analyzed data demonstrated systematic DNA methylation variation in 341 CpGs (FDR adjusted $P < 0.05$) and 1107 regions. Forty CpGs were located within known quantitative trait loci for gene expression traits in blood, and trait enrichment analysis suggested a strong association with immune-related, transcriptional control, and blood pressure regulation phenotypes. Decreasing fertility rates worldwide with the concomitant increased proportion of first-born children highlights a potential reflection of birth order-related epigenomic states on changing disease incidence trends.

Birth order, or the ordinal position of a child within their family, is associated with a wide variety of health outcomes. First-borns are at a higher risk for type 1 diabetes¹, high blood pressure², synovial sarcoma³, metabolic diseases⁴, immune diseases (including allergy⁵, eczema⁶, acute lymphoblastic leukemia^{7,8}, and lymphoma⁹). First-born children are at lower risks for other diseases including acute myeloid lymphoma¹⁰ and non-Hodgkin lymphoma¹¹. These risk associations are robust, being replicated in populations worldwide. The proportion of first-born children compared to later born children is increasing due to decreasing birth rates worldwide¹², suggest that some disease trends may be related to this changing demographic. Notably, most of the diseases listed above have exhibited increased incidence over the same time period as demographic changes leading to decreasing family size, for instance allergies and type 1 diabetes^{13,14}, suggesting that some proportion of the observed disease incidence trends can be attributed to this change.

Importantly, first-borns experience different gestational environments than their later-born siblings, as indexed using a variety of different biomarkers. These environments may impact later disease risk and support a biological basis for prenatal environmental conditions related to birth order. First-borns experience less sufficient placentation, higher estrogen levels, and lower insulin sensitivity, which could all contribute to subsequent post-birth disease risks^{15–18}. The means and mechanisms by which these factors (related to birth order) impact childhood outcomes are not currently understood but may be crucial to efforts at understanding disease etiology and prevention. Li et al. reported that DNA methylation using a genome-wide correlation analysis of array-based DNA methylation marks of sibling pairs born after a twin birth was more correlated than sibling pairs born before a twin birth from the same mother¹⁹. This study suggests that DNA methylation tends to be more consistent and stabilized for later born infants, subsequent to prior deliveries; however, this study did not examine the directionality of DNA methylation alterations after twin pregnancies. In another candidate gene study²⁰, the DNA methylation of genes in T-cell pathways were reported to be associated with birth order, and they could in turn affect immune functions of the newborn. However, these studies had small sample sizes and could not detect DNA methylation changes with birth order on a wider genomic scale.

Here we aimed to investigate associations between neonatal DNA methylation and birth order on a genome-wide scale for the first time that we are aware of, combining results from 16 cohorts from the Pregnancy and Childhood Epigenetics Consortium (PACE). The large number of studies allowed extensive replication and consistency of findings, yielding a veritable catalog of birth order associations. Investigating differentially methylated probes (DMPs), as well as differentially methylated regions (DMRs) in infants with different birth orders may provide mechanistic insights on how birth order could impact associated developmental differences and disease risks.

Methods

Participating cohorts. Sixteen cohorts from 12 countries (Germany, South Africa, Belgium, United Kingdom, Norway, Italy, Greece, Finland, Gambia, Spain, Netherlands, United States of America) were included in this study, including 8164 participants (Table 1). All studies used neonatal blood—for most this was derived from the umbilical cord, and for some from heel-prick blood spots. For a detailed description of each cohort, including DNA methylation extraction and data preprocessing steps, see Supplementary Note 1. Additional details on key birth characteristics particularly birthweight was published previously²¹.

Each cohort acquired individual site-specific ethics approval as well as informed consent. The overall analysis was approved by the University of Southern California Institutional Review Board in Health Science.

Definition of birth related variables. Birth order refers to the number of deliveries the mother had at the time of the subject's birth. It was coded as an ordinal variable (1, 2, 3, ...). Only singletons, and whose older siblings are also singletons, were included in this project, if such information is available. If multiple participants within a sample set were from the same family, only one of them was randomly included in this study to maintain independence of all study subjects. Miscarriages and abortions were not counted as delivery events. Stillbirth refers to fetal death at 20 weeks or later of pregnancy. If stillbirth information was available, it was included as a previous delivery.

DNA methylation measurement. Extraction of blood samples, isolation of genomic DNA, and DNA methylation array measurements were done separately by each cohort. See Supplementary Note 1 for a detailed description for each cohort. The Illumina450K array was used by 14 cohorts and the EPIC array by 8 cohorts.

Statistics and Reproducibility. Epigenome-wide association (EWAS) models were run in each cohort independently, with a prespecified pipeline using robust linear regression. If participants from a cohort included multiple ancestries, each ancestry was run separately. In total, there were 23 datasets, each including one ancestry from a specific cohort.

Briefly, winsorized DNA methylation beta value for each CpG was modeled as the dependent variable, with birth order (coded as 1, 2, 3,...) as a discrete independent variable. Covariates included child sex (male as 0, female as 1), technical variable to address potential batch effects, cell type proportional estimates based on the Salas et al. cord blood reference panel²², selection factor, maternal age (years), gestational age (weeks), birthweight (gram), and maternal smoking status (nonsmoker as 0, smoker as 1). Selection factor applies when there was selection on a phenotype to create the original DNA methylation dataset for each individual study—for instance leukemia status (case/control) in the CCLS study. Note that despite the selection factor, all children were not identified as such at birth—any conditions or diseases selected were diagnosed/developed later in childhood. The main model is as follows:

$$\begin{aligned} \text{Methylation } \beta \text{ value} \sim & \text{birth order (ordinal)} + \text{sex} + \text{gestational age (weeks)} \\ & + \text{Batch} + \text{selection factor} + \text{maternal age} + \text{birthweight} \\ & + \text{maternal smoking} + \text{deconvoluted cell proportion} \end{aligned}$$

In meta-analysis, CpGs on sex chromosomes, as well as those overlapping SNPs and probes with >5% minor allele frequency in the entire population, were not included. “IlluminaHumanMethylationEPICanno.ilm10b4.hg19”²³ was used to annotate CpGs including their locations, overlapping genes or closest genes, and their regulatory regions.

Meta-analysis of all cohorts was conducted using METAL²⁴ weighted by inverse of standard errors, assuming fixed-effects. There were 754,340 CpGs in the final analysis that were included in at least 1 cohort. A differentially methylated probe (DMP) was defined as a CpG with false discovery rate (FDR) adjusted *P* value < 0.05. Heterogeneity between cohorts was measured using heterogeneity *P* (*P*_{het}) value output by METAL. Differentially methylated regions (DMR) were identified using “ipdmr”²⁵ function from the ENmix²⁶ R package, with default parameters. Meta-analysis and shadow meta-analysis were done in two

Table 1 Description of participation cohorts.

