Epigenome-wide meta-analysis of blood DNA methylation in newborns and children identifies numerous loci related to gestational age.

Simon Kebede Merid et al.

Genome Med. 2020 Mar 2;12(1):25.

The official published article is available online at:
DOI: https://doi.org/10.1186/s13073-020-0716-9

Epigenome-wide meta-analysis of blood DNA methylation in newborns and children identifies numerous loci related to gestational age

Simon Kebede Merid[1]ł, Alexei Novoloaca[2]ł, Gemma C. Sharp[3,4]ł, Leanne K. Küpers[5-7]ł, Alvin T.Kho[8]ł, Ritu Roy[9-10], Lu Gao[11], Isabella Annesi-Maesano[12], Pooja Jain[13, 14], Michelle Plusquin[13, 15], Manolis Kogevinas[16-19], Catherine Allard[20], Florianne O. Vehmeijer[21-23], Nabila Kazmi[3, 6], Lucas A. Salas[24], Faisal I Rezwan[25], Hongmei Zhang[26], Sylvain Sebert[27-29], Darina Czamara[30], Sheryl L. Rifas-Shiman[31], Phillip E. Melton[32, 33], Debbie A Lawlor[3, 6, 34], Göran Pershagen[1, 35], Carrie V. Breton[11], Karen Huen[36], Nour Baiz[12], Luigi Gagliardi[37], Tim S Nawrot[13, 38], Eva Corpeleijn[7], Patrice Perron[20,39], Liesbeth Duijts[21-22], Ellen Aagaard Nohr[40], Mariona Bustamante[16-18], Susan L Ewart[41], Wilfried Karmaus[26], Shanshan Zhao[42], Christian M Page[43], Zdenko Herceg[2], Marjo-Riitta Jarvelin[27,28,44,45], Jari Lahti[46, 47], Andrea A. Baccarelli[48], Denise Anderson[49], Priyadarshini Kachroo[50], Caroline L Relton[3, 6, 34], Anna Bergström[1, 35], Brenda Eskenazi[51], Munawar Hussain Soomro[12], Paolo Vineis[52], Harold Snieder[7], Luigi Bouchard[20, 53, 54], Vincent W. Jaddoe[21-23], Thorkild I.A. Sørensen[3, 55, 56], Martine Vrijheid[16-18], S Hasan Arshad[57, 58], John W Holloway[25], Siri E Håberg[43], Per Magnus[43], Terence Dwyer[60], Elisabeth B. Binder[30, 61], Dawn L. DeMeo[50], Judith M Vonk[62, 63], John Newnham[64], Kelan G.Tantisira[50], Inger Kull[65], Joseph L Wiemels[66], Barbara Heude[67], Jordi Sunyer[16-19], Wenche Nystad[43], Monica C Munthe-Kaas[43, 68], Katri Räikkönen[43], Emily Oken[31], Rae-Chi Huang[49], Scott T Weiss[50], Josep Maria Antó[16-19], Jean Bousquet[69, 70], Ashish Kumar[1, 71, 72], Cilla Söderhäll[73], Catarina Almqvist[74, 75], Andres Cardenas[76], Olena Gruzieva[1, 35], Cheng-Jian Xu[77], Sarah E Reese[78], Juha Kere[79, 80], Petter Brodin[73, 81, 82], Olivia Solomon[36], Matthias Wielscher[44], Nina Holland[36], Akram Ghantous[2], Marie-France Hivert[20, 31, 83], Janine F. Felix[21-23], Gerard H. Koppelman[77], Stephanie J. London[78]‡, Erik Melén[1, 65, 84]‡*

ł Shared first authorship

‡ Shared last authorship

* Corresponding author: Erik Melén, MD, PhD
Department of Clinical Sciences and Education, Södersjukhuset
Karolinska Institutet
118 83 Stockholm, Sweden
Email: erik.melen@ki.se
Phone: +46706689194

1. Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
2. Epigenetics Group, International Agency for Research on Cancer, Lyon, France.
3. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
4. Bristol Dental School, University of Bristol, Bristol.
5. Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, the Netherlands.
6. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.
7. University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands.
8. Boston Children's Hospital and Harvard Medical School, Boston, MA, USA.
9. Computational Biology And Informatics, University of California, San Francisco, San Francisco, CA, USA.
10. HDF Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA.

11. Department of Preventive Medicine, University of Southern California, Los Angeles, USA.
12. Sorbonne Université and INSERM, Epidemiology of Allergic and Respiratory Diseases Department (EPAR), Pierre Louis Institute of Epidemiology and Public Health (IPLESP UMRS 1136), Saint-Antoine Medical School, Paris, France.
13. NIHR-Health Protection Research Unit, Respiratory Infections and Immunity, Imperial College London, London, UK.
14. Department of Epidemiology and Biostatistics, The School of Public Health, Imperial College London, London, United Kingdom.
15. Centre for environmental sciences, Hasselt University, Hasselt, Belgium.
16. ISGlobal, Barcelona Institute for Global Health, Barcelona, Spain.
17. Universitat Pompeu Fabra (UPF), Barcelona, Spain.
18. CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain.
19. IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain.
20. Centre de recherche du Centre hospitalier universitaire de Sherbrooke (CHUS), Sherbrooke, Qc, Canada.
21. The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
22. Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
23. Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
24. Department of Epidemiology, Geisel School of Medicine at Dartmouth College.
25. School of Water, Energy and Environment, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK
26. Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, USA.
27. Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland
28. Biocenter Oulu, University of Oulu, Finland
29. Department of Genomic of Complex diseases, School of Public Health, Imperial College London
30. Max-Planck-Institute of Psychiatry, Department of Translational Research in Psychiatry, Munich, Germany.
31. Division of Chronic Disease Research Across the Lifecourse (CoRAL), Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA.
32. School of Pharmacy and Biomedical Sciences, Faculty of Health Sciences, Curtin University, Bentley Australia.
33. Curtin/UWA Centre for Genetic Origins of Health and Disease, School of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Western Australia.
34. Bristol NIHR Biomedical Research Centre, Bristol, UK
35. Centre for Occupational and Environmental Medicine, Stockholm Region, Stockholm, Sweden.
36. Children's Environmental Health Laboratory, University of California, Berkeley, Berkeley, California.
37. Division of Neonatology and Pediatrics, Ospedale Versilia, Viareggio, AUSL Toscana Nord Ovest, Pisa, Italy.
38. Department of Public Health & Primary Care, Leuven University, Leuven, Belgium.
39. Department of Medicine, Université de Sherbrooke.
40. Research Unit for Gynaecology and Obstetrics, Department of Clinical Research, University of Southern Denmark, Odense, DK.
41. College of Veterinary Medicine, Michigan State University, East Lansing, USA.
42. National Institute of Environmental Health Sciences, National Institutes of Health, Dept. of Health and Human Services, RTP, NC, USA.
43. Norwegian Institute of Public Health, Oslo, Norway.
44. Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, UK.
45. Unit of Primary Care, Oulu University Hospital, Oulu, Finland
46. Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Finland.
47. Turku Institute for Advanced Studies, University of Turku, Turku, Finland.
48. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University Medical Center, New York, NY, USA.
49. Telethon Kids Institute, University of Western Australia, Australia.
50. Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
51. Center for Environmental Research and Children's Health (CERCH), University of California, Berkeley, Berkeley, California.

52. MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK.
53. Department of biochemistry, Université de Sherbrooke, Sherbrooke, Qc, Canada
54. Department of medical biology, CIUSSS-SLSJ, Saguenay, Qc, Canada
55. Novo Nordisk Foundation Center for Basic Metabolic Research, Section on Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
56. Department of Public Health, Section of Epidemiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
57. Clinical & Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK.
58. The David Hide Asthma and Allergy Research Centre, Newport, Isle of Wight, UK.
59. NIHR Biomedical Research Centre, University Hospitals Southampton NHS Foundation Trust, Southampton, UK.
60. The George Institute for Global Health, Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK.
61. Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, USA.
62. University of Groningen, University Medical Center Groningen, Dept. of Epidemiology, Groningen, The Netherlands.
63. University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, The Netherlands.
64. Faculty of Health and Medical Sciences, UWA Medical School, University of Western Australia.
65. Department of Clinical Sciences and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden; Sachs' Children's Hospital, Södersjukhuset, Stockholm, Sweden.
66. Center for Genetic Epidemiology, University of Southern California.
67. INSERM, UMR1153 Epidemiology and Biostatistics Sorbonne Paris Cité Center (CRESS), Research Team on Early life Origins of Health (EarOH), Paris Descartes University, Paris, France.
68. Department of Pediatric Oncology and Hematology, Oslo University Hospital, Norway.
69. University Hospital, Montpellier, France.
70. Department of Dermatology, Charité, Berlin, Germany.
71. University of Basel, Basel, Switzerland.
72. Swiss Tropical and Public Health Institute, Basel, Switzerland.
73. Department of Women´s and Children´s Health, Karolinska Institutet, Stockholm, Sweden.
74. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
75. Pediatric Allergy and Pulmonology Unit at Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden.
76. Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, CA, USA.
77. University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, GRIAC Research Institute Groningen, The Netherlands.
78. National Institute of Environmental Health Sciences, National Institutes of Health, Dept. of Health and Human Services, RTP, NC, USA.
79. Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden.
80. Folkhälsa Research Institute, Helsinki, and Stem Cells and Metabolism Research Program, University of Helsinki Finland.
81. Department of Newborn Medicine, Karolinska University Hospital, Stockholm, Sweden.
82. Science for Life Laboratory, Stockholm, Sweden.
83. Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA.
84. Sachs' Children's Hospital, South General Hospital, Stockholm, Sweden.

Abstract

Background: Preterm birth and shorter duration of pregnancy are associated with increased morbidity in neonatal and later life. As the epigenome is known to have an important role during fetal development, we investigated associations between gestational age and blood DNA methylation in children.

