Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight

Leanne K. Küpers[1–3], Claire Monnereau[4–6], Gemma C. Sharp[2,7], Paul Yousefi[2,3,8], Lucas A. Salas[9,10], Akram Ghantous[11], Christian M. Page[12,13], Sarah E. Reese[14], Allen J. Wilcox[14], Darina Czamara[15], Anne P. Starling[16], Alexei Novoloaca[11], Samantha Lent[17], Ritu Roy[18,19], Cathrine Hoyo[20,21], Carrie V. Breton[22], Catherine Allard[23], Allan C. Just[24], Kelly M. Bakulski[25], John W. Holloway[26,27], Todd M. Everson[28], Cheng-Jian Xu[29,30], Rae-Chi Huang[31], Diana A. van der Plaat[32], Matthias Wielscher[33], Simon Kebede Merid[34], Vilhelmina Ullemar[35], Faisal I. Rezwan[27], Jari Lahti[36,37], Jenny van Dongen[38], Sabine A.S. Langie[39–41], Tom G. Richardson[2,3], Maria C. Magnus[2,3,12], Ellen A. Nohr[42], Zongli Xu[14], Liesbeth Duijts[4,43,44], Shanshan Zhao[45], Weiming Zhang[46], Michelle Plusquin[47,48], Dawn L. DeMeo[49], Olivia Solomon[8], Joosje H. Heimovaara[1], Dereje D. Jima[21,50], Lu Gao[22], Mariona Bustamante[10,51–53], Patrice Perron[23,54], Robert O. Wright[24], Irva Hertz-Picciotto[55], Hongmei Zhang[56], Margaret R. Karagas[9,57], Ulrike Gehring[58], Carmen J. Marsit[28], Lawrence J. Beilin[59], Judith M. Vonk[32], Marjo-Riitta Jarvelin[33,60–62], Anna Bergström[34,63], Anne K. Örtqvist[35], Susan Ewart[64], Pia M. Villa[65], Sophie E. Moore[66,67], Gonneke Willemsen[38], Arnout R.L. Standaert[39], Siri E. Håberg[12], Thorkild I.A. Sørensen[2,68,69], Jack A. Taylor[14], Katri Räikkönen[37], Ivana V. Yang[70], Katerina Kechris[46], Tim S. Nawrot[47,71], Matt J. Silver[66], Yun Yun Gong[72], Lorenzo Richiardi[73,74], Manolis Kogevinas[10,52,53,75], Augusto A. Litonjua[49], Brenda Eskenazi[8,76], Karen Huen[8], Hamdi Mbarek[77], Rachel L. Maguire[20,78], Terry Dwyer[79], Martine Vrijheid[10,52,53], Luigi Bouchard[80,81], Andrea A. Baccarelli[82,83], Lisa A. Croen[84], Wilfried Karmaus[56], Denise Anderson[31], Maaike de Vries[32], Sylvain Sebert[60,61,85], Juha Kere[86–88], Robert Karlsson[35], Syed Hasan Arshad[26,89], Esa Hämäläinen[90], Michael N. Routledge[91], Dorret I. Boomsma[38,92], Andrew P. Feinberg[93], Craig J. Newschaffer[94], Eva Govarts[39], Matthieu Moisse[95,96], M. Daniele Fallin[97], Erik Melén[34,98], Andrew M. Prentice[66], Eero Kajantie[99–101], Catarina Almqvist[35,102], Emily Oken[103], Dana Dabelea[104], H. Marike Boezen[32], Phillip E. Melton [105,106], Rosalind J. Wright[24], Gerard H. Koppelman[29], Letizia Trevisi[107], Marie-France Hivert[54,103,108], Jordi Sunyer[10,52,53,75], Monica C. Munthe-Kaas[109,110], Susan K. Murphy[111], Eva Corpeleijn[1], Joseph Wiemels[112], Nina Holland[8], Zdenko Herceg[11], Elisabeth B. Binder[15,113], George Davey Smith[2,3], Vincent W.V. Jaddoe[4–6], Rolv T. Lie[114], Wenche Nystad[115], Stephanie J. London[14], Debbie A. Lawlor[2,3]#*, Caroline L. Relton[2,3]#*, Harold Snieder[1]#*, Janine F. Felix[4–6]#*

* To whom correspondence should be addressed:

j.felix@erasmusmc.nl
+31 (0)107043997
P.O. Box 2040, room NA-2818, 3000 CA Rotterdam, the Netherlands

h.snieder@umcg.nl
+31 (0)503610738
PO Box 30.001, 9700 RB Groningen, The Netherlands

caroline.relton@bristol.ac.uk
+44 (0)1173310098
Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom

d.a.Lawlor@bristol.ac.uk
+44 (0)1173310098
Oakfield House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom

# equal contributions as last author

Word count (incl. methods): 4941 words

Affiliations:

1. Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.
2. MRC Integrative Epidemiology Unit, University of Bristol , Bristol, United Kingdom.
3. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom.
4. The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
5. Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
6. Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
7. School of Oral and Dental Sciences, University of Bristol, Bristol, UK.
8. Children's Environmental Health Laboratory, Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley.
9. Department of Epidemiology, Geisel School of Medicine at Dartmouth College, Hanover, NH, USA.
10. ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain.
11. Epigenetics Group, International Agency for Research on Cancer, Lyon, France.
12. Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway.
13. Oslo Centre for Biostatisitcs and Epidemology, Oslo University Hospital, Oslo, Norway.
14. Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, United States of America.
15. Department of Translational Research in Psychiatry, Max-Planck-Institute of Psychiatry, Munich, Germany.
16. Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA.
17. Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.
18. HDF Comprehensive Cancer Center, University of California, San Francisco, USA.
19. Computational Biology Core.
20. Department of Biological Sciences, North Carolina State University, Raleigh, NC, USA.
21. Center for Human Health and the Environment, North Carolina State University, Raleigh, NC, USA.
22. Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA.
23. Centre de recherche du Centre hospitalier universitaire de Sherbrooke, QC, Canada.
24. Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, United States.
25. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA.
26. Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.
27. Human Development and Health, Faculty of Medicine, University of Southampton, Southampton General Hospital, Southampton, UK.
28. Department of Environmental Health, Rollins School of Public Health at Emory University, Atlanta, GA, USA.
29. University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, GRIAC Research Institute Groningen, The Netherlands.
30. University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands.
31. Telethon Kids Institute, University of Western Australia, Perth, Australia.
32. Department of Epidemiology and Groningen Research Institute for Asthma and COPD (GRIAC),

University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
33. Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, London, UK.
34. Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
35. Department of Medical Epidemiology and Biostatistics, and the Karolinska Institutet, Stockholm, Sweden.
36. Helsinki Collegium for Advanced Studies, University of Helsinki, Helsinki, Finland.
37. Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Finland.
38. Department of Biological Psychology, Netherlands Twin Register, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.
39. VITO - Sustainable Health, Mol, Belgium.
40. Theoretical Physics, Faculty of Sciences, Hasselt University, Hasselt, Belgium.
41. Centre for Environmental Sciences, Hasselt University, Hasselt, Belgium.
42. Research Unit for Gynaecology and Obstetrics, Department of Clinical Research, University of Southern Denmark, Odense, DK.
43. Department of Pediatrics, Division of Respiratory Medicine and Allergology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
44. Department of Pediatrics, Division of Neonatology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.
45. Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, North Carolina, United States of America.
46. Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA.
47. Centre for Environmental Sciences, Hasselt University, Diepenbeek, Belgium.
48. MRC/PHE Centre for Environment and Health School of Public Health Imperial College London, St Mary's Campus, Norfolk Place, London, UK.
49. Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
50. Bioinformatics Research Center, North Carolina State University, Raleigh, NC, USA.
51. Genomics and Disease Group, Bioinformatics and Genomics Program, Centre for Genomic Regulation (CRG), Barcelona, Spain.
52. Universitat Pompeu Fabra (UPF), Barcelona, Spain.
53. CIBER Epidemiología y Salud Pública (CIBERESP), Spain.
54. Department of Medicine, Universite de Sherbrooke, QC, Canada.
55. Department of Public Health Sciences, School of Medicine, University of California Davis MIND Institute, California, USA.
56. Division of Epidemiology, Biostatistics and Environmental Health, School of Public Health, University of Memphis, Memphis, USA.
57. Children's Environmental Health & Disease Prevention Research Center at Dartmouth, Hanover, NH, USA.
58. Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands.
59. Medical School. University of Western Australia. Perth, Australia.
60. Center for Life Course Health Research, Faculty of Medicine, University of Oulu, 90014 Oulu, Finland.
61. Biocenter Oulu, University of Oulu, Finland.
62. Unit of Primary Care, Oulu University Hospital, Oulu, Finland.
63. Center for Occupational and Environmental Medicine, Stockholm County Council, Stockholm, Sweden.
64. College of Veterinray Medicine, Michigan State University, East Lansing, MI, USA.
65. Obstetrics and Gynaecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.
66. Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine, London, United Kingdom.
67. Department of Women and Children's Health, King's College London, London, United Kingdom.
68. Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
69. Department of Public Health, Section of Epidemiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark.
70. Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA.