Cohort	Country	Illumina array	Ethnicity	Sample size (n)	Mean birth weight (SD) (g)	Mean gestational age (SD) (g)	Mean maternal age (SD) (g)	Sex (n)	Birth order (n)		
LiNA	Germany	450 K	European	472	3434.97 (470.75)	39.78 (1.5)	30.58 (4.52)	M 249	1 317		
								F 223	2 116		
									≥4 7		
DCHS	South Africa	450 K	Black African/Mixed	117	3101.03 (512.36)	38.71 (1.77)	26.61 (5.56)	M 50	1 42		
								F 67	2 44		
									≥4 12		
	EPIC	African/Mixed	146	3090.03 (559.66)	38.99 (2.68)	27.51 (5.95)	M 79	1 49			
							F 67	2 52			
								≥4 13			
ENVIRONAGE	Belgium	450 K	European	182	3393.71 (486.92)	39.10 (1.66)	29.28 (4.41)	M 96	1 101		
								F 86	2 81		
									≥4 0		
	EPIC	European	326	3425.00 (482.00)	39.16 (1.63)	30.18 (4.31)	M 160	1 174			
							F 166	2 118			
								≥4 3			
POSEIDON	Germany	450 K	European	286	3415.08 (464.4)	39.21 (1.21)	31.58 (4.81)	M 136	1 148		
								F 150	2 113		
									≥4 7		
									≥4 3		
ALSPAC	United Kingdom	450 K	European	744	3487.03 (483.68)	39.56 (1.51)	29.72 (4.40)	M 363	1 354		
								F 381	2 278		
									≥4 27		
									≥4 27		
MoBa 1	Norway	450 K	European	984	3645.40 (544.00)	39.50 (1.60)	29.90 (4.40)	M 524	1 423		
								F 460	2 2		
									≥4 3		
124 ≥4	37	MoBa 3	Norway	450 K	European	238	3671.81 (555.48)	39.65 (1.66)	29.62 (4.41)	M 119	1 118
F 119	2 81										
	≥4 11										
Piccolipiù	Italy	450 K	European	98	3221.07 (433.28)	39.58 (1.57)	33.32 (4.44)	M 53	1 45		
								F 45	2 44		
									≥4 1		
									≥4 1		
Rhea	Greece	450 K	European	91	3277.80 (442.32)	38.54 (1.30)	29.78 (4.74)	M 50	1 29		
								F 41	2 40		
									≥4 0		
									≥4 0		
PREDO	Finland	450 K	European	822	3546.00 (556.60)	39.80 (1.60)	33.30 (5.80)	M 433	1 248		
								F 389	2 369		
									≥4 51		
	EPIC	European	147	3457.10 (517.40)	39.80 (1.40)	32.10 (4.90)	M 75	1 62			
							F 72	2 57			
								≥4 8			
HERO-G	Gambia	EPIC	African	115	3020.00 (350.00)	39.80 (1.20)	30.70 (7.30)	M 68	1 12		
								F 47	2 21		
									≥4 13		
									≥4 69		
INMA	Spain	450 K	European	383	3271.00 (419.56)	39.77 (1.39)	30.36(4.09)	M 195	1 170		
								F 188	2 101		
									≥4 14		

Table 1 (continued)

Cohort	Country	Illumina array	Ethnicity	Sample size (n)	Mean birth weight (SD) (g)	Mean gestational age (SD) (g)	Mean maternal age (SD) (g)	Sex (n)	Birth order (n)
Generation R	Netherlands	450 K	European	1249	3550.00 (509.00)	40.20 (1.50)	31.70 (4.20)	M 636	≥4 0
								F 613	1 755
									2 372
Finngedi	Finland	EPIC	European	527	3703.00 (470.50)	39.93(1.32)	31.98 (5.28)	M 255	≥4 16
								F 272	1 245
									2 150
Isle of wight (IOW)	United Kingdom	450 K	European	104	3430.99 (517.40)	39.40 (1.66)	23.24 (2.59)	M 46	≥4 80
								F 58	1 66
									2 25
		EPIC	European	570	3430.00 (530.00)	40.00 (1.50)	27.20 (5.04)	M 274	≥4 2
								F 296	1 234
									2 203
California Childhood Leukemia Study (CCLS)	United States	450 K	European	163	3562.50 (581.26)	39.21 (1.90)	31.45 (5.75)	M 95	≥4 0
								F 68	1 72
									2 54
		450 K	Latino	133	3466.80 (663.16)	39.22 (2.40)	27.18 (6.19)	M 79	≥4 7
								F 54	1 56
									2 35
EPIC	European	98	3396.50 (554.64)	39.30 (2.29)	30.38 (6.24)	M 58	≥4 16		
						F 40	1 50		
							2 30		
EPIC	Latino	169	3366.44 (617.21)	39.26 (2.18)	26.58 (6.17)	M 93	≥4 5		
						F 76	1 73		
							2 53		
									≥4 17

different institutions, one from USA (USC), the other France (IARC). Comb-p²⁷ was also used to identify DMRs, using meta-analyzed DMPs, to test the robustness of the “ipdmr” function.

Gene pathway enrichment analyses of DMPs were performed with “methylGSA” R package²⁸, using all FDR adjusted significant CpGs from the meta-analysis as inputs. Gene Ontology (GO)²⁹ and Kyoto Encyclopedia of Genes and Genomes (KEGG)³⁰ databases were both used and pathways with FDR-corrected *P* value < 0.05 were considered significant. Enrichment analyses of DMRs were carried out using the database for annotation, visualization and integrated discovery (DAVID)^{31,32}, with genes overlapping DMRs as input, focusing on GO and KEGG results.

We investigated if there was a significant increase or decrease in overlap with transcription factor (TF) binding sites among top hits. TF data for 161 transcription factors from 91 cell types were downloaded from the ENCODE project (wgEncodeRegTfbsClusteredV3.bed). The number of CpGs among significant hits overlapping TF binding sites were compared to that of array-wide CpGs with the Fisher’s exact test³³.

EWAS Open Platform³⁴ was used to conduct trait enrichment analysis, in order to identify if significant CpGs from our study were associated with other phenotypes included in EWAS Atlas³⁵. Associations of DNA methylation levels in the blood and brain were inferred using BECon³⁶. Expression levels of genes related to DMPs in different tissues were queried on the GTEx portal³⁷.

Finally, the associations between methylation levels of significant CpGs and expressions of nearby genes (cis-expression quantitative trait methylation, cis-eQTM) in the blood were

queried from published results³⁸ from the Human Early Life Exposome (HELIX) project³⁹.

Sensitivity analyses. We conducted several sensitivity analyses to test the robustness of our results.