Methods: We performed meta-analysis of Illumina's HumanMethylation450-array associations between gestational age and cord blood DNA methylation in 3,648 newborns from 17 cohorts without common pregnancy complications, induced delivery or Caesarean section. We also explored associations of gestational age with DNA methylation measured at 4-18 years in additional pediatric cohorts. Follow-up analyses of DNA methylation and gene expression correlations were performed in cord blood. DNA methylation profiles were also explored in tissues relevant for gestational age health effects: fetal brain and lung.

Results: We find that DNA methylation in cord blood is associated with gestational age (range 27-42 weeks) at 8,899 CpGs including 3,343 novel CpG ($P_{Bonferroni} < 1.06\times10-7$). After restricting findings to at least three significant adjacent CpGs, we identified 1,276 CpGs annotated to 325 genes. Results were generally consistent when analyses were restricted to term births. Cord blood findings tended not to persist into childhood and adolescence. Pathway analyses identified enrichment for biological processes critical to embryonic development. Follow-up of identified genes showed correlations between gestational age

and DNA methylation levels in fetal brain and lung tissue, as well as correlation with expression levels.

Conclusions: We identified numerous CpGs differentially methylated in relation to gestational age at birth that appear to reflect fetal developmental processes across tissues. These findings may contribute to understanding mechanisms linking gestational age to health effects.

Keywords

Development, epigenetics, gestational age, preterm birth, transcriptomics.

Background

Preterm birth (birth before 37 weeks' gestation) is associated with increased neonatal morbidity and mortality[1,2], as well as later health[3-6]. In children born at very young gestational ages, bronchopulmonary dysplasia, retinopathy and neurodevelopmental impairment are major health challenges[7-12]. Lower lung function is observed in children born moderately preterm, i.e. between 32 and 36 completed weeks, compared to those born at term[13]. Even variation in gestational age within the normal range (37-41 weeks) is related to various health outcomes, including neurological and cognitive development [14-17] and respiratory disease[4]. Mechanisms for many of these findings are not well understood.

The epigenome is known to have an important role during fetal development. The best studied epigenetic modification is methylation. DNA methylation patterns have been associated with environmental factors relevant to preterm birth, including smoking, air pollution exposure, microbial and maternal nutritional factors[18-22]. Such exposure-related epigenetic patterns potentially influence gene expression profiles and/or susceptibility to chronic disease during the lifecourse[23,24]. Further, DNA methylation in whole blood at birth may also reflect development across fetal life. It is possible that DNA methylation changes at birth may contribute to the myriad immediate and late health outcomes that have been associated with gestational age.

Knowledge about DNA methylation and gene expression profiles associated with length of gestation may help to better understand both the molecular basis of abnormal processes related to prematurity as well as normal human development. Several studies have

reported associations of gestational age among both term and preterm births with cord blood DNA methylation[25-29]. In the largest EWAS to date (n = 1,753 newborns), 5,474 CpGs in cord blood were associated with gestational age[30]. While these individual studies have identified widespread associations of DNA methylation patterns at birth with gestational age, meta-analysis of results from multiple individual cohorts increases sample size and, thus, greatly increases power to detect robust differential methylation signals.

We examined DNA methylation levels in newborns in relation to gestational age in a large-scale meta-analysis and also examined functional effects on expression of nearby genes of potential relevance for later health. We meta-analysed harmonised cohort specific EWAS results of the association of gestational age with cord blood DNA methylation levels from the Pregnancy And Childhood Epigenetics (PACE) Consortium of pregnancy and childhood cohorts[31]. We also examined associations with continuous gestational age limited to term newborns. CpGs that were differentially methylated in cord blood in relation to gestational age were then analyzed in two fetal tissues (lung and brain), with relevance for health impacts of low gestational age[7-12]. We conducted analyses to explore whether associations of CpG methylation with gestational age persisted in older children aged 4-18 years. DNA methylation status at the identified CpGs was analysed for association with gene expression patterns of nearby genes in cord blood during different developmental stages. Finally, we performed pathway and functional network analysis of identified genes to gain insight into the biological implications of our findings.

Methods

Figure 1 gives an outline of the design of this study.

Study population

A total of 11,000 participants in 26 independent cohorts were included in our study. In the "all births model" meta-analysis we included n=6,885 newborns from 20 cohorts. In our main "no complications model" we excluded participants with maternal complications (maternal pre-eclampsia or diabetes or hypertension) and Caesarean section delivery or delivery start with induction, leaving 3,648 newborns from 17 cohorts for this analysis (Additional file 1: Table S1). For the additional look-up of persistent differential methylation at later ages, we used participants from 4 cohorts with whole blood DNA methylation in early childhood (4-5y; n=453), 5 cohorts with whole blood DNA methylation at school age (7-9y; n=899) and 5 cohorts with whole blood DNA methylation in adolescence (16-18y; n=1129). Detailed methods for each cohort are provided in Additional file 2: Supplementary information. All cohorts acquired ethics approval and informed consent from participants prior to data collection through local ethics committees (Additional file 2: Supplementary information).

Gestational age

In each cohort, information on gestational age at birth was obtained from birth certificates (n=725), medical records using ultrasound estimation (n=1,931), or last menstrual period date (n=468), or combined estimate from ultrasound and last menstrual period date

(n=6,630), or otherwise from self-administrated questionnaires (n=1,246). Gestational age was analysed in days. Women with a gestational age of more than 42 weeks (294 days) were excluded from all models. Additionally, multiple births were also excluded from the analysis.

Methylation measurements and quality control

DNA methylation from newborns and older children was measured using the Illumina450K platform. Each cohort conducted their own quality control and normalisation of DNA methylation data, as detailed in Additional file 1: Table S2. Cohorts corrected for batch effects in their data using surrogate variables, ComBat[32] or by including a batch covariate in their models. To reduce the impact of severe outliers in the DNA methylation data on the meta-analysis, cohorts trimmed the methylation beta values by removing, for each CpG, observations more than three times the interquartile range below the 25th percentile or above the 75th percentile[33]. Cohorts retained all CpGs that passed quality control and removed CpGs that were mapped to the X (n = 11,232) or Y (n = 416) chromosomes and control probes (n = 65), leaving a maximum total of 473,864 CpGs included in the meta-analysis.

Cohort-specific statistical analyses

Each cohort performed independent EWAS according to a common, pre-specified analysis plan. Robust linear regression (rlm in the MASS R package[34]) was used to model gestational age as the exposure and DNA methylation beta values as the outcome. In the primary analysis, gestational age was used as a continuous variable excluding cohorts that had term-

264 only infants. In secondary models, we modelled term-only children defined as a gestational
265 age ≥ 37 weeks (≥ 259 days), but less or equal with 42 weeks. All models were adjusted for
266 sex, maternal age (years), maternal social class (variable defined by each individual cohort;
267 Additional file 1: Table S2), maternal smoking status (the preferred categorization was into
268 three groups: no smoking in pregnancy, stopped smoking in early pregnancy, smoking
269 throughout pregnancy, but a binary categorization of any versus no smoking was also
270 acceptable), parity (the preferred categorization was into two groups: no previous children,
271 one or more previous children), birth weight in grams, age of the child (years) included for
272 older children, batch or surrogate variables. Optionally, cohorts could include ancestry,
273 and/or selection covariates, if relevant to their study. We also adjusted for potential
274 confounding by cell type using estimated cell type proportions calculated from a cord blood
275 cell type reference panel[35] for newborn cohorts or the adult blood cell type reference
276 panel[36] for cohorts with older children using the *estimateCellCounts* function in the *minfi* R
277 package[37].
278
279 ## Meta-analysis
280 We performed fixed-effects meta-analysis weighted by the inverse of the variance with
281 METAL[38]. A shadow meta-analysis was also conducted independently by a second study
282 group (see author contribution) and the results were compared[39] (and confirmed). All
283 downstream analyses were conducted using R version 2.5.1 or later[40]. Multiple testing was
284 accounted for by applying Bonferroni correction level for 473,864 tests ($P < 1.06\times10^{-7}$). A
285 random effects models were performed using the METASOFT tool[41]. We explored

heterogeneity between studies using the I^2 statistic[42]. *A priori*, we defined I^2>50% as reflecting a high level of between-study variation. In case of I^2>50%, we replaced values with random effects estimates as these are attenuated in the face of heterogeneity and thus more conservative. To focus functional analyses and bioinformatics efforts on genes and loci that were found to be robustly associated with gestational age, we selected regions that had at least three adjacent Bonferroni significant CpGs ($P < 1.06\times10^{-7}$)[43].

Analyses of differentially methylated regions

Differentially methylated regions (DMRs) were identified using two methods available for meta-analysis results comb-p[44] and DMRcate[45]. Input parameters used for the DMR calling in both algorithms are provided in the supplementary text. Comb-p uses a 1-step Šidák correction[44] and DMRcate uses an FDR correction[45] per default. The selected regions were defined based on the following criteria: the minimum number of CpGs in a region had to be 2, regional information can be combined from probes within 1,000 bp and the multiple-testing corrected $P < 0.01$ (Šidák-corrected $P < 0$ .01 from comb-p and FDR < 0.01 from DMRcate).

Analyses of embryonic DNA methylation

DNA methylation from lung tissue of 74 fetuses (estimated ages 59 to 122 days post conception[46]) were used for analyses of differentially methylated CpGs (three or more adjacent Bonferroni significant CpGs, $P < 1.06\times10^{-7}$; n=1,276) from the newborn meta-analysis. A linear regression model adjusted for sex and *in utero* smoke exposure (IUS) was

applied. A Bonferroni look-up level correction (0.05/1,030; $P < 4.85\times10^{-5}$) considered as significance threshold, followed by a comparison of the direction of effect with that in the cord blood meta-analysis. We also performed look-up analyses of selected 1,276 CpGs in another organ, fetal brain tissue, from 179 fetuses collected between 23 and 184 days post-conception[47]. For these analyses, we kept the available Bonferroni correction $P < 1.06\times10^{-7}$ as significance threshold, followed by a comparison of the direction of effect with that in the cord blood meta-analysis.