71. Department of Public Health & Primary Care, Leuven University, Leuven, Belgium.
72. School of Food Sciences and Nutrition, University of Leeds, Leeds, United Kingdom.
73. Department of Medical Sciences, University of Turin, Turin, Italy.
74. AOU Citta della Salute e della Sceinza, CPO Piemonte, Turin, Italy.
75. IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain.
76. Center for Environmental Research and Children's Health, School of Public Health, University of California, Berkeley.
77. Department of Biological Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.
78. Department of Community and Family Medicine, Duke University Medical Center, Raleigh, NC, USA.
79. The George Institute for Global Health, Nuffield Department of Obstetrics & Gynaecology, University of Oxford, Oxford, UK.
80. Department of Biochemistry, Université de Sherbrooke, QC, Canada.
81. ECOGENE-21 Biocluster, Chicoutimi Hospital, Saguenay, QC, Canada.
82. Laboratory of Precision Environmental Biosciences, Columbia University Mailman School of Public Health, New York, NY, USA.
83. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, United States.
84. Division of Research, Kaiser Permanente Northern California, California, USA.
85. Department for Genomics of Common Diseases, School of Public Health, Imperial College London, UK.
86. Folkhälsan Institute of Genetics, Helsinki, and Research Programs Unit, Molecular Neurology, University of Helsinki, Helsinki, Finland.
87. Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden.
88. School of Basic and Medical Biosciences, King's College London, Guy's Hospital, London, UK.
89. David Hide Asthma and Allergy Research Centre, Isle of Wight, United Kingdom.
90. HUSLAB and the Department of Clinical Chemistry, University of Helsinki, Helsinki, Finland.
91. LICAMM, School of Medicine, Univeristy of Leeds, Leeds, United Kingdom.
92. Amsterdam Public Health Institute, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.
93. Center for Epigenetics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.
94. AJ Drexel Autism Institute, Drexel University, Philadelphia, Pennsylvania, USA.
95. KU Leuven – University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Institute for Neuroscience and Disease (LIND), Leuven, Belgium.
96. VIB, Center for Brain & Disease Research, Laboratory of Neurobiology, Leuven, Belgium.
97. Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA.
98. Sachs' Children's Hospital, Stockholm, Sweden.
99. National Institute for Health and Welfare, Helsinki and Oulu, Finland.
100. Hospital for Children and Adolescents, Helsinki University Hospital and University of Helsinki, Helsinki, Finland.
101. PEDEGO Research Unit, MRC Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland.
102. Pediatric Allergy and Pulmonology Unit at Astrid Lindgren Children's Hospital, Karolinska University Hospital, Stockholm, Sweden.
103. Department of Population Medicine, Harvard Medical School, Harvard Pilgrim Health Care Institute, Boston, MA, USA.
104. Department of Epidemiology, Colorado School of Public Health, and Department of Pediatrics, University of Colorado School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA.
105. Centre for Genetic Origins of Health and Disease, School of Biomedical Sciences, University of Western Australia, Perth, Australia.
106. School of Pharmacy and Biomedical Sciences, Curtin University, Perth, Australia.
107. Department of Global Health and Social Medicine, Harvard Medical School, Boston, United States.
108. Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA.
109. Norwegian Institute of Public Health, Oslo, Norway.
110. Department of Pediatric Oncology and Hematology, Oslo University Hospital, Norway.
111. Department of Obstetrics and Gynecology, Duke University Medical Center.
112. Department of Epidemiology and Biostatistics, University of California, San Francisco, USA.
113. Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, USA.

114. Norwegian Institute of Public Health, Oslo, Norway, Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway.
115. Department for Non-Communicable Diseases, Norwegian Institute for Public Health, Oslo, Norway.

Abstract

Birthweight is associated with health outcomes across the life course, DNA methylation may be an underlying mechanism. In this meta-analysis of epigenome-wide association studies of 8,825 neonates from 24 birth cohorts in the Pregnancy And Childhood Epigenetics Consortium, DNA methylation in neonatal blood is associated with birthweight at 914 sites, with a difference in birthweight ranging from -183 to 178 grams per 10% increase in methylation ($P_{Bonferroni}<1.06x10^{-7}$). In additional analyses in 7,278 participants, <1.3% of birthweight-associated differential methylation is also observed in childhood and adolescence, but not adulthood. Birthweight-related CpGs overlap with some CpGs that were previously reported to be related to maternal smoking (55/914, $p_{enrichment}=6.12x10^{-74}$) and BMI in pregnancy (3/914, $p_{enrichment}=1.13x10^{-3}$), but not with those related to famine or folate levels in pregnancy. Whether the associations that we observe are causal or explained by confounding or fetal growth influencing DNA methylation (i.e. reverse causality) requires further research.

Introduction

Intrauterine exposures, such as maternal smoking, pre-pregnancy body mass index (BMI), hyperglycaemia, hypertension, folate and famine are associated with fetal growth and hence birthweight[1–6]. Observational studies show that birthweight is also associated with later-life health outcomes, including cardio-metabolic and mental health, some cancers and mortality[7–11]. In these long term associations, birthweight may act as a proxy for potential effects of intrauterine exposures[12,13]. Several mechanisms may explain the associations of intrauterine exposures with birthweight and later-life health as we illustrate in **Figure 1**. Our overall conceptual framework in this study was that the intrauterine environment induces epigenetic alterations, which influence fetal growth and hence correlate with birthweight. This is partly supported by previous large-scale epigenome-wide association studies (EWAS) that have reported associations of relevant maternal pregnancy exposures, including smoking, air pollution and BMI, with DNA methylation in offspring neonatal blood[14–16]. However, whilst four previous EWAS have observed associations of DNA methylation with birthweight[18–21], the evidence to date has been limited in scale and power with sample sizes ranging from approximately 200 to 1,000.