For the top 20 CpGs from the meta-analysis, we conducted leave-one-out (LOO) analyses, excluding one cohort at a time, to observe if the results were driven by one specific cohort. Forest plots showing LOO results were plotted with ‘forestplot’ function in the ‘forestplot’ R package⁴⁰.

To investigate whether the associations between DNA methylation and birth order were different in different ancestries, we repeated the meta-analyses in European participants (*n* = 7484) and African participants (*n* = 378) separately. Ancestry specific analysis was not run in Latinos because Latino participants had the smallest sample size, and they all came from one cohort.

In addition, miscarriages and abortions have arguably smaller physiological effects than full term pregnancy; however, their effects on neonatal DNA methylation of future babies are unclear. Therefore, for cohorts with miscarriage/abortion information available, we carried out sensitivity analysis counting miscarriages and abortions as a delivery.

Lastly, maternal weight gain was reported to be associated with placental DNA methylation alterations⁴¹, which in turn could affect neonatal methylation. To test this, we adjusted for maternal weight gain as an additional sensitivity analysis in cohorts with this information.

Reporting summary. Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Meta-analysis identified significant CpG probes associated with birth order. Our meta-analysis included all 23 datasets from 16 cohorts identified 341 CpGs differentially methylated at FDR adjusted P value < 0.05 (Fig. 1, Supplementary Data 1). In these and all data presented, positive coefficients refer to higher (hyper-) DNA methylation with later birth order compared to earlier, and negative coefficients refer to lower (hypo-) methylation with later birth order compared to earlier. The most significant CpG (cg09249800, FDR adjusted P value = 7.24×10^{-6}) was in a CpG island in the gene body of *ACOT7*. The second most significant CpG was located in the transcription start site (TSS) of *LOC650226*, located in a Chromosome 7 peak overlapping shore and island regions of *LOC650226* and *ZNF727* genes (Fig. 1, Supplementary Data 2). CpG sites in the promoter regions of *FAM169A* (cg04654716, FDR adjusted P value = 4.40×10^{-4}) and *LIF* (cg19539004, FDR adjusted P value = 4.40×10^{-4}) were also among the top hits. See Table 2 for annotation of all significant hits including their genomic coordinates, meta-analysis I^2 value and additional outputs. We also computed the top associations with a statistical model examining first birth versus all subsequent births as a group (bivariate analysis) (Supplementary Data 3).

A total of 1 KEGG and 43 GO pathways were enriched among these 341 DMPs (Supplementary Data 4), including those involved in cell growth development (germ cell development, multicellular organism reproduction, growth factor activity etc.) and leukocyte activation and migration (leukocyte transendothelial migration, positive regulation of B cell activation, regulation of leukocyte chemotaxis etc.).

We collected data on all 161 transcription factors (TFs) from ENCODE ChIP-seq database and tested if birth order related CpGs were more or less likely to overlap with TF bindings sites. As a result, 10 TF binding sites (MAZ, CTCF, POLR2A, RAD21, EZH2, ZBTB7A, GATA3, GATA2, TAL1, POU5F1) were enriched, while 13 (ATF1, CREB1, NFYA, GTF2F1, CEBPD, ELK1, RFX5, TAF7, RELA, KDM5B, E2F4, PML, SIN3AK20)

were depleted (i.e., significantly under-represented) among these CpGs.

Trait enrichment analysis suggested that birth order-associated hits were also associated with 69 other traits (Supplementary Data 5), the top 4 of which were all immune-related phenotypes including allergic sensitization (P value = 5.90×10^{-96}), fractional exhaled nitric oxide (P value = 1.53×10^{-69}), childhood asthma (P value = 3.42×10^{-60}), and atopy (P value = 3.42×10^{-60}). Smoking (P value = 1.98×10^{-38}), maternal smoking (P value = 5.28×10^{-23}), down syndrome (P value = 2.58×10^{-20}) and neurodevelopmental presentations and congenital anomalies (ND/Cas) (P value = 8.31×10^{-18}) were also among top enriched traits.

Forty out of the 341 significant CpGs (11.73%) were previously reported to be cis-eQTM in blood (Supplementary Data 6), with some of the CpGs associated with multiple transcripts. This proportion was much higher than that for all CpGs on the 450 K array (2.37%). For example, the methylation level of cg04654716 was reported to be positively associated with *FAM169A* expression level (eQTM P value = 6.24×10^{-7}).

DNA methylation levels are often tissue-specific, and because we analyzed DNA from blood, we wanted to evaluate whether we could infer DNA methylation levels of these 341 birth-order related CpGs in the brain, because trait hits above seemed to be very relevant to neural functions. By querying published dataset by Edgar et al.³⁶ which reported concordance of DNA methylation in the blood and the brain, 277 birth order related CpGs had blood-brain association data available (Supplementary Data 7), and 113 (40.79%) CpGs among them had an absolute Spearman correlation coefficient bigger than 0.2. Interestingly most of the genes we mentioned as top birth order-associated hits exhibited enhanced gene expression in brain tissues compared with other tissues (including *PRRT1*, *PLEKHB1*, *ACOT7*, *FAM169A*, *ZBED9*) (Supplementary Fig. 1–5).

Differentially methylated regions associated with birth order.

We identified 1,107 DMRs associated with birth order (Table 3, Supplementary Data 8). Functional annotations with genes overlapping these DMRs by DAVID³¹ identified 17 significant pathways (adjusted P value < 0.05). Eleven (64.70%) of them are related to DNA transcription regulation, 3 of them likely related