Look-up analyses in older ages

Differentially methylated CpGs (three or more adjacent CpGs below the Bonferroni correction $P < 1.06\times10^{-7}$; n=1,276) from the newborn meta-analyses were analysed with a look-up approach using data from four early childhood, five school age, and five adolescence cohorts. Cohorts included the same covariates in these analyses as in the cord blood analyses and child age. We performed fixed effects inverse variance weighted meta-analyses using METAL[38] for these three age groups. For this hypothesis-driven analysis, CpG methylation association with gestational age was considered statistically significant at nominal $P < 0.05$, followed by a comparison of the direction of effect with that in the cord blood meta-analysis.

Longitudinal analysis

Longitudinal DNA methylation data from birth to early childhood and from birth to adolescence were analysed for the three or more adjacent Bonferroni significant 1,276 CpGs

found to be associated with gestational age. DNA methylation from two time points (birth and 4 years) in INMA and three time points (birth, 7 and 17 years) in ALSPAC were analysed separately. To estimate changes in DNA methylation, we applied linear mixed models with repeated measurement taking into account the within-person time effect. The models were adjusted for covariates and estimated cell count similar to cross-sectional analysis. Interaction terms between age and gestational age were included in the model to capture differences in methylation change between (birth and 4 years), (birth and 7 years) and (7 and 17 years) per day increase in gestational age at delivery, respectively. The stable CpGs that did not change significantly from birth to adolescence had no association with age (at nominal $P < 0.05$), and no interaction between gestational age and childhood age (at nominal $P < 0.05$).

Enrichment and functional analysis

CpGs were annotated using *FDb.InfiniumMethylation.hg19* R package, with enhanced annotation for nearest genes within 10Mb of each site, as previously described[20]. Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using the overrepresentation analysis (ORA) tool ConsensusPathDB (http://consensuspathdb.org/ [48,49]). P-values for enrichment were adjusted for multiple testing using the FDR method.

DNA methylation in relation to gene expression

Correlations between DNA methylation and gene expression levels were tested using paired

DNA methylation and gene expression data in publicly available datasets. We tested transcript levels of genes within a 500 kb region of the 1,276 three adjacent CpGs (250 kb upstream and 250 kb downstream). The mRNA gene expression (Affymetrix Human Transcriptome Array 2.0) and methylation (Illumina Infinium® HumanMethylation450 BeadChip assay) were measured in cord-blood samples from 38 newborns[50-52]. First, we created residuals for mRNA expression and residuals for DNA methylation and used linear regression models to evaluate correlations between expression residuals and DNA methylation residuals. These residual models were adjusted for covariates, estimated white blood cell proportions, and technical variation. We corrected these analyses for multiple testing using Bonferroni correction.

Results

Study characteristics

We meta-analyzed Illumina's HumanMethylation450-array results from 17 independent cohorts with data on newborn DNA methylation status, and 10 cohorts with data on DNA methylation in older children (age 4 to 18 years), including 4 cohorts with DNA methylation data both at birth and at an older age (Figure 1). Table 1 summarizes the characteristics of participating cohorts. A summary of methods used by each cohort is provided in Additional file 1: Table S1 and Table S2. In our main "no complications" model, we excluded participants exposed to maternal pregnancy complications (maternal diabetes, hypertension or pre-eclampsia) and whose labour was induced or who were delivered by Caesarean section. With continuous gestational age in number of days as the exposure (gestational age

range 186-294 days corresponding to 27-42 weeks), we analysed results from 3,648 newborns and from 2,481 older children. This model was selected as the main model because associations of DNA methylation with gestational age related to pregnancy complications or potentially influenced by obstetric interventions, may be less reflective of normal developmental processes than newborns with spontaneous uncomplicated delivery. However, we also analysed a larger dataset of 6,885 newborns from 20 independent cohorts, including pregnancies with pregnancy complications and obstetric interventions, referred to as the "all births model" (see below).

Associations between gestational age and newborn DNA methylation

Our main "no complications" model identified DNA methylation at 8,899 CpGs in cord blood annotated to 4,966 genes associated with gestational age after Bonferroni correction for 473,864 tests ($P < 1.06\times10^{-7}$). CpGs associated with gestational age had a modest predominance of negative (60%) versus positive (40%) direction of effect, with an overall absolute median difference in mean methylation of 0.36% per gestational week, IQR = [0.26%-0.49%] (Figure 2A). In general, results were highly homogeneous; evidence of high between-study heterogeneity, using a criterion of $I^2 > 50\%$, was seeing for only 319 of the 8,899 CpGs (Additional file 1: Table S3). Leave one out analyses did not indicate an influential effect on meta-analysis results of any single study. However, we replaced fixed effects values with random effects estimates for those CpGs with between study $I^2 > 50\%$, as these are more conservative in the case of heterogeneity.

Differentially methylated CpGs spanned all chromosomes (Figure 2B). The CpG with the lowest p-value ($P = 2.7\times10^{-129}$ for cg16103712; Table 2) was annotated to *MATN2* on chr 8, and the difference in mean methylation at this CpG was 2.13% lower per additional gestational week (equal to 0.30% per day). The CpG with the largest negative association was cg04347477, annotated to *NCOR2* on chr 12 (Table 3), with a lower mean methylation of 2.53% per additional gestational week. *B3GALT4* (chr 6) had the largest number of significant CpGs negatively associated with gestational age (21 out of 52 (40%) tested CpGs annotated to *B3GALT4*). The largest positive association was observed for cg13036381 annotated to *LOC401097* (chr 3) (Table 3) with a difference in mean methylation of 1.95% per additional gestational week. *DDR1* (chr 6) had the largest number of significant CpGs positively associated with gestational age (26/95 (27%) CpGs). A complete list of associated CpGs is presented in Additional file 1: Table S3 and the CpG variation across cohorts in Additional file 3: Figure S1 (top CpGs).

We performed a sensitivity analysis by excluding cohorts that were included in previous EWAS of gestational age[29,30] (three cohorts: MoBa1, MoBa2 and ALSPAC) in order to evaluate associations not driven by previous results, and found a high correlation (r=0.89) of effect estimates (Additional file 3: Figure S2) compared with results from all cohorts included in the no complication model.

Next, we performed a meta-analysis of the larger dataset of 6,885 participants from 20 studies without excluding maternal complications and Caesarean section delivery or induced delivery. In this “all births model”, 17,095 CpGs located in or near 7,931 genes were

associated with gestational age after Bonferroni correction ($P < 1.06\times10^{-7}$). Not surprisingly given the higher levels of statistical significance in this much larger data set, we found somewhat more between-study heterogeneity than in the no complications model, but high levels ($I^2 > 50\%$) were observed for only 1,784 out of these 17,095 CpGs (Additional file 1: Table S4). We also observed a considerable overlap of CpGs between the two models with 93% of the 8,899 CpGs in the no complication model also reaching Bonferroni-significance in the all births model and showing the same direction of effect.

CpG localization and regulatory region analyses

The 8,899 differentially methylated CpGs in relation to continuous gestational age in the no complications model were enriched for localization to CpG island shores (33% of the 8,899 CpGs are in shores, whereas 23% of all CpGs on the 450K array are in shores, $P_{enrichment} = 4.1\times 10^{-100}$, Figure 3), open sea (45% versus 37% , $P_{enrichment} = 1.4 \times 10^{-63}$), enhancers (37% versus 22%, $P_{enrichment} = 1.05 \times 10^{-236}$), DNase hypersensitivity sites (18% versus 12%, $P_{enrichment} = 1.3\times 10^{-56}$) and CpG island shelves (12% versus 10%, $P_{enrichment} = 1.2 \times 10^{-11}$) (Figure 3). In contrast, we found relative depletion in CpG islands (10% versus 31%, $P_{enrichment} = 2.2 \times 10^{-308}$), FANTOM 4 promoters (2.3% versus 6.7%, $P_{enrichment} = 6.7 \times 10^{-79}$), and promoter-associated regions (11% versus 19%, $P_{enrichment} = 2.2 \times 10^{-104}$).

Analysis restricted to term-births

To evaluate whether observed DNA methylation differences in relation to continuous gestational age were driven by preterm birth, we repeated the no complications model

including only infants born at term (gestational age 37 to 42 weeks). In this analysis, we meta-analysed results from 18 cohorts (one additional cohort with term-birth data only was included; GEN3G) (n = 3,593). We identified 5,930 sites significantly associated with gestational age at Bonferroni correction ($P < 1.06\times10^{-7}$, median difference in mean methylation per additional gestational week = 0.43%, IQR = [0.32%-0.58%]). The vast majority (5,399; 91%) of these differentially methylated CpGs overlapped with those found in the main analyses (no complications model) without exclusion of those born preterm (Figure 4).

Selection of CpGs for downstream analyses

Given the large number of significant associations in our main model (8,899 CpGs), we focused subsequent analyses on loci including at least three adjacent CpGs that survived Bonferroni correction[43]. There were 1,276 differentially methylated CpGs in 325 unique genes that fulfilled this criterion (Additional file 1: Table S5). As in the overall data, we observed a slight predominance of negative (n=702; 55%) versus positive (n=574; 45%) directions of effect (Figure 2A). The lowest p-value, $P = 1.2\times10^{-93}$, was observed for cg04276536 (*CCDC102A,* chromosome 16). As for the full EWAS results, the largest negative and positive association effect sizes were observed for cg04347477 (*NCOR2*) and cg13036381 (*LOC401097*), respectively. These 1,276 CpGs had the same CpG localization enrichment pattern as the full set of Bonferroni-significant CpGs (n=8,899), except that there was a relative depletion in CpG island shelves (7.6% versus 10% overall, $P_{enrichment} = 2.3\times10^{-12}$) and open sea (32% versus 37%, $P_{enrichment} = 2.4\times10^{-12}$) (Figure 3).