In this study, we hypothesised that there are associations between DNA methylation and birthweight. We further aimed to explore if these epigenetic alterations are associated with later disease outcomes (**Figure 1**). If birthweight is a proxy for a range of adverse prenatal exposures, we might expect neonatal blood DNA methylation to be associated with birthweight. However, we acknowledge that any associations of DNA methylation with birthweight may be explained by confounding[17] or reflect reverse

We present a large meta-analysis of multiple EWAS to explore associations between neonatal blood DNA methylation and birthweight. In further analyses, we explore whether any birthweight-associated differential methylation persists at older ages. To aid functional interpretation we: (i) explore the overlap of identified cytosine-phosphate-guanine sites (CpGs) that

256 are differentially methylated in relation to birthweight with those known to be associated with
257 intrauterine exposure to smoking, famine and different levels of BMI and folate; (ii) associate DNA
258 methylation at identified CpGs with gene expression and (iii) explore potential causal links with
259 birthweight and later-life health using Mendelian randomization (MR)[22]. We show that DNA
260 methylation in neonatal blood is associated with birthweight and some of the differential
261 methylation is also observed in childhood and adolescence, but not in adulthood. Also, we show
262 overlap between birthweight-related CpGs and CpGs related to intrauterine exposures. Potential
263 causality of the associations needs to be studied further.

Results

Participants

We used data from 8,825 neonates from 24 studies in the Pregnancy And Childhood Epigenetics (PACE) Consortium, representing mainly European, but also African and Hispanic ethnicities with similar proportions of males and females. Details of participants used in all analyses are presented in **Table 1, Supplementary Table 1** and study-specific **Supplementary Methods**.

Meta-analysis

Primary, secondary and follow-up analyses are outlined in the study design in **Figure 2**. Methylation at 8,170 CpGs, measured in neonatal blood using the Illumina Infinium® HumanMethylation450 BeadChip assay and adjusted for cell type heterogeneity[23–25], was associated with birthweight (False Discovery Rate (FDR) < 0.05), of which 1,029 located in or near 807 genes survived the more stringent Bonferroni correction ($p<1.06\text{x}10^{-7}$, **Supplementary Table 2**). We observed both positive (45%) and negative (55%) directions of associations between methylation levels of these 1,029 CpGs and birthweight (**Figure 3**) and these CpGs were spread throughout the genome (orange track (1) in **Figure 4** and **Supplementary Figure 1**). We found evidence of between-study heterogeneity ($I^2>50\%$) for 115 of the 1,029 sites (**Supplementary Table 2**), thus we prioritised 914 CpGs, located in or near 729 genes, based on $p<1.06\text{x}10^{-7}$ and $I^2>50\%$ for further analyses (**Figure 3** and orange track (1) in **Figure 4**). The CpG with the largest positive association was cg06378491 (in the gene body of *MAP4K2*). For each 10% increase in methylation at this site, birthweight was 178g higher (95% confidence interval (CI): 138, 218g). The CpG with the largest negative association was cg10073091 (in the gene body of *DHCR24*), which showed a 183g decrease in birthweight per 10% increase in methylation (95% CI: -225, -142g). The CpG with the smallest P-value and $I^2\leq50\%$ was cg17714703 (in the gene body of *UHRF1*), which showed a 130g increase in birthweight for 10% increase in methylation (95% CI: 109, 151g).

Findings were consistent with results from our main analyses when restricted to participants of European ethnicity, with a Pearson correlation coefficient for effect estimates of 0.99 for the 914 birthweight-associated CpGs (**Supplementary Figure 2**, blue track (2) in **Figure 4** and **Supplementary Table 3**) and 0.90 for all 450k CpGs. Comparing the main meta-analyses to the four Hispanic cohorts and the two African cohorts revealed that 94.9% and 74.0% of the 914 CpGs showed consistent direction of association, with Pearson correlation coefficients for point estimates of 0.82 and 0.48, respectively (**Supplementary Table 3**). In leave-one-out analyses, in which we reran the main meta-analysis repeatedly with one of the 24 studies removed each time, there was no strong evidence that any one study influenced findings consistently across the 914 differentially methylated CpGs that passed Bonferroni correction and for which between study heterogeneity had an I2<50%. For 139/914 CpGs (15.2%) the difference in mean birth weight for a 10% greater methylation at that site varied by ≥20% with removal of a study, but the study resulting in the change was different for different CpGs. **Supplementary Figures 3.1-3.20** show the results for a random 10 plots where removal of one study changed the result by 20% or more and a random 10 where this was not the case; full results are available on request from the authors. Findings were broadly consistent when birthweight was categorised to high (>4,000g, n=1,593) versus normal (2,500-4,000g, n=6,377) (**Supplementary Table 4,** yellow track (5) in **Figure 4**) and when we did not exclude neonates born preterm or to women with pre-eclampsia or diabetes (**Supplementary Figure 4** and **Supplementary Tables 5A and 5C**, and red track (3) in **Figure 4**). Without these exclusions, we were able to examine associations with low (<2,500g, n=178) versus normal (2,500-4,000g, n=4,197) birthweight, though statistical power was still limited. Four CpGs were associated with low versus normal birthweight (Bonferroni-corrected threshold), none of which overlapped with the 914 CpGs from the main analysis (**Supplementary Table 5B,** purple track (4) in **Figure 4**). We identified that 161 of the 914 differentially methylated CpGs potentially contained a single-nucleotide polymorphism (SNP) at cytosine or guanine positions (i.e. polymorphic CpGs; **Supplementary Table 6**). Polymorphic CpGs may affect probe binding and hence measured DNA methylation levels[26,27]. We used one of the