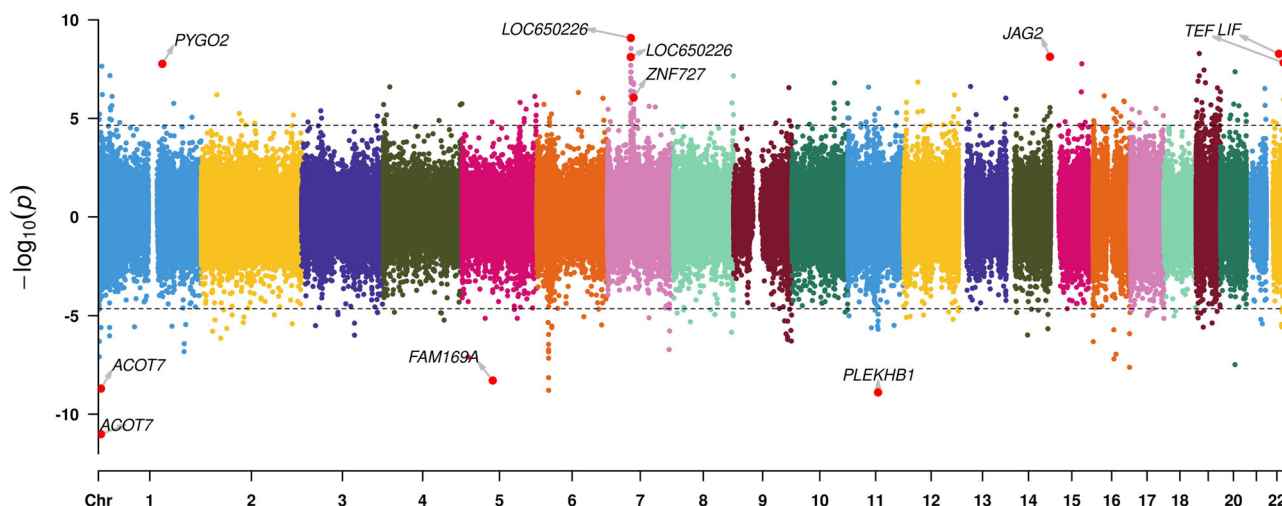


Fig. 1 Bi-directional Miami plot showing associations between DNA methylation and birth order. Bi-directional Miami plot showing the results of meta-analysis of the association between DNA methylation and birth order, adjusting for sex, maternal age, gestational age, birthweight, maternal smoking, batch effects, selection factor, and cell proportions. Directions of the associations were shown on the Y-axis, with positive associations above $Y = 0$ and negative associations below. Threshold of significance after false-discovery rate (FDR) correction is shown in a dashed horizontal line. A total of 341 CpGs were significant after FDR multiple correction, the threshold of which was shown with a dashed line.

to transcription regulation (17.65%), and only 2 (GO:0098978 glutamatergic synapse, and GO:0005887 integral component of plasma membrane) (11.76%) are not related to this function (Supplementary Data 9).

Sensitivity analyses. We ran several sensitivity analyses to test the robustness of our results. We did leave-one-out analyses for top 20 hits from our analysis, each time excluding one dataset from the meta-analysis, to test if results were heavily influenced by any one dataset. For all top 20 CpGs, leaving datasets out one by one did not change the significance of our results. Effect sizes were all in the same direction as the main model, and none of the 95% confidence interval (CI) of the meta-analysis estimates crossed zero (Supplementary Figs. 6–9).

Since our participants were from different ancestral groups but predominantly European, we conducted meta-analyses in European ($n = 7484$) and African ($n = 378$) ancestries separately to observe ancestry-specific birth order related CpGs (Supplementary Data 1, Supplementary Data 10). In European participants, there were 316 significant CpGs after multiple correction, while in African participants alone, only 1 CpG remained significant (Supplementary Data 2), likely due to small sample size. 117 of the 341 significant CpGs from the main model were also significant in European participants, and all CpGs had the same direction of effects (Supplementary Data 1). However, in participants of African ancestry, 273 CpGs out of the 341 CpGs (80.06%) were in the same direction as the main model, and none of these 341 CpGs were significant in African participants alone (Supplementary Data 1).

We also controlled for maternal weight gain as an additional variable, and results were highly consistent, including the 341 significant CpGs from the original model (Fig. 2a). We also counted abortions/ miscarriages as a birth event, and results were similar to our main models (Fig. 2b).

Discussion

In this study, we combined multiple cohorts from 12 countries, including participants of European, African and Latino self-reported ancestries, and identified 341 CpGs whose DNA methylation levels were associated with birth order. This was the first multi-cohort large-scale EWAS study investigating the associations between neonatal DNA methylation and birth order. As no single cohort was specifically designed to examine DNA methylation and birth order our results may be considered exploratory, however the strength of the PACE Consortium allows confirmatory replication and validation.

Birth order has been associated with multiple diseases and does not have a genetic cause. Therefore, it is of interest to investigate whether epigenetic alterations, especially DNA methylation observable at birth, is associated with birth order. These epigenetic alterations may mediate the impact of birth order on disease risk, and can serve as a roadmap of candidate biomarkers to investigate such risk. To establish a robust set of birth order-associated biomarkers, we conducted an EWAS meta-analysis including multiple datasets from cohorts around the world. We found numerous CpGs differentially methylated in relation to birth order, with some associated with gene expression in tissues that have birth order disease associations such as the brain, immune system, and cardiovascular system. The dramatic fall in fertility rates worldwide over the preceding decades and projections for the future are leading to a higher proportion of first-born individuals with certain future continuation of such trends; in addition the contribution of variance in DNA methylation impacted by birth order and its associated diseases is of strong

interest to the Developmental Origin of Health and Disease (DOHaD) community.

The most significant CpG was cg09249800 (adjusted P value 7.24×10^{-6}), in the gene body region of *ACOT7*. The encoded protein hydrolyzes palmitoyl-coenzyme A (palmitoyl-CoA), and was reported to be associated with mesial temporal lobe epilepsy⁴². Interestingly, a previous GWAS study⁴³ identified a SNP (rs11121611) within *ACOT7* to be associated with “asthma exacerbation measurement, response to corticosteroid”. However, cg09249800 (chr1:6341287, Hg19) was about 25 kb upstream of rs11121611 (chr1:6367119, Hg19), and it was not reported to be a cis-eQTM of *ACOT7* (Supplementary Data 6).

While cg09249800 was the most statistically significant association, its effect size (-2.8×10^{-3}) was nearly 4 times smaller than strongest effect size CpG which was cg26865747, (coefficient = 0.0105), proximal to the *SCAND3* gene, a zinc finger transcription factor implicated in tumor proliferation and invasion⁴⁴. Significant individual CpG sites ranged in effect sizes from 0.0001 to 0.01, namely over two orders of magnitude, and 70 of the 341 CpGs had larger effect sizes than the most significant single CpG site at *ACOT7*. Other significant hits were of interest. For example, there was a prominent cluster of 10 CpGs overlapping the *LOC650226* and *ZNF727* genes. In addition, all significant CpGs overlapping *ZNF727* were reported to be cis-eQTMs, meaning their DNA methylation levels were associated with expression levels of the *ZNF727* gene. The reasons for their associations with birth order requires further investigation, although it is interesting that all 10 CpGs were also reported by Häberg et al.⁴⁵ to have a significantly lower DNA methylation level in babies born with assisted reproductive technology (ART) compared to naturally conceived babies (FDR adjusted P value $< 9.86 \times 10^{-5}$). Interestingly in our study, later-borns were more methylated than first-borns in this region. It was not clear why DNA methylation patterns vary in this manner. A potential explanation is that the later order a child was born, the more established the pregnancy process becomes including placentation, leading to more stable nutrition status promoting physiologic homeostatic DNA methylation patterns. We did not evaluate the relationship of *ZNF727* DNA methylation to post-natal outcomes, but such an effort would be valuable, particularly in the modern era as family size in some countries has decreased compared to historical trends. Either way, additional data is needed to elucidate answers to whether DNA methylation alterations may mediate some of the disease associations ascribed to birth order.