Differentially methylated region (DMR) analyses

Using two different methods for DMR analysis of gestational age in relation to newborn DNA methylation, we identified 4,479 significant (Šidák-corrected P < 0 .01) DMRs from the comb-p method and 14,671 significant (FDR P < 0 .01) DMRs from DMRcate, respectively, including 2,375 DMRs (representing 11,861 CpGs) that were significant based on both approaches (Additional file 1: Table S6). Out of the 8,899 bonferroni significant single CpGs, 2,289 CpGs overlapped with CpGs in identified in the combined DMR analyses (11,861 CpGs). Moreover, from loci included by the three or more adjacent CpG selection (n=1,276), 521 CpGs overlapped with those identified in the combined DMR analyses. Of note, out of the 1,276 CpGs, 1,223 and 1,231 CpGs were captured by DMRs identified using the comb-p and DMRcate independent approaches, respectively.

Assessment of CpG methylation in earlier embryonic stages

We examined whether the CpGs detected in cord blood (that originate from embryonic germ layer mesoderm) were differentially methylated in relation to gestational age in other fetal tissues, lung and brain, that originate from the two other embryonic germ layers, ectoderm and endoderm, respectively, collected prenatally[46,47]. To this end, we performed look-up analyses in DNA methylation data for 74 fetal lung samples representing gestational age 59 to 122 days (~8 to 17 completed gestational weeks)[46]. Out of the 1,276 CpGs, selected based on three or more adjacent CpGs from our no complications model, 1,030 CpGs were available in the fetal lung dataset. We observed associations at Bonferroni look-

485 up level correction significance (0.05/1,030; $P < 4.85\times10^{-5}$) between DNA methylation levels
486 in fetal lung tissue and gestational age at tissue collection for 151 (15%) CpGs (Additional file
487 1: Table S7). Of these 151 (58 negatively and 93 positively associated), 78 showed the same
488 direction of association with gestational age in cord blood and fetal lung tissue. The look-up
489 analyses of fetal brain tissue were undertaken in 179 samples representing 23 to 184 days
490 (~3 to 26 completed weeks)[47]. Out of the 1,276 CpGs, we found significant associations
491 (using Bonferroni correction $P < 1.06\times10^{-7}$ cut-off since only this data was available for
492 analyses; Additional file 1: Table S8) for 268 CpGs (21%) in relation to gestational age at
493 tissue collection. Of these 268 sites, 227 had same direction of effect in the cord blood and
494 fetal brain data. We found enrichment more than expected by chance for our cord blood
495 gestational age associated CpGs (n=1,276) in fetal lung ($P = 2.1 \times 10^{-4}$) and brain ($P = 3.9 \times$
496 10^{-57}) tissue. Thirty CpGs showed significant associations with gestational age in all three
497 tissues (cord blood, fetal lung and fetal brain).
498
499 ## Assessment of CpG methylation in older children
500 We examined whether the differentially methylated CpGs detected in cord blood samples
501 were associated with gestational age at birth in whole blood from older children. We
502 conducted three separate meta-analyses (no complications model) reflecting different age
503 periods in a total of 2,481 children: (i) Early childhood (4-5y; n = 453 from 4 cohorts); (ii)
504 school age (7-9 y; n = 899 from 5 cohorts) and (iii) adolescence (16-18 y; n = 1,129 from 5
505 cohorts), Additional file 1: Table S1. Of the 1,276 three or more adjacent genome wide
506 significant CpGs from our analyses in cord blood, 1,258 CpGs were available for analyses in
507 all older age groups. Out of these CpGs, we observed 40 sites in early childhood, 60 sites in

school age, and 60 sites in adolescence to be associated with gestational age at the nominal significance level, $P < 0.05$ with same direction of effect (Additional file 1: Table S9). However, no CpG survived Bonferroni look-up level correction (0.05/1,258; $P < 3.97\times10^{-5}$). One CpG (cg26385222 annotated to *TMEM176B*) previously associated with gestational age at birth[27], was nominally significant in all age groups with same direction of effect.

Longitudinal analysis

The results of the longitudinal analyses of blood DNA methylation in the INMA Study (n = 177 with paired samples from birth and 4 years) and the ALSPAC Study (n = 281 with samples collected at birth, 7 and 17 years) are provided in Additional file 1: Table S10. The vast majority of gestational age associated CpGs (n=1,054/1,276; 83%) underwent changes in methylation levels with age. Both increasing and decreasing patterns of change during early childhood (4 y) were observed, followed by stabilization during school age (7 y). For example, for cg08943494 in *PRR5L* on chr 11, an initial level of 61.5% and 51.4% in cord blood DNA methylation in INMA and ALSPAC respectively, decreased by 8.2% per year on average during early childhood in INMA and by 3.3% per year on average up to school age in ALSPAC, but then negligible further changes were seen from 7 to 17 years (Figure 5A). In contrast, increasing levels were seen for cg18183624 (chr 17; *IGF2BP1),* from an initial 48.8% and 38.7% in cord blood DNA methylation in INMA and ALSPAC, respectively, with a 5.1% per year on average between birth to 4 years in INMA and 1.9% per year on average between birth to 7 years, but after that no changes from 7 to 17 years. (Figure 5B).

Of the 1,054 CpGs displaying changes in DNA methylation levels with age, there were 589 CpGs where gestational age was associated with changes in DNA methylation levels (i.e. where an interaction between gestational age and age was found) from birth to 4 years (INMA) and 460 CpGs with changes from birth to 7 years (ALSPAC). However, only 30 of the 1,054 CpGs changed significantly in DNA methylation between 7 and 17 years (ALSPAC), suggesting that gestational age-related changes in DNA methylation levels had largely stabilized by age 7.

We identified 222 stable CpGs out of 1,276 (17%) that did not change appreciably from birth to adolescence. As an example, the stable DNA methylation at cg27058497 (*RUNX3*, chromosome 1) is shown in Figure 5C. A much lower proportion of the gestational age associated CpGs were stable from birth to adolescence compared to all CpGs on the array (17% versus 71%, $P_{enrichment} = 2.23\times 10^{-308}$).

Enrichment for biological processes and pathways

Using the complete list of 8,899 CpGs annotated to 4,966 genes, these were enriched for 1,784 GO terms including regulation of cellular and biological processes, system development, different signalling pathways and organ development (Additional file 1: Table S11). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses revealed 124 significant terms at FDR<0.05 representing a variety of human diseases, most notably various cancers, viral infections, metabolic processes and immune-related disorders (Additional file 1: Table S12). The 325 genes annotated to the 1,276 CpGs, selected by virtue

552 of three or more CpGs being localized to the same gene, were enriched for 198 Gene
553 Ontology (GO) terms very similar to those identified using Bonferroni significant CpGs
554 (Additional file 1: Table S13). When restricting analyses to the 222 longitudinally stable
555 CpGs, corresponding to 139 genes, 13 significant KEGG terms were revealed, primarily
556 representing infection- and immune-related disorders (Additional file 1: Table S14). For 186
557 genes annotated to the 1,054 CpGs changing with postnatal age, only one KEGG terms were
558 identified as statistically significant ($P = 1.2 \times 10^{-3}$ for the term MAPK signalling pathways;
559 Additional file 1: Table S14).
560
561 ## Correlation of DNA methylation and gene expression
562 For the 1,276 CpGs differentially methylated in relation to gestational age with at least 3
563 adjacent CpGs, we assessed correlations between DNA methylation and gene expression
564 (*cis*-eQTMs). From a publicly available dataset of expression and DNA methylation measured
565 in 38 cord blood samples[50-52], 1,174 out of the 1,276 CpGs were located within a 500kb (+/-
566 250 kb) window of a transcript cluster. Of these 1,174, 246 unique CpGs (367 total CpG-
567 transcript associations) correlated significantly with gene expression (Bonferroni $P< 0.05$,
568 Additional file 1: Table S15). Forty-six percent of these DNA methylation-expression
569 correlations were negative, with the lowest $P = 3.55\times10^{-6}$ coeff = -6.03 for cg01332054 and
570 *SEMA7A* expression and the largest negative effect estimate (-12.69) for cg26179948 and
571 *JAZF1* expression (Additional file 3: Figure S3 A, B). Fifty-four percent were positive, with
572 the lowest $P = 1.04\times10^{-5}$ coeff = 2.88 for cg20139800 and *MOG* expression and the largest
573 positive effect estimate (19.35) for cg03665259 and *CDSN* expression (Additional file 3:

574 Figure S3 C, D).

575

576 Discussion

577 In this large consortium-based meta-analysis, we identified 8,899 sites across the genome
578 where gestational age at birth was associated with cord blood DNA methylation. We also
579 identified numerous unique differentially methylated regions (DMRs) associated with
580 gestational age by applying two independent methods. The results were consistent when
581 restricted to births at term, demonstrating that the majority of our results were not driven
582 by preterm births. We confirmed many of the findings from previously published EWAS of
583 gestational age[23,26,27,29,30,53] and found a very high correlation between the significant CpG
584 point estimates in previously published datasets compared to our study (e.g. corr=0.92
585 between Hannon et al CpGs and our data; Additional file 1: Table S16), but importantly, we
586 also found 3,343 CpGs corresponding to 2,577 genes that had not been described
587 previously. There was a general lack of stability of the cord blood findings into childhood
588 and adolescence. However, there was a significant overlap of differentially methylated CpGs
589 in cord blood, fetal brain and lung tissues.

590

591 We found that various functional elements were enriched among gestational age-associated
592 CpGs. CpG island shores, enhancers, and DNase I hypersensitive sites were particularly
593 susceptible to DNA methylation changes in relation to gestational age, suggesting that these
594 differentially methylated sites are of functional importance[54].