largest studies (ALSPAC; N=633) to explore this. We found no indication of bimodal distributions for any of the 161 CpGs suggesting SNPs had not markedly affected methylation measurements at these sites (dip test p-values: 0.299 to 1.00)[28–30].

Analyses at older ages

We took the 914 neonatal blood CpGs that were associated with birthweight at Bonferroni corrected statistical significance and with I^2<50% and examined their associations with birthweight when measured in blood taken in childhood (2-13y; N = 2,756 from 10 studies), adolescence (16-18y; N = 2,906 from 6 studies) and adulthood (30-45y; N = 1,616 from 3 studies). Only participants from ALSPAC, CHAMACOS and Generation R had also contributed to the main neonatal blood EWAS. In childhood, adolescence and adulthood, we observed 87, 49 and 42 of the 914 CpGs to be nominally associated with birthweight (p<0.05). All these CpGs showed consistent directions of association. Ten CpGs showed differential methylation across all 4 age periods. However, only a minority survived Bonferroni correction for 914 tests (p<5.5x10^{-5}): 12 (1.3%), 1 (0.1%) and 0 CpGs in childhood, adolescence and adulthood, respectively (**Supplementary Table 7**; the 12 CpGs that persisted in childhood are presented in the green track (6) in **Figure 4**). Of the 914 CpGs, 50%, 52% and 49% CpGs showed consistency in direction of association in childhood, adolescence and adulthood, but these 914 CpGs were only weakly correlated with methylation levels in neonatal blood (Pearson correlation coefficients 0.15, 0.06 and 0.02, respectively for methylation level correlations with neonatal blood for blood taken in childhood, adolescence and adulthood).

Intrauterine factors

We observed enrichment of previously published maternal smoking-related CpGs in the birthweight-associated CpGs[14] (55/914 (6.0%) $p_{enrichment}$=6.12x10^{-74}, of which cg00253658 and cg26681628 also showed persistent methylation differences in the look-up in childhood). We additionally found enrichment of maternal BMI-related CpGs in the list of birthweight-related CpGs[15] (3/914 (0.3%)

342 $p_{enrichment}$=1.13x10^{-3}). All directions of association were consistent with the birthweight-lowering
343 influence of maternal smoking or the positive association of maternal BMI with birthweight
344 (**Supplementary Table 8**). We did not find evidence for overlap with plasma folate[31], and no famine
345 exposure related CpGs were previously presented at the Bonferroni-corrected level of P<1.06x10^{-7} [32].
346 In additional analyses for overlap between all FDR hits from the birthweight EWAS with those FDR
347 hits presented in the smoking, maternal BMI, folate and famine EWAS, we found an overlap of
348 430/8,170 CpGs (5.3%, $p_{enrichment}$=7.38x10^{-132}) for smoking, 584/8,170 CpGs (7.1%, $p_{enrichment}$=3.34x10^{-}
349 ^{62}) for maternal BMI and 14/8,170 (0.2%, $p_{enrichment}$=0.02) for folate. For famine we did not observe
350 overlap.

351

352 *Metastable epialleles and imprinted genes*

353 We tested the birthweight-associated CpGs for enrichment of metastable epialleles (loci for which
354 the methylation state is established in the periconceptional period[33,34]). We additionally tested for
355 enrichment of CpGs annotated to imprinted genes (loci that depend on the maintenance of
356 parental-origin-specific methylation marks in the pre-implantation embryo, some of which are
357 known to regulate fetal growth[35,36]). We did not find evidence of enrichment for metastable
358 epialleles (3/1,936 metastable epialleles overlap a birthweight-associated CpG), imprinting control
359 regions (0/741) or imprinted gene transcription start sites (5/1,728) (**Supplementary Table 9**).

360

361 *Comparison with recent GWAS findings for newborn birthweight*

362 To compare these EWAS results to those from genetic studies, we used the 60 recently published
363 fetal SNPs associated with birthweight in a GWAS meta-analysis of 153,781 newborns[37] and mapped
364 the CpG sites identified in the EWAS to these SNPs to seek evidence of co-localisation of genetic and
365 epigenetic variation (**Supplementary Table 10**). We repeated this for the 10 recently published
366 maternal SNPs associated with birthweight in a GWAS meta-analysis of 86,577 women[38]
367 (**Supplementary Table 11**). We observed that one or more of the 914 birthweight-associated CpGs

were within 4Mb (+/- 2Mb) of 34/60 fetal and all 10 maternal birthweight-associated SNPs were within 4Mb (+/- 2Mb) of one or more of the 914 birthweight-associated CpGs. Of the 34 fetal SNPs, three were located in the same gene as the CpG, as was one of the ten maternal SNPs. Ten fetal and four maternal SNPs were within 100kb of identified CpGs. In a look-up of the fetal and maternal SNPs from GWAS of birthweight in an online cord blood methylation quantitative trait loci (mQTL) database (mqtldb.org[39]), 35 fetal and 4 maternal SNPs affected methylation at some CpG(s), but none at the 914 birthweight-associated CpGs specifically.