The CpG site cg04654716 (effect size -0.0023 , adjusted P value 5.12×10^{-9}) in the transcription start site of *FAM169A* is also of interest. Similar to CpGs overlapping *ZNF727*, cg04654716 was also reported to be a cis-eQTM, whose DNA methylation level was positively related to *NSA2* expression level. SNPs in *NSA2* were also associated with metabolism-related traits in many GWAS studies (low density lipoprotein cholesterol measurement^{46,47}, total cholesterol measurement^{46,47}, linoleic acid measurement⁴⁸, omega-6 polyunsaturated fatty acid measurement^{48,49}, and HMG CoA reductase inhibitor use measurement⁵⁰). Interestingly, metabolic function was also related to birth order⁴, and its causal pathway is worth further investigation.

Most of the participants in this study were of European ancestry, and unsurprisingly, in European participants alone, the effect sizes of all significant CpGs were in the same directions as the main model, while in African participants, about 80% significant CpGs were in the same direction (Supplementary Data 2). Further investigation on how these CpGs were related to birth order in other ancestries including Asians and Latinos is required.

Table 3 Top 10 DMRs associated with birth order from meta-analysis.

chr	Start	End	Width	P	FDR adjusted P value	# of probes	Gene symbol	Distance2TSS
chr6	28601270	28602543	1274	3.70E-36	4.17E-33	11	ZBED9	-46158
chr7	56514964	56516425	1462	9.68E-33	5.46E-30	11	LOC650226	0
chr11	73357018	73357612	595	9.44E-25	3.55E-22	9	PLEKHB1	0
chr6	32120202	32121261	1060	3.25E-22	9.17E-20	26	PRRT1	0
chr7	63360692	63361617	926	2.13E-20	4.79E-18	6	LINC01005	128863
chr7	63385989	63387147	1159	3.36E-19	5.60E-17	8	LINC01005	103333
chr7	57483672	57484819	1148	3.48E-19	5.60E-17	6	MIR3147	10941
chr7	63505637	63506148	512	4.39E-19	6.18E-17	5	ZNF727	0
chr7	57471758	57473294	1537	6.07E-19	7.60E-17	6	MIR3147	0
chr9	131154346	131156014	1669	5.21E-18	5.87E-16	7	MIR219B	0

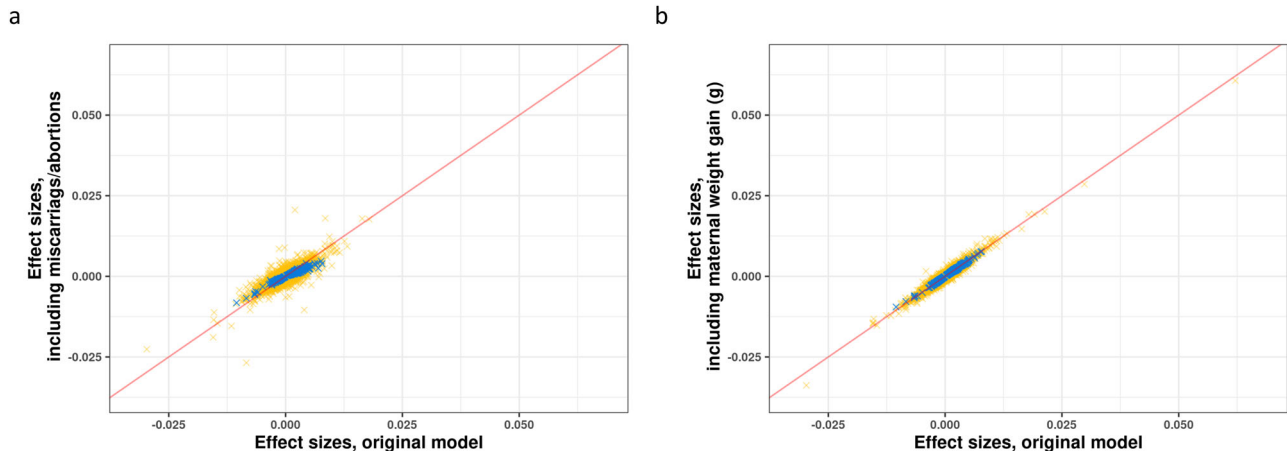


Fig. 2 Comparisons of results from sensitivity models to the main model. **a** Effect sizes from the main models (X-axis) plotted against effect sizes from models including maternal weight gain as an additional covariate (Y-axis). $Y=X$ line was plotted in a red line. The 341 significant CpGs from the main models were colored blue, and all other non-significant CpGs were colored yellow. **b** Effects sizes from the main models (X-axis) plotted against effect sizes from models counting miscarriage/abortion as a birth event (Y-axis), similar to that of **(a)**.

In trait enrichment analyses, birth order related CpGs were also associated with 69 traits (Supplementary Data 5), especially allergy-related features including allergic sensitization, childhood asthma, atopy, serum immunoglobulin E levels, allergic asthma, wheeze, respiratory allergies, primary Sjögren's Syndrome, systemic lupus erythematosus, multiple sclerosis and cow's milk allergy. Interestingly, first-borns were also reported to have a higher risk of allergy⁵¹ and eczema was previously reported to be associated with birth order^{6,51}. Other immune related features were also significantly associated with birth order, for example, psoriasis, B acute lymphoblastic leukemia (B-ALL) and acute chorioamnionitis (aCA). Among those, B-ALL was reported to be more common in first-borns⁷.

Several other traits previously reported to be related to birth order were also identified, including blood pressure² (cardiac autonomic responses (deceleration capacity), diastolic blood pressure, systolic blood pressure, atherosclerosis, preeclampsia, maternal hypertensive disorders in pregnancy), metabolism⁴ (serum liver enzyme levels (alanine aminotransferase, ALT), serum liver enzyme levels (gamma-glutamyl transferase, GGT), metabolic trait, hepatic steatosis, hepatic fat), birth weight⁵² and body mass index (BMI)⁵³. The top association in our DMR analysis was *ZBED9*, recently identified as a regulatory gene for blood pressure⁵⁴.

Additional traits were associated with birth order related CpGs in our study, but have not previously reported to be associated with birth order. These include abnormal karyotype related traits (Down syndrome, Klinefelter syndrome), and several neural function-related traits (soluble tumor necrosis factor receptor 2 levels in plasma, neurodevelopmental presentations and

congenital anomalies (ND/CAs), schizophrenia, myalgic encephalomyelitis/chronic fatigue syndrome, leukoaraiosis). More investigation could reveal whether these traits are also related to birth order, and how birth order related epigenetic changes might contribute to such relationships.