We found clear overlap of differentially methylated CpGs in cord blood, fetal brain and fetal lung tissues in relation to gestational age. Thus, our cord blood findings seem to partly capture the epigenomic plasticity of prenatal development across tissues. The gene with the largest negative magnitude of association with cord blood DNA methylation in relation to gestational age, *NCOR2,* was also differentially methylated in brain and lung fetal tissues. *NCOR2* is involved in vitamin A metabolism and has previously been associated in GWAS with lung function[55]. Vitamin A supplementation is suggested to reduce the risk of bronchopulmonary dysplasia in extremely preterm born children[56]. Differential methylation of *NCOR2* in neurons associated with aging has been reported[57]. The gene with the second largest magnitude of negative association with methylation at birth, *PRR5L,* has been linked in GWAS to allergic diseases, found downregulated (expression) in osteoarthritis, and differentially methylated in type II diabetes[58-60]. The gene with the lowest p-value in our EWAS, *MATN2* plays a critical role in the differentiation and maintenance of skeletal muscles, peripheral nerves, liver and skin during development and regeneration[61] and is suggested as a potential biomarker in the early stage of osteoarthritis[62].

Differentially methylated CpGs associated with gestational age in cord blood were also present in our childhood and adolescence analyses. The only CpG (cg26385222, *TMEM176B*) that was associated with gestational age at all three time points (birth, childhood and adolescence) has been associated with gestational age in cord blood in previous studies[27]. The protein encoded by *TMEM176B* has also been suggested as a potential biomarker for

various cancers[63]. The low number of significant associations with gestational age at older ages with no CpG surviving multiple test correction may be partially explained by smaller sample sizes in childhood and adolescence than at birth and by the fact that many later exposures may obscure the association. However, in agreement with the cross-sectional analyses, our longitudinal analyses showed that DNA methylation at gestational age-associated CpGs typically undergoes dynamic changes during early childhood to a much higher degree than overall for CpGs on the 450K array. For the majority of these dynamics CpGs, change was most prominent during the first years of life, with many sites tending stabilize in methylation levels by school age. We also identified a subset of the CpGs differential methylated at birth (17%) which seem stable over time. For these CpGs, the early alteration of methylation levels by length of gestation was found stable postnatally across childhood and into adolescence.

In recent analyses by Xu *et al*, 14,150 CpGs related to childhood age were identified[64] and we found 280 overlapping with these CpGs among our 1,276 CpG list. Moreover, a study by Acevedo *et al*, showed 794 age-modified CpGs within 3 to 60 month after birth and 57 CpGs were overlapping with our 1,276 CpG list[65]. Thus, a proportion of gestational age-related CpGs are also associated with postnatal ageing. But similar to results from Simpkin *et al*.[66], we observed very little overlap (only 3 CpGs) with the CpGs used to derive epigenetic age by the Hannum and Horvath approach[67,68] or the epigenetic clock for gestational age at birth (10 CpGs overlapping)[28]. It should be noted that these studies primarily used the Illumina 27K array for analyses, which makes comparison difficult.

In the functional analyses we observed significant enrichment for several GO terms related to embryonic development, regulation of process and immune system development. The pathway analyses identified a subset of these genes linked to diseases also associated with low gestational age, for example asthma[69], inflammatory bowel disease[70], Type I/II diabetes[71] and cancer (leukemia)[72]. Importantly, genes annotated to CpGs found stable across childhood also showed enrichment for infection- and immune-related conditions. Whether cord blood DNA methylation at these CpGs affects later disease risk remains to be studied. Interestingly, differentially methylated loci in relation to asthma development have been recently identified in newborns[73]. The stable CpG cg27058497 (*RUNX3*), has been associated with *in utero* tobacco smoking exposure[74], childhood asthma[75], esophagus squamous cell carcinoma[76] and chronic fatigue syndrome[77]. Despite adjustment for maternal smoking in our gestational age EWAS model, we observed overlap between all FDR hits from our gestational age EWAS with those FDR hits presented in the maternal smoking related DNA methylation[20] with an overlap of 2,302/47,324 CpGs (4.9%, $P_{enrichment} < 2.2x10^{-308}$). This overlap likely reflects some pregnant women under reporting their smoking behaviour and the fact that smoking-related CpGs capture quantitative smoking history better than self-report[78,79]. However, we cannot rule out the possibility that some overlapping CpGs could be involved in biologic pathways linking smoking to the well-established consequence of shorter gestational length[80]. Other potential confounders not accounted for in this study such as maternal obesity and alcohol intake may influence offspring DNA methylation although we have found in the PACE consortium that their

impact on methylation[81,82] is very modest compared with maternal smoking in pregnancy which was included in our models.

This paper aimed at identifying CpGs associated with gestational age while adjusting for birth weight. In a recent PACE paper, we found 1,071 CpGs at Bonferroni significant levels association with birth weight[83]. Even after adjustment of birth weight in our gestational age EWAS, we observed overlap between the birth weight EWAS and the current gestational age EWAS for 373/1,071 CpGs (34.9% $P_{enrichment}$ <2.2x10^{-308}). These two perinatal factors, birth weight and gestational age, may have a shared impact on DNA methylation in newborns. However, it is difficult to disentangle the effects of these correlated factors.

To further investigate a potential functional impact of our differentially methylated CpGs, we examined correlations with gene expression in cord blood. We found multiple *cis*-eQTMs among the gestational age-related CpGs where methylation was strongly correlated with gene expression in cord blood, implying that the identified CpGs may have a direct functional effect in newborns. *IGF2BP1,* known to be involved in adiposity and cardiometabolic disease risk[84], and to play an essential role in embryogenesis and carcinogenesis[85,86], was the most significant positively differentially methylated CpG in cord blood. Low gestational age is a well-established risk factor for later cardiometabolic disease[87]. Our expression findings likely reflect relevant for health outcomes associated with low gestational age.

There are potential study limitations in our study including heterogeneity in normalisation and quality control (QC) protocols since individual cohorts performed their own QC and normalisation. However, one of our previous EWAS meta-analysis reported robust results comparing the non-normalised methylation and different data processing methods used across the cohorts for normalisation[20]. Furthermore, between study heterogeneity at our pre-specified threshold was observed for only a minority of differentially methylated CpGs. Cohorts collected gestational age data from medical records, birth certificates or questionnaires in two ways, either ultrasound estimates and/or according to last menstrual period (or combined estimates), which may introduce bias. However, gestational age determined by ultrasound correlates well with last menstrual period data[88]. Despite a large sample size, we had few extreme premature births included in our dataset. Interpretation of effects of DNA methylation on gene expression was done for *cis*-effects only, not *trans*-effects. Since our analyses were primarily cross-sectional, we cannot infer the temporality in the associations and we cannot assume associations are causal[89]. We recognize the possibility that the observed methylation patterns represent fetal maturity, accompanying a 'normal' developmental process or determining time *in utero*; it was however not possible to include fetuses who did not survive pregnancy most of whom will have been delivered very early. The majority of study participants were of European-ancestry and very few cohorts were Hispanic. We were unable to explore ethnic differences in detail since that would require large sample sizes for each ethnic group. However, when analyses were restricted to European-ancestry cohorts the results were essentially identical with correlation coefficient 0.97 (Additional file 3: Figure S4) to those with all cohorts included. Finally, we acknowledge a potential limitation by applying a filter (regions with at least three

706 or more adjacent CpGs with a Bonferroni-corrected p-value <0.05) in order to capture a set
707 of genes robustly affected by gestational age, which may have led to potentially important
708 single CpGs not being included in the functional analyses. In addition, genes with few CpGs
709 represented on the 450K array are likely under-represented in the downstream analyses.
710 The strengths of our study are large sample size, the comprehensive analyses using robust
711 statistical methods, as well as the availability of samples at multiple ages and our ability to
712 compare our findings with those in fetal tissue datasets. To account for potential cell type
713 effects, we adjusted our models for estimated cell counts using cord blood and adult whole
714 blood references[35,36]. However, we acknowledge the limitations of available blood cell type
715 reference data sets and recognize that some of the signals we identified as effects of
716 gestational age might reflect differences in cell type composition that we did not completely
717 control. Larger panels that better capture cell type composition across the range of
718 gestational age would be a useful advance. Although we present data on all available
719 participants in our all births model, we based our study conclusions on the main no
720 complications model results, after excluding samples related to delivery induced by medical
721 interventions (induction and/or Caesarean section) and maternal complications.
722

723 ## Conclusions

724 we show that DNA methylation at numerous CpG sites and DMRs across the genome is
725 associated with gestational age at birth. Our results provide a comprehensive catalogue of
726 differential methylation in relation to this important factor, which may serve as utility to the
727 growing community of researchers studying the developmental origins of adult disease.

Identified CpGs were linked to multiple functional pathways related to human diseases and enriched for several categories of biological processes critical to fetal development. As such, many sites might capture epigenomic plasticity of fetal development across tissues. We also found that blood DNA methylation levels in identified CpGs change over time for a majority of CpGs and that levels stabilize after school age. Taken together, our findings provide new insight into epigenetics related to preterm birth and gestational age.

Declarations

Ethics approval and consent to participate

All cohorts acquired ethics approval and informed consent from participants prior to data collection through local ethics committees; detailed information for each cohort found in Additional file 2: Supplementary information.

Consent for publication

Not applicable.