Functional analyses

We compared the 914 birthweight-related CpGs with a recently published list of 18,881 expression quantitative trait methylation sites (cis-eQTMs, +/- 250kb around the transcription start site), CpG sites known to correlate with gene expression, from whole blood samples of 2,101 Dutch adult individuals. We found that 82 of the 914 birthweight-associated CpGs were associated with gene expression of 98 probes (cis-eQTMs)[40] ($p_{enrichment}<1.73x10^{-11}$, **Supplementary Table 12**). Additionally, in 112 Spanish four-year-olds[41], we observed that 19 CpGs were inversely associated with whole blood mRNA gene expression and 4 CpGs were positively associated with gene expression (FDR<0.05, **Supplementary Table 13**). Of these 23 CpGs, 13 were also found in the publicly available cis-eQTM list[40]. In 84 Gambian children (age 2 years)[42], we found 2 CpGs that were inversely associated with whole blood mRNA gene expression, but neither were found in the Spanish results or the publicly available cis-eQTM list. The 914 birthweight-associated CpGs showed no functional enrichment of Gene Ontology (GO) terms or Kyoto Encyclopedia of Genes and Genomes (KEGG) terms (FDR<0.05).

Mendelian randomization

We aimed to explore causality using MR analysis, in which genetic variants associated with methylation levels (methylation quantitative trait loci (mQTLs)) are used as instrumental variables to

appraise causality. For 788 (86%) of the 914 birthweight-associated CpGs, no mQTLs were identified in a publicly available mQTL database[39]. For 108 (86%) of the remaining 126 CpGs, only 1 mQTL was identified and for the remainder none had more than four mQTLs (**Supplementary Table 14** provides a complete list of all mQTLs identified for these 126 CpGs). Many of the currently available methods that can be used as sensitivity analyses to explore whether MR results are biased by horizontal pleiotropy (a single mQTL influencing multiple traits) require more than one genetic instrument (here mQTLs) and even with two or three this can be difficult to interpret[43]. Having determined that it was not possible to undertake MR analyses of 86% of the birthweight-related differentially methylated CpGs (because we did not identify any mQTLs), and for the majority of the remaining CpGs we would not have been reliably able to distinguish causality from horizontal pleiotropy (because only 1 mQTL could be identified), we decided not to pursue MR analyses further.

Discussion

This large-scale meta-analysis shows that birthweight is associated with widespread differences in DNA methylation. We observed some enrichment of birthweight-associated CpGs among sites that have previously been linked to smoking during pregnancy[14] and pre-pregnancy BMI[15], consistent with the hypothesis that epigenetic pathways may underlie the observational associations of those prenatal exposures with birthweight[17,44,45]. However, the actual overlap in this analysis was modest, likely explained by the adjustments for maternal smoking and BMI in the EWAS analyses. The overlap that we observed with pregnancy smoking related CpGs may reflect the possibility that smoking-related CpGs capture smoking better than self-report[46,47], in line with expectations of pregnant women underreporting their smoking behaviour. Adjustment for maternal smoking and BMI may have masked a greater level of overlap between our results and EWAS of these two maternal exposures. The fact that we find an association of DNA methylation across the genome with birthweight provides some support for our conceptual framework shown in **Figure 1**. However, we acknowledge that the associations that we have observed may also be explained by causal effects of maternal pregnancy exposures on both DNA methylation and fetal growth, as well as subtle inflammatory responses in cell type proportions associated with maternal smoking that might not have been completely captured with the currently available cell type estimation methods.

The differential methylation associated with birthweight in neonates persisted only minimally across childhood and into adulthood. Larger (preferably longitudinal) studies are needed to explore persistent differential methylation in more detail and with better power at older ages. It is possible that inclusion of the Gambia study in the childhood EWAS (which was the only non-European study in these analyses and was not included in the main meta-analyses with neonatal blood) might have impacted these results, although this study made up just 7% of the total child follow-up sample. A rapid attenuation of differential methylation in relation to birthweight in the first years after birth has previously been reported[20], but our sample size for these analyses may have been too small to detect persistence. This rapid decrease, if real, may indicate a reduction in

the dose of the child's exposure to maternal factors such as smoking once the offspring is delivered, with that reduction continuing as the child ages. Persistence of birthweight-related differential DNA methylation may not necessarily be a prerequisite for long-term effects, as transient differential methylation in early life may cause lasting functional alterations in organ structure and function that predispose to later adverse health effects.