Neurological traits previously assessed in relationship to birth order include intelligence, in which first-borns tended to display higher levels⁵⁵⁻⁵⁷. As our samples were collected from blood, these enriched neural related traits also led us to investigate the consistency of methylation levels of birth order associated CpGs in blood and brain. Of all the significant CpGs whose blood-brain association data were available, 40.79% had modest to strong associations (absolute Spearman correlation coefficient >0.2). When we confined our trait enrichment analysis to CpGs whose methylation levels were highly correlated in blood and brain only, similar traits were enriched, and neurodevelopmental presentations and congenital anomalies (ND/CAs) became the second most enriched phenotype (Supplementary Data 5).

We identified 1,107 DMRs associated with birth order. Enrichment analysis of these genes showed that almost all significant pathways were related to regulation of gene transcription. More work is needed to understand what proteins were regulated and in which direction to elucidate more specific information that could impact human development. Only 2 enriched pathways were not transcription related, one of them being the glutamatergic synapse pathway. Glutamatergic synapse is involved in neural network development, and is essential for transferring and processing information⁵⁸, which may contribute to the association between birth order and intelligence.

There were some shortcomings with this study. While investigating birth order, we examined unrelated individuals, instead of same-family siblings. This inevitably introduced noise and decreased the reliability of our findings due to genetic, socio-economic status, and cultural differences between study participants both within and across study cohorts. The choice to remove any genetically related study subjects (by family relationships) maintains the independence of every study subject, randomizing unmeasured confounders whereas we statistically controlled for key known confounders—including sex, cell type distributions, maternal age, gestational age, birthweight, and maternal tobacco use (when available). We were not able to control for socio-economic status characteristics (such as maternal education or family income), characteristics which will impact disease risk and potentially DNA methylation profiles. Moreover, our study was designed to identify associations only, and is not capable of demonstrating causality or mediation by DNA methylation and birth order-related diseases. Instead, the data provides a catalog of candidates for future research. Also, despite the worldwide scope and large number of studies in the PACE Consortium, we lack extensive data on some race/ethnic groups particularly non-Europeans including Latinos, Africans and Asians. It would be of interest to investigate how birth order is associated with DNA methylation changes in these groups, and how they vary from the conclusions in this study. Another weakness is the variability of average family size among the various PACE consortium studies (which vary with regards to fertility rates) which might affect power to detect effects from higher birth orders. Also, our focus on neonatal blood exclusively may limit the discovery aspects of our data for other tissues, along with the limited coverage of the epigenome afforded by the Illumina array platforms used. Strengths of this study include the large sample size afforded by the PACE on a common analysis platform, and the presumed consistency in our main predictor variable (birth order) which should have a universal description worldwide and similar physiologic impacts across countries. The use of cord blood is also a major strength in that a molecular phenotype was captured before onset of many of the associated traits. The consistency of our top CpG hits across cohorts argues for true and meaningful associations which may prove to have further resources for the maternal/fetal health research community.

We note that first-borns, compared to their later-born siblings, have a variety of postnatal environmental differences which may also impact disease risk. First-borns are typically exposed to infectious agents later in their childhood development; indeed, birth order was often used as a proxy for infection timing^{59–61}. Such postnatal experiences including child rearing practices and postnatal infections are commonly conjectured to be mediators for birth order's health and disease impacts. These factors are not likely to be related to pre-birth environments and are not a subject of the current analysis but are important postnatal mediators. Future studies should robustly evaluate both the prenatal and postnatal mediators of birth order on disease risk—including DNA methylation at birth.

In conclusion, our results from multiple datasets showed with high confidence that birth order has a widespread and consistent association with DNA methylation in the cord blood of newborns. These differences provide a catalog of associations which can be assessed as causal mediators in the etiology of health conditions related to birth order.

Data availability

Blood samples and raw genetic data of neonatal subjects from each cohort are governed by their respective institutions and/or government agencies, and mostly could not be shared publicly without specific approvals. For example, for data from first author cohort,

California Childhood Leukemia Study (CCLS), we respectfully are unable to share raw, individual genetic data freely with other investigators. Should we be contacted by other investigators who would like to use the data; we will direct them to the California Department of Public Health Institutional Review Board to establish their own approved protocol to utilize the data, which can then be shared peer-to-peer.

Received: 19 May 2023; Accepted: 12 December 2023;

Published online: 09 January 2024

References

- Cardwell, C. R. et al. Birth order and childhood type 1 diabetes risk: a pooled analysis of 31 observational studies. *Int. J. Epidemiol.* **40**, 363–374 (2011).
- Wells, J. C. K. et al. Associations of birth order with early growth and adolescent height, body composition, and blood pressure: prospective birth cohort from Brazil. *Am. J. Epidemiol.* **174**, 1028–1035 (2011).
- Wiemels, J. L. et al. Birth characteristics and risk of early-onset Synovial Sarcoma. *Cancer Epidemiol. Biomark. Prev.* **29**, 1162–1167 (2020).
- Siervo, M., Horta, B. L., Stephan, B. C. M., Victora, C. G. & Wells, J. C. K. First-borns carry a higher metabolic risk in early adulthood: evidence from a prospective cohort study. *PLoS ONE* **5**, e13907 (2010).
- Karmaus, W. & Botezan, C. Does a higher number of siblings protect against the development of allergy and asthma? A review. *J. Epidemiol. Community Health* **56**, 209–217 (2002).
- Hughes, A. M. et al. Eczema, birth order, and infection. *Am. J. Epidemiol.* **167**, 1182–1187 (2008).
- Hjalgrim, L. L. et al. Birth weight and risk for childhood leukemia in Denmark, Sweden, Norway, and Iceland. *J. Natl Cancer Inst.* **96**, 1549–1556 (2004).
- Westergaard, T. et al. Birth characteristics, sibling patterns, and acute leukemia risk in childhood: a population-based cohort study. *J. Natl Cancer Inst.* **89**, 939–947 (1997).
- Jones, S. J. et al. Birth order, sibship size, childhood environment and immune-related disorders, and risk of lymphoma in lymphoid cancer families. *Cancer Epidemiol. Biomark. Prev.* **29**, 1168–1178 (2020).
- Von Behren, J. et al. Birth order and risk of childhood cancer: a pooled analysis from five US States. *Int. J. Cancer* **128**, 2709–2716 (2011).
- Grulich, A. E. et al. Birth order, atopy, and risk of non-hodgkin lymphoma. *J. Natl Cancer Inst.* **97**, 587–594 (2005).
- Roser, M. Fertility Rate. *Our World Data* (2014).
- Shin, Y. H. et al. Global, regional, and national burden of allergic disorders and their risk factors in 204 countries and territories, from 1990 to 2019: a systematic analysis for the Global Burden of Disease Study 2019. *Allergy* **78**, 2232–2254 (2023).
- Imperatore, G., Mayer-Davis, E. J., Orchard, T. J. & Zhong, V. W. Prevalence and Incidence of Type 1 Diabetes among Children and Adults in the United States and Comparison with Non-U.S. Countries. in *Diabetes in America* (eds. Cowie, C. C. et al.) (National Institute of Diabetes and Digestive and Kidney Diseases (US), 2018).
- Gluckman, P. D. & Hanson, M. A. Maternal constraint of fetal growth and its consequences. *Semin. Fetal Neonatal Med.* **9**, 419–425 (2004).
- Ayyavoo, A., Savage, T., Derraik, J. G. B., Hofman, P. L. & Cutfield, W. S. First-born children have reduced insulin sensitivity and higher daytime blood pressure compared to later-born children. *J. Clin. Endocrinol. Metab.* **98**, 1248–1253 (2013).
- Bernstein, L. et al. Higher maternal levels of free estradiol in first compared to second pregnancy: early gestational differences. *J. Natl Cancer Inst.* **76**, 1035–1039 (1986).
- Panagiotopoulou, K. et al. Maternal age, parity, and pregnancy estrogens. *Cancer Causes Control* **1**, 119–124 (1990).
- Li, S. et al. Twin birth changes DNA methylation of subsequent siblings. *Sci. Rep.* **7**, 8463 (2017).
- Nilormee, O. et al. Maternal DNA methylation in second half of pregnancy in Th1, Th2, Th17 and Treg pathway genes differs with birth-order. *J. Allergy Clin. Immunol.* **139**, AB376 (2017).
- Küpers, L. K. et al. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nat. Commun.* **10**, 1893 (2019).
- Salas, L. A. et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol.* **19**, 64 (2018).
- IlluminaHumanMethylationEPICanno.ilm10b4.hg19. *Bioconductor* <http://bioconductor.org/packages/illuminaHumanMethylationEPICanno.ilm10b4.hg19/>.
- Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- Xu, Z., Xie, C., Taylor, J. A. & Niu, L. ipDMR: identification of differentially methylated regions with interval *P*-values. *Bioinformatics* **37**, 711–713 (2021).