Competing financial interests

DA Lawlor declares grants from Medtronic Ltd and Roche Diagnostics and EBB; A Ghantous is identified as personnel of the IARC, the author alone is responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views

of the IARC. The remaining authors declare that they have no competing interests

Authors' contributions

EM and SJL conceived and designed the study with input from the project group (SKM, GHK, JF, M-FH, AG, NH, MW, OS, PB, JK, SER, C-JX, AC, OG, CAM, CS, AK and LKK). GCS (ALSPAC and GOYA), SKM (BAMSE, EDEN and PIAMA), RR (CBC), OS (CHAMACOS), LG (CHS), PJ (EXPOSOMICS: Environage, PiccoliPlus and RHEA), LKK (GECKO), CA (Gen3G), FOV (Generation R), LAS (INMA), FIR (IOW F1), HZ (IOW F2), SER (MoBa1 and MoBa2), AN (MoBa3), MW (NFBC86), DC (PREDO), AC (Project Viva) and PEM (Raine) conducted the cohort-specific analyses. Longitudinal analyses were performed by SKM (INMA, with support from MB) and GSC (ALSPAC). ATK performed analyses on fetal lung data sets. SKM meta-analyses all results with AN as shadow analyst. SKM performed expression and DNA methylation follow-up analyses and bioinformatics analysis. SKM, EM and SJL wrote the first draft of the manuscript. All authors (SKM, AN, GCS, LKK, ATK, RR, LG, IA, PJ, MP, MK, CA, FOV, NK, LAS, FIR, HZ, SS, DC, SLR-S, PEM, DAL, GP, CVB, KH, NB, LG, TSN, EC, PP, LD, EAN, MB, SLE, WK, SZ, CMP, ZH, M-RJ, JL, AAB, DA, PK, CLR , AB, BE, MHS, PV, HS, LB, VWJ, TIAS, MV, SHA, JWH, SEH, PM, TD, EBB, DLD, JMV, JN, KGT, IK, JLW, BH, JS, WN, MCM-K, KR, EO, R-CH, STW, JMA, JB, AK, CS, CA, AC, OG, C-JX, SER, JK, PB, OS, MW,NH, AG, M-FH, JFF, GHK, SJL, EM) read and critically revised subsequent drafts, and approved the final version. Correspondence and material requests should be addressed to EM (erik.melen@ki.se).

Acknowledgements

For all studies, acknowledgements and funding information can be found in Additional file 2: Supplementary information.

Availability of data and materials

Genome wide DNA methylation meta-analysis summary statistics corresponding to all analyses presented in this manuscript (from cohorts approving open access repository) will be uploaded in an open-access data repository after acceptance. Individual cohort level data may be available by application to the relevant institutions after obtaining required approvals.

References

1. Goldenberg, R.L., Culhane, J.F., Iams, J.D. & Romero, R. Epidemiology and causes of preterm birth. *Lancet (London, England)* **371**, 75-84 (2008).
2. Engle, W.A. Morbidity and mortality in late preterm and early term newborns: a continuum. *Clinics in perinatology* **38**, 493-516 (2011).
3. Leung, J.Y., Lam, H.S., Leung, G.M. & Schooling, C.M. Gestational Age, Birthweight for Gestational Age, and Childhood Hospitalisations for Asthma and Other Wheezing Disorders. *Paediatric and perinatal epidemiology* **30**, 149-159 (2016).
4. Raby, B.A., *et al.* Low-normal gestational age as a predictor of asthma at 6 years of age. *Pediatrics* **114**, e327-332 (2004).
5. Been, J.V., *et al.* Preterm birth and childhood wheezing disorders: a systematic review and meta-analysis. *PLoS medicine* **11**, e1001596 (2014).
6. den Dekker, H.T., *et al.* Early growth characteristics and the risk of reduced lung function and asthma: A meta-analysis of 25,000 children. *The Journal of allergy and clinical immunology* **137**, 1026-1035 (2016).
7. Parets, S.E., Bedient, C.E., Menon, R. & Smith, A.K. Preterm birth and its long-term effects: methylation to mechanisms. *Biology* **3**, 498-513 (2014).
8. Kwinta, P. & Pietrzyk, J.J. Preterm birth and respiratory disease in later life. *Expert review of respiratory medicine* **4**, 593-604 (2010).
9. Hille, E.T., *et al.* Functional outcomes and participation in young adulthood for very preterm and very low birth weight infants: the Dutch Project on Preterm and Small for Gestational Age Infants at 19 years of age. *Pediatrics* **120**, e587-595 (2007).
10. Geldof, C.J., van Wassenaer, A.G., de Kieviet, J.F., Kok, J.H. & Oosterlaan, J. Visual perception and visual-motor integration in very preterm and/or very low birth weight children: a meta-analysis. *Research in developmental disabilities* **33**, 726-736 (2012).
11. Kerkhof, G.F., Breukhoven, P.E., Leunissen, R.W., Willemsen, R.H. & Hokken-Koelega, A.C. Does preterm birth influence cardiovascular risk in early adulthood? *The Journal of pediatrics* **161**, 390-396.e391 (2012).
12. Aarnoudse-Moens, C.S., Weisglas-Kuperus, N., van Goudoever, J.B. & Oosterlaan, J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics* **124**, 717-728 (2009).
13. Thunqvist, P., *et al.* Lung Function at 8 and 16 Years After Moderate-to-Late Preterm Birth: A Prospective Cohort Study. *Pediatrics* **137**(2016).
14. Ghartey, K., *et al.* Neonatal respiratory morbidity in the early term delivery. *American journal of obstetrics and gynecology* **207**, 292.e291-294 (2012).
15. Noble, K.G., Fifer, W.P., Rauh, V.A., Nomura, Y. & Andrews, H.F. Academic achievement varies with gestational age among children born at term. *Pediatrics* **130**, e257-264 (2012).
16. Talge, N.M., Allswede, D.M. & Holzman, C. Gestational Age at Term, Delivery Circumstance, and Their Association with Childhood Attention Deficit Hyperactivity Disorder Symptoms. *Paediatric and perinatal epidemiology* **30**, 171-180 (2016).
17. Yang, S., Bergvall, N., Cnattingius, S. & Kramer, M.S. Gestational age differences in health and development among young Swedish men born at term. *International journal of epidemiology* **39**, 1240-1249 (2010).
18. Gruzieva, O., *et al.* Epigenome-Wide Meta-Analysis of Methylation in Children

Related to Prenatal NO2 Air Pollution Exposure. *Environmental health perspectives* **125**, 104-110 (2017).
19. Joubert, B.R., *et al.* Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. *Nature communications* **7**, 10577 (2016).
20. Joubert, B.R., *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet* **98**, 680-696 (2016).
21. Gruzieva, O., *et al.* Prenatal Particulate Air Pollution and DNA Methylation in Newborns: An Epigenome-Wide Meta-Analysis. *Environmental health perspectives* **127**, 57012 (2019).
22. Pan, W.H., *et al.* Exposure to the gut microbiota drives distinct methylome and transcriptome changes in intestinal epithelial cells during postnatal development. *Genome medicine* **10**, 27 (2018).
23. Cruickshank, M.N., *et al.* Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. *Genome medicine* **5**, 96 (2013).
24. Cutfield, W.S., Hofman, P.L., Mitchell, M. & Morison, I.M. Could epigenetics play a role in the developmental origins of health and disease? *Pediatric research* **61**, 68r-75r (2007).
25. Lee, H., *et al.* DNA methylation shows genome-wide association of NFIX, RAPGEF2 and MSRB3 with gestational age at birth. *International journal of epidemiology* **41**, 188-199 (2012).
26. Schroeder, J.W., *et al.* Neonatal DNA methylation patterns associate with gestational age. *Epigenetics* **6**, 1498-1504 (2011).
27. Parets, S.E., *et al.* Fetal DNA Methylation Associates with Early Spontaneous Preterm Birth and Gestational Age. *PloS one* **8**, e67489 (2013).
28. Knight, A.K., *et al.* An epigenetic clock for gestational age at birth based on blood methylation data. *Genome biology* **17**, 206 (2016).
29. Simpkin, A.J., *et al.* Longitudinal analysis of DNA methylation associated with birth weight and gestational age. *Human molecular genetics* **24**, 3752-3763 (2015).
30. Bohlin, J., *et al.* Prediction of gestational age based on genome-wide differentially methylated regions. *Genome biology* **17**, 207 (2016).
31. Felix, J.F., *et al.* Cohort Profile: Pregnancy And Childhood Epigenetics (PACE) Consortium. *International journal of epidemiology* **47**, 22-23u (2018).
32. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
33. Hoaglin, D.C., Iglewicz, B. & Tukey, J.W. Performance of Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association* **81**, 991-999 (1986).
34. Venables, W.R., B. . Modern Applied Statistics with S. *Springer* (2002).
35. Bakulski, K.M., *et al.* DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics* **11**, 354-362 (2016).
36. Reinius, L.E., *et al.* Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PloS one* **7**, e41361 (2012).
37. Aryee, M.J., *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics (Oxford, England)*

30, 1363-1369 (2014).
38. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics (Oxford, England)* **26**, 2190-2191 (2010).
39. Rice K, H.J., Lumley T. A re-evaluation of fixed effect(s) meta-analysis. *J. R. Statist. Soc. A* **181**, 205–227 (2018).
40. R Core Team and R Development Core Team.R: A language and environment for statistical computing. . *(R Foundation for Statistical Computing 2012)*.
41. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *American journal of human genetics* **88**, 586-598 (2011).
42. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Statistics in medicine* **21**, 1539-1558 (2002).
43. Hannula-Jouppi, K., *et al.* Differentially methylated regions in maternal and paternal uniparental disomy for chromosome 7. *Epigenetics* **9**, 351-365 (2014).
44. 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 (Oxford, England)* **28**, 2986-2988 (2012).
45. Peters, T.J., *et al.* De novo identification of differentially methylated regions in the human genome. *Epigenetics & chromatin* **8**, 6 (2015).
46. Chhabra, D., *et al.* Fetal lung and placental methylation is associated with in utero nicotine exposure. *Epigenetics* **9**, 1473-1484 (2014).
47. Spiers, H., *et al.* Methylomic trajectories across human fetal brain development. *Genome research* **25**, 338-352 (2015).
48. Kamburov, A., Wierling, C., Lehrach, H. & Herwig, R. ConsensusPathDB—a database for integrating human functional interaction networks. *Nucleic acids research* **37**, D623-D628 (2009).
49. Kamburov, A., *et al.* ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic acids research* **39**, D712-D717 (2011).
50. Rojas, D., *et al.* Prenatal arsenic exposure and the epigenome: identifying sites of 5-methylcytosine alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes. *Toxicological sciences : an official journal of the Society of Toxicology* **143**, 97-106 (2015).
51. Rager, J.E., *et al.* Prenatal arsenic exposure and the epigenome: altered microRNAs associated with innate and adaptive immune signaling in newborn cord blood. *Environmental and molecular mutagenesis* **55**, 196-208 (2014).
52. Barrett, T., *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic acids research* **41**, D991-995 (2013).
53. Hannon, E., *et al.* Variable DNA methylation in neonates mediates the association between prenatal smoking and birth weight. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **374**, 20180120 (2019).
54. Ziller, M.J., *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477-481 (2013).
55. Minelli, C., *et al.* Association of Forced Vital Capacity with the Developmental Gene NCOR2. *PloS one* **11**, e0147388 (2016).
56. Garg, B.D., Bansal, A. & Kabra, N.S. Role of vitamin A supplementation in prevention of bronchopulmonary dysplasia in extremely low birth weight neonates: a