Methylation is known to be associated with gene expression[48]. However, we found no consistent associations between birthweight-related methylation and gene expression in two childhood studies. This could be due to the relatively small sample sizes, differences in ethnicities, age, or platforms to measure gene expression. The use of blood, which is likely only a possible surrogate tissue for fetal growth phenotypes, for gene expression analysis might also explain the lack of findings. We did find multiple cis-eQTMs among the birthweight-related CpGs at which methylation was related to gene expression in blood when using a publicly available database from a larger adult sample[40], providing some evidence that birthweight-related differentially methylated CpGs may be associated with gene expression. These initial *in silico* association analyses need further exploration to establish any underlying causal mechanisms.

In observational studies, birthweight has repeatedly been associated with a range of later-life diseases. Changes in DNA methylation has been hypothesized as a potential mechanism linking early exposures, birthweight and later health (**Figure 1**). We originally aimed to explore this using MR analysis. For the vast majority of the birthweight-associated CpGs, no genetic instrumental variables were available. For the remaining 126 CpGs, only 1 mQTL was available, which would make it impossible to disentangle causality from horizontal pleiotropy. To ensure a strong basis for future MR analyses on this topic, there is a clear need for a more extensive mQTL resource.

Strengths of this study are its large sample size and the extensive analyses that we have undertaken. In a post-hoc power calculation based on the sample size of 8,825 with a weighted mean birthweight of 3560g (weighted mean standard deviation (SD): 483g) and with an alpha set at

456 the Bonferroni-corrected level of $P<1.06x10^{-7}$ we had 80% power, with a two-sided test, to detect a
457 minimum difference of 0.13 SD (63 grams) in birth weight for each SD increase in methylation. The
458 difference in methylation corresponding to a 1 SD increase differs per CpG, as it depends on the
459 distribution of the methylation values. We acknowledge that smaller differences which might be
460 clinically or biologically relevant may not have been identified in the current analysis. Nonetheless,
461 to our knowledge this analysis has brought together all studies currently available with relevant data
462 and is the largest published study of this association. DNA methylation patterns in neonatal blood,
463 whilst easily accessible in large numbers, may not reflect the key tissue of importance in relation to
464 birthweight. DNA methylation and gene expression in placental tissue may be important targets for
465 future studies. DNA methylation varies between leukocyte subtypes[49] and we used an adult whole
466 blood reference to correct for this in the main analyses[23,24], as the study-specific analyses were
467 completed before the widespread availability of specific cord blood reference datasets[50,51]. However,
468 we observed very similar findings in two studies (Generation R and GECKO) when we compared the
469 results with those using one of the currently available cord blood references[50]. Although we adjusted
470 for potential major confounders that may affect both methylation and fetal growth, we acknowledge
471 that the main results cannot ascertain causality. That is, whilst we have hypothesised that variation
472 in fetal DNA methylation influences fetal growth and hence birthweight, and undertaken the
473 analyses accordingly, we cannot exclude the possibility that differences in neonatal blood DNA
474 methylation are caused by variation in fetal growth itself, or that the association is confounded by
475 factors, including maternal smoking and BMI, that independently influence both fetal growth and
476 DNA methylation (as suggested in **Figure 1**). The 450k array that was used to measure genome-wide
477 DNA methylation only covers 1.7% of the total number of CpGs present in the genome and
478 specifically targets CpGs in promoter regions and gene bodies[52]. We removed the CpGs that were
479 flagged as potentially cross-reactive, as the measured methylation levels may represent methylation
480 at either of the potential loci. Also, although we did not find evidence for polymorphic effects for the
481 161 potentially polymorphic CpGs in ALSPAC, we cannot completely exclude these potential

polymorphic effects in the meta-analysed results. The majority of participants were of European ethnicity and when analyses were restricted to those of European ethnicity the results were essentially identical to those with all studies included. Direct comparisons of the main analysis with analyses in those of Hispanic or of African ethnicity for the 914 hits suggested strong correlations with Hispanic but weaker with African ethnicity. However, these results need to be treated with caution. First, we had very few studies of Hispanic and African populations. Second, we only compared the initial hits from the main meta-analysis with all ethnicities included. A detailed exploration of ethnic differences would require similar large samples for each ethnic group and within ethnic EWAS, which is beyond the scope of the data currently available.

Neonatal blood DNA methylation at many sites across the genome is associated with birthweight. Further research is required to determine if these are causal and if so whether they mediate any long-term effect of intrauterine exposures on future health.

Methods

Participants

In the main EWAS meta-analysis we explored associations of neonatal blood DNA methylation with birthweight using data from 8,825 neonates from 24 studies in the PACE Consortium[53] (**Table 1**). We removed multiple births from all analyses and excluded preterm births (<37 weeks) and offspring of mothers with pre-eclampsia or diabetes (three major pathological causes of differences in fetal growth). In follow-up analyses, we explored whether any sites found in the main analysis were discernible in relation to birthweight when examined in DNA from blood drawn during childhood (2-13y; 2,756 children from 10 studies), adolescence (16-18y; 2,906 adolescents from 6 studies) or adulthood (30-45y; 1,616 adults from 3 studies), see **Supplementary Table 1B**. Informed consent was obtained from all participants, and all studies received approval from local ethics committees. Study-specific methods and ethical approval statements are provided in **Supplementary Methods**.