26. Xu, Z., Niu, L. & Taylor, J. ENmix: Quality control and analysis tools for Illumina DNA methylation BeadChip. <https://doi.org/10.18129/B9.bioc.ENmix> (2022).
27. Pedersen, B. S., Schwartz, D. A., Yang, I. V. & Kechris, K. J. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated *P*-values. *Bioinformatics* **28**, 2986–2988 (2012).
28. Ren, X. & Kuan, P. F. methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. *Bioinformatics* **35**, 1958–1959 (2019).
29. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
30. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
31. Sherman, B. T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkac194> (2022).
32. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
33. Lee, S.-T. et al. Epigenetic remodeling in B-cell acute lymphoblastic leukemia occurs in two tracks and employs embryonic stem cell-like signatures. *Nucleic Acids Res.* **43**, 2590–2602 (2015).
34. Xiong, Z. et al. EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. *Nucleic Acids Res.* **50**, D1004–D1009 (2022).
35. Li, M. et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res.* **47**, D983–D988 (2019).
36. Edgar, R. D., Jones, M. J., Meaney, M. J., Turecki, G. & Kobor, M. S. BECon: a tool for interpreting DNA methylation findings from blood in the context of brain. *Transl. Psychiatry* **7**, e1187–e1187 (2017).
37. Carithers, L. J. et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv. Biobank.* **13**, 311–319 (2015).
38. Ruiz-Arenas, C. et al. Identification of autosomal cis expression quantitative trait methylation (cis eQTM) in children's blood. *eLife* **11**, e65310 (2022).
39. Maitre, L. et al. Human Early Life Exposome (HELIX) study: a European population-based exposome cohort. *BMJ Open* **8**, e021311 (2018).
40. Gordon, M., Lumley T. Advanced Forest Plot Using 'grid' Graphics. *CRAN Repository*. <https://cran.r-project.org/web/packages/forestplot/forestplot.pdf> (2022).
41. Shrestha, D., Ouidir, M., Workalemahu, T., Zeng, X. & Tekola-Ayele, F. Placental DNA methylation changes associated with maternal pregnancy BMI and gestational weight gain. *Int. J. Obes.* **44**, 1406–1416 (2020).
42. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
43. Hernandez-Pacheco, N. et al. Genome-wide association study of inhaled corticosteroid response in admixed children with asthma. *Clin. Exp. Allergy J. Br. Soc. Allergy Clin. Immunol.* **49**, 789–798 (2019).
44. Zhang, X. et al. ZNF452 facilitates tumor proliferation and invasion via activating AKT-GSK3 β signaling pathway and predicts poor prognosis of non-small cell lung cancer patients. *Oncotarget* **8**, 38863–38875 (2017).
45. Häberg, S. E. et al. DNA methylation in newborns conceived by assisted reproductive technology. *Nat. Commun.* **13**, 1896 (2022).
46. Wojcik, G. L. et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).
47. Zhou, L. et al. A genome wide association study identifies common variants associated with lipid levels in the Chinese population. *PLoS ONE* **8**, e82420 (2013).
48. Borges, M. C. et al. Role of circulating polyunsaturated fatty acids on cardiovascular diseases risk: analysis using Mendelian randomization and fatty acid genetic association data from over 114,000 UK Biobank participants. *BMC Med.* **20**, 210 (2022).
49. Richardson, T. G. et al. Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. *PLoS Biol.* **20**, e3001547 (2022).
50. Sakae, S. et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
51. Bernsen, R. M. D., de Jongste, J. C. & van der Wouden, J. C. Birth order and sibship size as independent risk factors for asthma, allergy, and eczema. *Pediatr. Allergy Immunol.* **14**, 464–469 (2003).
52. Bohn, C. et al. Birth weight increases with birth order despite decreasing maternal pregnancy weight gain. *Acta Paediatr. Oslo Nor.* **110**, 1218–1224 (2021).
53. Yokoyama, Y. et al. Twin's Birth-Order Differences in Height and Body Mass Index From Birth to Old Age: A Pooled Study of 26 Twin Cohorts Participating in the CODATwins Project. *Twin Res. Hum. Genet.* **19**, 112–124 (2016).
54. Akbarzadeh, M. et al. The AGT epistasis pattern proposed a novel role for ZBED9 in regulating blood pressure: Tehran Cardiometabolic genetic study (TCGS). *Gene* **831**, 146560 (2022).
55. Belmont, L. & Marolla, F. A. Birth order, family size, and intelligence. *Science* **182**, 1096–1101 (1973).
56. Kristensen, P. & Bjerkedal, T. Explaining the relation between birth order and intelligence. *Science* **316**, 1717 (2007).
57. Sulloway, F. J. Psychology. Birth order and intelligence. *Science* **316**, 1711–1712 (2007).
58. Di Maio, V. The glutamatergic synapse: a complex machinery for information processing. *Cogn. Neurodyn* **15**, 757–781 (2021).
59. Kuper, H. et al. Birth order, as a proxy for age at infection, in the etiology of hepatocellular carcinoma. *Epidemiol. Camb.* **11**, 680–683 (2000).
60. Blaser, M. J., Chyou, P. H. & Nomura, A. Age at establishment of *Helicobacter pylori* infection and gastric carcinoma, gastric ulcer, and duodenal ulcer risk. *Cancer Res.* **55**, 562–565 (1995).
61. Westergaard, T. et al. Birth order, sibship size and risk of Hodgkin's disease in children and young adults: a population-based study of 31 million person-years. *Int. J. Cancer* **72**, 977–981 (1997).