systematic review of randomized trials. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet*, 1-8 (2018).
57. Gasparoni, G., *et al.* DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenetics & chromatin* **11**, 41 (2018).
58. Ferreira, M.A.R., *et al.* Eleven loci with new reproducible genetic associations with allergic disease risk. *The Journal of allergy and clinical immunology* (2018).
59. Wang, X., Ning, Y. & Guo, X. Integrative meta-analysis of differentially expressed genes in osteoarthritis using microarray technology. *Molecular medicine reports* **12**, 3439-3445 (2015).
60. Al Muftah, W.A., *et al.* Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clinical epigenetics* **8**, 13 (2016).
61. Korpos, E., Deak, F. & Kiss, I. Matrilin-2, an extracellular adaptor protein, is needed for the regeneration of muscle, nerve and other tissues. *Neural regeneration research* **10**, 866-869 (2015).
62. Zhang, S., *et al.* Matrilin-2 is a widely distributed extracellular matrix protein and a potential biomarker in the early stage of osteoarthritis in articular cartilage. *BioMed research international* **2014**, 986127 (2014).
63. Cuajungco, M.P., *et al.* Abnormal accumulation of human transmembrane (TMEM)-176A and 176B proteins is associated with cancer pathology. *Acta histochemica* **114**, 705-712 (2012).
64. Xu, C.J., *et al.* The emerging landscape of dynamic DNA methylation in early childhood. *BMC genomics* **18**, 25 (2017).
65. Acevedo, N., *et al.* Age-associated DNA methylation changes in immune genes, histone modifiers and chromatin remodeling factors within 5 years after birth in human blood leukocytes. *Clinical epigenetics* **7**, 34 (2015).
66. Simpkin, A.J., *et al.* Prenatal and early life influences on epigenetic age in children: a study of mother-offspring pairs from two cohort studies. *Human molecular genetics* **25**, 191-201 (2016).
67. Horvath, S. DNA methylation age of human tissues and cell types. *Genome biology* **14**, R115 (2013).
68. Hannum, G., *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* **49**, 359-367 (2013).
69. Goyal, N.K., Fiks, A.G. & Lorch, S.A. Association of late-preterm birth with asthma in young children: practice-based study. *Pediatrics* **128**, e830-838 (2011).
70. Sonntag, B., *et al.* Preterm birth but not mode of delivery is associated with an increased risk of developing inflammatory bowel disease later in life. *Inflammatory bowel diseases* **13**, 1385-1390 (2007).
71. Li, S., *et al.* Preterm birth and risk of type 1 and type 2 diabetes: systematic review and meta-analysis. *Obesity reviews : an official journal of the International Association for the Study of Obesity* **15**, 804-811 (2014).
72. Wang, Y.F., Wu, L.Q., Liu, Y.N., Bi, Y.Y. & Wang, H. Gestational age and childhood leukemia: A meta-analysis of epidemiologic studies. *Hematology (Amsterdam, Netherlands)* **23**, 253-262 (2018).
73. Reese, S.E., *et al.* Epigenome-wide Meta-analysis of DNA Methylation and

Childhood Asthma. *The Journal of allergy and clinical immunology* (2018).
74. Maccani, J.Z., Koestler, D.C., Houseman, E.A., Marsit, C.J. & Kelsey, K.T. Placental DNA methylation alterations associated with maternal tobacco smoking at the RUNX3 gene are also associated with gestational age. *Epigenomics* **5**, 619-630 (2013).
75. Yang, I.V., *et al.* DNA methylation and childhood asthma in the inner city. *The Journal of allergy and clinical immunology* **136**, 69-80 (2015).
76. Zheng, Y., Zhang, Y., Huang, X. & Chen, L. Analysis of the RUNX3 gene methylation in serum DNA from esophagus squamous cell carcinoma, gastric and colorectal adenocarcinoma patients. *Hepato-gastroenterology* **58**, 2007-2011 (2011).
77. de Vega, W.C., Herrera, S., Vernon, S.D. & McGowan, P.O. Epigenetic modifications and glucocorticoid sensitivity in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *BMC medical genomics* **10**, 11 (2017).
78. Reese, S.E., *et al.* DNA Methylation Score as a Biomarker in Newborns for Sustained Maternal Smoking during Pregnancy. *Environmental health perspectives* **125**, 760-766 (2017).
79. Valeri, L., *et al.* Misclassified exposure in epigenetic mediation analyses. Does DNA methylation mediate effects of smoking on birthweight? *Epigenomics* **9**, 253-265 (2017).
80. Warren, G.W., Alberg, A.J., Kraft, A.S. & Cummings, K.M. The 2014 Surgeon General's report: "The health consequences of smoking--50 years of progress": a paradigm shift in cancer care. *Cancer* **120**, 1914-1916 (2014).
81. Sharp, G.C., *et al.* Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: findings from the pregnancy and childhood epigenetics (PACE) consortium. *Human molecular genetics* **26**, 4067-4085 (2017).
82. Sharp, G.C., *et al.* Maternal alcohol consumption and offspring DNA methylation: findings from six general population-based birth cohorts. *Epigenomics* **10**, 27-42 (2018).
83. Kupers, L.K., *et al.* Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nature communications* **10**, 1893 (2019).
84. Lu, Y., *et al.* New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nature communications* **7**, 10495 (2016).
85. Mahaira, L.G., *et al.* IGF2BP1 expression in human mesenchymal stem cells significantly affects their proliferation and is under the epigenetic control of TET1/2 demethylases. *Stem cells and development* **23**, 2501-2512 (2014).
86. Huang, X., *et al.* Insulin-like growth factor 2 mRNA-binding protein 1 (IGF2BP1) in cancer. *Journal of hematology & oncology* **11**, 88 (2018).
87. Cooper, R., Atherton, K. & Power, C. Gestational age and risk factors for cardiovascular disease: evidence from the 1958 British birth cohort followed to mid-life. *International journal of epidemiology* **38**, 235-244 (2009).
88. Hoffman, C.S., *et al.* Comparison of gestational age at birth based on last menstrual period and ultrasound during the first trimester. *Paediatric and perinatal epidemiology* **22**, 587-596 (2008).
89. Dyke, S.O.M., *et al.* Points-to-consider on the return of results in epigenetic research. *Genome medicine* **11**, 31 (2019).

Figure legends

Figure 1. An overview of the study design.

Figure 2A, B: Volcano (A) and Manhattan (B) plots for the meta-analysis of gestational age and offspring DNA methylation association at birth, after adjustment for covariates and estimated cell proportions. The effect size represents methylation change per gestational week.

Figure 3. Position enrichment analyses for CpGs. Black: all CpGs in the Illumina450k annotation file, grey: CpGs significantly associated with GA after Bonferroni correction ($p<1.06\times10^{-7}$) and light grey: three or more adjacent CpGs associated with GA after Bonferroni correction ($p<1.06\times10^{-7}$). "**" represent significant two-sided doubling mid p-value of the hypergeometric test.

Figure 4. Overlap between Bonferroni-significant CpG sites from two different analyses after exclusion of maternal and delivery start with induction or caesarean section ("no complication" model). The blue color represents the continuous gestational age main model, the green represents the continuous model restricted to term only. Overlap of findings alters the color.

Figure 5. Change in DNA methylation during childhood and adolescence for selected CpG sites associated with gestational age. A: Decreasing methylation levels from birth to childhood (A.1) and stabilization during adolescence (A.2). B: Increasing methylation levels from birth to childhood and stabilization during adolescence. C: Stable CpGs that did not change during childhood or adolescence; (1) INMA from birth to early childhood, (2) ALSPAC from birth to adolescence. The figures show representative single CpGs for each category (A-C).

Table 1. Characteristics of each cohort included in the association meta-analysis between gestational age (GA) and DNA methylation in newborns and older children.