Acknowledgements

We thank all the participants and their families for this study. We also acknowledge all funding agencies of cohorts involved in this study, as mentioned in the supplementary methods. The overall study at University of Southern California was funded by the National Institute of Environmental Health Sciences (P01ES018172, R01ES09137, R01CA185058), the National Cancer Institute (R01CA175737), and the US Environmental Protection Agency (RD83451101, RD83615901).

Author contributions

Conceptualization, methodology: J.L.W., S.J.L., S.L., A.G.; Meta-analysis and Shadow analysis: S.L., N.S.; Cohort-specific analysis: S.L., N.S., A.G., S.A., B.R., I.M., M.J.S., S.A.-L., N.K., F.I.R., S.R., L.S., J.T., G.M., G.C.S., C.M., L.M., D.J.S., H.J.Z., R.A., T.N., C.W., E. Kajantie, E. Keikkala, S.M., J.R., S.S., W.S., M.V., V.W.V.J., R.M.B., A.M.P., M.C.-T., T.D., S.E.H., Z.H., M.C.M., M.C.M.-K., C.M.P., M.V., M.G., T.S., S.W., L.Z., L.G., L.R., D.C., K.R., L.C., M.V., S.H.A., S.E., M.P., J.F.F., S.E.M., M.V., J.W.H., W.K., G.H., A.Z., F.S., J.L., A.H., T.T.H., S.J.L., J.L.W.; Writing—Original Draft: S.L., J.L.W.; Writing—Review and Editing: S.L., N.S., A.G., S.A., B.R., I.M., M.J.S., S.A.-L., N.K., F.I.R., S.R., L.S., J.T., G.M., G.C.S., C.M., L.M., D.J.S., H.J.Z., R.A., T.N., C.W., E. Kajantie, E. Keikkala, S.M., J.R., S.S., W.S., M.V., V.W.V.J., R.M.B., A.M.P., M.C.-T., T.D., S.E.H., Z.H., M.C.M., M.C.M.-K., C.M.P., M.V., M.G., T.S., S.W., L.Z., L.G., L.R., D.C., K.R., L.C., M.V., S.H.A., S.E., M.P., J.F.F., S.E.M., M.V., J.W.H., W.K., G.H., A.Z., F.S., J.L., A.H., T.T.H., S.J.L., J.L.W.; Visualization: S.L.; Project administration: J.W., S.L.

Competing interests

The authors declare no competing interests. Where authors are identified as personnel of the International Agency for Research on Cancer/ World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/ World Health Organization.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-023-05698-x>.

Correspondence and requests for materials should be addressed to Joseph L. Wiemels.

Peer review information *Communications Biology* thanks Paola Galdi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Chia Chuen Khor and George Inglis. A peer review file is available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

¹Center for Genetic Epidemiology, Department of Population and Public Health Sciences, University of Southern California, Los Angeles, California, USA. ²Epigenomics and Mechanisms Branch, International Agency for Research on Cancer, Lyon, France. ³Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA. ⁴Centre for Environmental Sciences, UHasselt, Agoralaan, Building D, 3590 Diepenbeek, Belgium. ⁵The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. ⁶Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. ⁷Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine, London, UK. ⁸ISGlobal, Institute for Global Health, Barcelona, Spain. ⁹Universitat Pompeu Fabra (UPF), Barcelona, Spain. ¹⁰CIBER Epidemiología y Salud Pública, Madrid, Spain. ¹¹Human Development and Health, Faculty of Medicine, Southampton General Hospital, University of Southampton, Southampton, UK. ¹²Department of Computer Science, Aberystwyth University, Aberystwyth, Ceredigion SY23 3DB, UK. ¹³Department of Environmental Immunology, Helmholtz Centre for Environmental Research -UFZ, Leipzig, Germany. ¹⁴Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. ¹⁵Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland. ¹⁶MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. ¹⁷School of Psychology, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK. ¹⁸School of Public Health, University of California Berkeley, Berkeley, California, USA. ¹⁹SAMRC Unit on Risk & Resilience in Mental Disorders, Dept of Psychiatry & Neuroscience Institute, University of Cape Town, Rondebosch, South Africa. ²⁰Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital, University of Cape Town, Rondebosch, South Africa. ²¹Clinical Medicine Research Unit, Medical Research Center Oulu, Oulu University, Hospital and University of Oulu, Oulu, Finland. ²²Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway. ²³Pediatric Research Centre, Children's Hospital, Helsinki University Hospital and University of Helsinki, Helsinki, Finland. ²⁴Population Health Unit, Department of Public Health and Welfare, Finnish Institute for Health and Welfare, Oulu, Finland. ²⁵Research Unit of Population Health, Faculty of Medicine, University of Oulu, Oulu, Finland. ²⁶Department of Anthropology and Institute of Behavioral Science, University of Colorado Boulder, Boulder, CO, USA. ²⁷MRC Unit The Gambia at the London School of Hygiene & Tropical Medicine, Fajara, The Gambia. ²⁸Nuffield Department of Women's & Reproductive Health, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK. ²⁹Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway. ³⁰Department of Pediatric Oncology and Hematology, Oslo University Hospital, Norwegian Institute of Public Health, Oslo, Norway. ³¹Department of Physical Health and Aging, Division for Mental and Physical Health, Norwegian Institute of Public Health, Oslo, Norway. ³²Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany. ³³Woman and Child Health Department, Ospedale Versilia, AUSL Toscana Nord Ovest, Pisa, Italy. ³⁴Department of Medical Sciences, University of Turin, CPO Piemonte, Turin, Italy. ³⁵Department of Population and Public Health Sciences, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA. ³⁶Department of Social Medicine, Faculty of Medicine, University of Crete, Heraklion, Greece. ³⁷Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK. ³⁸David Hide Asthma and Allergy Research Centre, Isle of Wight, UK. ³⁹College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA. ⁴⁰Department of Women & Children's Health, King's College London, London, UK. ⁴¹Division of Epidemiology, Biostatistics, and Environmental Health, University of Memphis, Memphis, TN, USA. ⁴²Perinatal Immunology, Medical Faculty, Saxonian Incubator for Clinical Translation (SIKT), University of Leipzig, Leipzig, Germany. ⁴³Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA. ⁴⁴Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC, USA. ✉email: wiemels@usc.edu