Study population	Cohort	N	N, PRE-TERM*	N TERM	Age mean (SD)	Maternal age mean (SD)	Mean GA (Days)	SD GA	Min GA	Max GA	Ethnicity
Newborn	ALSPAC**	249	10	239	0	29.8 (4.6)	277	10.78	224	294	European
	CBC (Hispanic)	128	10	118	0	27.3 (5.8)	273	17.70	196	294	Hispanic
	CBC (European)	132	11	121	0	31.9 (5.7)	273	16.10	189	294	European
	CHS	120	7	113	0	29.4 (5.6)	277	11.20	230	294	Mixed
	CHAMACOS	110	11	99	0	25.3 (5.0)	272	10.66	210	294	Hispanic
	EDEN	100	2	98	0	30.8 (5.0)	276	10.11	217	287	European
	EXPOSOMICS (Environage + PiccoliPlus + RHEA)	252	17	235	0	30.5 (4.8)	273	10.50	217	294	European
	Generation R	486	22	464	0	31.9 (4.2)	280	9.00	239	294	European
	INMA	134	2	132	0	30.5 (4.1)	278	9.57	234	286	European
	IOW F2	93	2	91	0	23.2 (2.6)	278	10.95	236	294	European
	MoBa1**	749	18	731	0	29.9 (4.3)	279	10.36	209	294	European
	MoBa2**	460	15	445	0	30.0 (4.5)	278	10.49	209	294	European
	MoBa3	177	3	174	0	29.6 (4.4)	279	10.38	199	294	European
	PREDO	308	5	303	0	33.4 (5.7)	278	11.20	186	294	European
	Project Viva	150	3	147	0	33.2 (4.5)	278	10.11	216	294	European
	Meta-analysis	**3648**	**138**								
Early childhood	BAMSE	145	10	135	4.3 (0.2)	31.2 (4.4)	275	16.22	187	293	European
	EDEN	89	2	87	5.6 (0.1)	30.8 (5.1)	276	9.23	245	287	European
	INMA	71	1	70	4.4 (0.2)	30.6 (4.3)	279	8.70	249	288	European
	PIAMA	148	4	144	4.1 (0.2)	30.6 (3.6)	278	10.51	233	294	European
	Meta-analysis	**453**	**17**								
School age	ALSPAC	273	12	261	7.5 (0.1)	29.9 (4.6)	277	10.99	224	294	European
	BAMSE	141	10	131	8.4 (0.4)	31.4 (4.5)	276	15.96	197	293	European
	BAMSE_EpiGene	232	8	224	8.3 (0.5)	30.8 (4.4)	278	11.47	209	294	European
	PIAMA	134	3	131	8.1 (0.3)	30.5 (3.6)	278	10.61	233	294	European
	Project Viva	119	2	117	7.8 (0.7)	33.5 (4.4)	278	10.32	216	294	European
	Meta-analysis	**899**	**35**								
Adolescence	ALSPAC	272	13	259	17.2 (1.0)	29.9 (4.6)	277	11.04	224	294	European
	BAMSE	159	7	152	16.7 (0.4)	31.2 (4.4)	278	12.70	187	294	European
	IOW F1	97	2	95	17.1 (0.5)	27.1 (5.1)	280	9.83	238	294	European
	NFBC86	287	9	276	16.1 (0.4)	29.0 (5.1)	280	8.65	237	294	European
	RAINE	314	9	305	17.0 (0.3)	29.0 (5.8)	274	11.90	196	294	European
	Meta-analysis	**1129**	**40**								

*Preterm birth categorised as GA less than 37 full weeks or 259 days and as term greater than 37 weeks or 259 days (but less than 42 full weeks).

** This study was included previous EWAS of gestational age[29,30]

Table 2. The top 10 Bonferroni-significant CpGs from the meta-analysis on the association between continuous GA and offspring DNA methylation at birth adjusted for estimated cell proportions.

CpGID	Chr	Genomic coordinates	Gene (Illumina annotation)	Relation to Island	Distance to nearest gene	UCSC Known Gene	Coefficient*	PVALUE	Direction of effect in each cohort**
cg16103712	8	99023869	*MATN2*	OpenSea	7355	*MATN2*	-0.0030	2.70E-129	----------
cg04685228	5	172462626		OpenSea	726	*ATP6V0E1*	-0.0028	8.55E-109	------?--------
cg04276536	16	57567813	*CCDC102A*	N_Shelf	0	*CCDC102A*	-0.0012	1.20E-93	------?--------
cg19744173	2	112913178	*FBLN7*	N_Shelf	0	*FBLN7*	-0.0016	4.91E-92	----------
cg27518892	16	57566936	*CCDC102A*	N_Shelf	0	*CCDC102A*	-0.0018	1.29E-89	----------
cg13924996	11	67053829	*ADRBK1*	S_Shore	0	*ADRBK1*	-0.0016	8.59E-89	------?--------
cg04494800	6	149775853	*ZC3H12D*	N_Shore	1923	*ZC3H12D*	-0.0016	4.52E-82	------?--------
cg27295118	14	22902226		OpenSea	-500	*AK125397*	-0.0024	1.20E-81	------?--------
cg26433582	11	68848232	*TPCN2*	N_Shore	917	*TPCN2*	-0.0019	1.31E-81	------?--------
cg18183624	17	47076904	*IGF2BP1*	S_Shore	0	*IGF2BP1*	0.0028	8.36E-80	+++++++++++++++

* Coefficient corresponding to methylation change per additional day of gestational age.

** Order of included cohorts in the meta-analysis: MoBa1, MoBa2, MoBa3, EDEN, EXPOSOMICS (Environage+PiccoliPlus+RHEA), CHS, IOWF2, Generation R, Project Viva, CBC (Hispanic), CBC (White), ALSPAC, PREDO, CHAMACOS and INMA."?" Means that CpG was not measured in that cohort.

Table 3. The top 10 Bonferroni-significant CpGs ranked by the magnitude of positive and negative effect (5 CpGs each) from the meta-analysis on the association between continuous GA and offspring DNA methylation at birth adjusted for estimated cell proportions.

CpGID	Chr	Genomic coordinates	Gene (Illumina annotation)	Relation to Island	Distance to nearest gene	UCSC Known Gene	Coefficient*	PVALUE	Direction of effect in each cohort**
cg13036381	3	1.6E+08	*LOC401097*	N_Shore	-927	*C3orf80*	0.00278	1.01E-47	+++++-+++++++++
cg18183624	17	47076904	*IGF2BP1*	S_Shore	0	*IGF2BP1*	0.00277	8.36E-80	++++++++++++++++
cg04213841	13	49792685	*NA*	N_Shore	-1788	*MLNR*	0.00245	3.60E-43	+++++?+++++++++
cg07738730	17	47077165	*IGF2BP1*	S_Shore	0	*IGF2BP1*	0.00217	2.87E-65	++++++++++++++-+
cg09476997	16	2087932	*SLC9A3R2*	N_Shore	0	*SLC9A3R2*	0.00208	2.41E-49	++++++++++++++++
cg04347477	12	1.25E+08	*NCOR2*	Island	833	*NCOR2*	-0.00361	3.38E-32	----------
cg08943494	11	36422615	*PRR5L*	OpenSea	69	*PRR5L*	-0.00360	1.95E-24	----------
cg20334115	1	2.26E+08	*PYCR2*	N_Shelf	0	*PYCR2*	-0.00350	1.40E-35	----------
cg16725984	16	89735184	*C16orf55*	Island	0	*C16orf55*	-0.00325	3.70E-26	----------
cg16103712	8	99023869	*MATN2*	OpenSea	7355	*MATN2*	-0.00304	2.70E-129	----------

* Coefficient corresponding to methylation change per additional day of gestational age.

** Order of included cohorts in the meta-analysis: MoBa1, MoBa2, MoBa3, EDEN, EXPOSOMICS (Environage+PiccoliPlus+RHEA), CHS, IOWF2, Generation R, Project Viva, CBC (Hispanic), CBC (White), ALSPAC, PREDO, CHAMACOS and INMA."?" Means that CpG was not measured in that cohort.

Additional files

Additional file 1:
Table S1: Cohort-specific results from epigenome-wide association analyses of gestational age.
Table S2: Normalisation technique and phenotype definitions used by each cohort.
Table S3: Bonferroni-significant CpGs from the meta-analysis on the association between continuous gestational age (no complications model) and offspring DNA methylation at birth adjusted for estimated cell counts
Table S4: Bonferroni-significant CpGs from the meta-analysis on the association between continuous gestational age (all births model) and offspring DNA methylation at birth adjusted for estimated cell counts.
Table S5: Gene regions that had at least three consecutive Bonferroni significant CpG sites from the continuous gestational age analyses (no complications model).
Table S6: DMRs (n = 2,375) for gestational age in relation to newborn methylation (no complication model) identified by using both comb-p ($P < 0.01$) and DMRcate (FDR < 0.01) methods.
Table S7: DNA methylation analyses in fetal lung tissue using the no complication gestational age three or more consecutive CpG list.
Table S8: DNA methylation analyses in fetal brain tissue using the no complication gestational age three or more consecutive CpG list.
Table S9: Methylation look-up analyses in older children using the no complication gestational age three or more consecutive CpG list.
Table S10: Longitudinal analysis of methylation levels in the INMA and ALSPAC studies using the no complication gestational age three or more consecutive CpG list.
Table S11: Gene Ontology (GO) term enrichment analyses for bonferroni-significant CpGs from the meta-analysis (no complications model).
Table S12: KEGG pathway analyses for bonferroni-significant CpGs from the meta-analysis (no complications model).
Table S13: Gene Ontology (GO) term enrichment analyses for three or more CpGs being localized to the same gene.
Table S14: KEGG pathway analyses for stable and dynamic CpGs.
Table S15: Correlation between methylation and gene expression levels in cord blood (cis-effects).
Table S16: The replication of bonferroni-significant CpGs from the meta-analysis (no complications model) in previous publication. (XLSX 6.97 MB)

Additional file 2:
Supplementary information. (. word 301 KB)

Additional file 3:
Figure S1. Forest plot for the top 10 Bonferroni-significant CpGs from the meta-analysis on the association between continuous GA and offspring DNA methylation at birth adjusted for estimated cell proportions.

Figure S2. Sensitivity analysis: Correlation of the point estimates for the no complications model main association of DNA methylation with gestational age (y-axis representing 3,648 participants from 17 cohorts) with point estimates for a meta-analysis after excluding three cohorts (MoBa1, MoBa2 and ALSPAC) that were included in a previous publication1,2 (x-axis representing 2,190 participants from 14 cohorts).

Figure S3. Correlations between methylation and gene expression levels for selected four pairs. First, we created residuals for mRNA expression and residuals for DNA methylation and used linear regression models to evaluate correlations between expression residuals and methylation residuals. These residual models were adjusted for covariates, estimated white blood cell proportions, and technical variation.

Figure S4. Sensitivity analysis: Correlation of the point estimates for the no complications model main association of DNA methylation with gestational age (y-axis representing 3,648 participants from 17 cohorts) with point estimates for a meta-analysis after excluding Non-European three cohorts (CBC, CHS and CHAMACOS) (x-axis representing 3,290 participants from 14 cohorts). (. word 181 KB)