Birthweight, DNA methylation and covariates

Our primary outcome was birthweight on a continuous scale (grams), adjusted for gestational age, and measured immediately after birth or retrospectively reported by mothers in questionnaires. In secondary analyses, we categorised and compared associations with high (>4,000g, n=1,593) versus normal (2,500-4,000g, n=6,377) birthweight. We also explored all associations with (continuous and categorical) birthweight in analyses that did not exclude women with preeclampsia, diabetes or preterm delivery, which also resulted in enough cases to explore low (<2,500g, n=178) versus normal (2,500-4,000g, n=4,197) birthweight (**Supplementary Table 1C** shows the characteristics of participants). Primary, secondary and follow-up analyses are outlined in the study design in **Figure 2**. DNA methylation was measured in neonatal blood samples using the Illumina Infinium® HumanMethylation450 BeadChip assay. All participants had cord blood samples except for three studies with heel stick blood spots (n=1,254 [14.2%]). After study-specific laboratory analyses, quality control, normalisation, and removal of control probes (N=65) and probes that mapped to the

X (N=11,232) and Y (n=370) chromosomes, we included 473,864 CpGs. DNA methylation is expressed as the proportion of cells in which the DNA was methylated at a specific site and hence takes values from zero to one. We converted this to a percentage and present differences in mean birthweight per 10% higher DNA methylation level at each CpG. All analyses were adjusted for gestational age at delivery, child sex, maternal age at delivery, parity (0/≥1), smoking during pregnancy (no smoking / stopped in early pregnancy / smoking throughout pregnancy), pre-pregnancy BMI, socio-economic position, technical variation, and estimated white blood cell proportions (B-cells, CD8+ T-cells, CD4+ T-cells, granulocytes, NK-cells and monocytes)[23–25]. In studies with participants from multiple ethnic groups, each group was analysed separately and results were added to the meta-analyses as separate studies. Further details are provided in the study-specific **Supplementary Methods**.

Statistical methods

Robust linear (birthweight as a continuous outcome) or logit (binary birthweight outcomes) regression EWAS were undertaken within each study according to a pre-specified analysis plan. Quality control, normalisation and regression analyses were conducted independently by each study. After confirming comparability of study-specific summary statistics[54], we combined results using a fixed effects inverse variance weighted meta-analysis[55]. The meta-analysis was done independently by two study groups and the results were compared in order to minimise the likelihood of human error. We show (two-sided) results after correcting for multiple testing using both the FDR<0.05[56] and the Bonferroni correction ($p<1.06\times10^{-7}$). We completed follow-up analyses for differentially methylated CpGs that reached the Bonferroni-adjusted threshold and did not show large between-study heterogeneity[57] ($I^2\leq50\%$). We annotated the nearest gene for each CpG using the UCSC Genome Browser build hg19[58,59]. We explored whether between-study heterogeneity might be explained by differences in ethnicity between studies, by repeating the meta-analysis including only participants of European ethnicity, which was by far the largest ethnic subgroup (n=6,023 from 17 studies) (**Figure 2**). Ethnicity was defined using maternal or self-report, unless

specified otherwise in study-specific **Supplementary Methods**. We also did meta-analyses only including the Hispanic studies and only including the African American studies and present those results for illustrative purposes only, given the much smaller sample size. All analyses were performed using R[60], except for the meta-analysis which was performed using METAL[55]. We removed CpGs that co-hybridised to alternate sequences (i.e. cross-reactive sites), because we cannot distinguish whether the differential methylation is at the locus that we have reported or at the one that the probe cross-reacts with. We compared the birthweight-related CpGs to lists of CpGs that are potentially influenced by a SNP (polymorphic sites)[26,27]. For these CpGs, we determined if DNA methylation levels were influenced by nearby SNPs, by assessing whether their distributions deviated from unimodality using Hartigans' dip test[28,29] and visual inspection of density plots in n=742 cord blood samples in the ALSPAC study.

Analyses at older ages

Analyses of the associations with DNA methylation in blood collected in childhood, adolescence and adulthood followed the same covariable adjustment and methods as for the main analyses ($p<5.5*10^{-5}$ for 914 tests). All participants and studies in these analyses at older ages had not been included in the main meta-analysis in neonatal blood, except for ALSPAC (N=633 in neonatal analyses, N=605 in childhood and N=526 in adolescence), CHAMACOS (N= 283 in neonatal analyses and N=191 in childhood) and Generation R (N=717 in neonatal analyses and N=372 in childhood). Characteristics are shown in study-specific **Supplementary Methods** and **Supplementary Table 1B**.

Intrauterine factors

We used a hypergeometric test to explore the extent to which any of the birthweight-related CpGs overlapped with those previously associated with intrauterine exposure to smoking[14] (n=568 CpGs), BMI[15] (n=104 CpGs) and plasma folate[31] (n=48 CpGs), using the same (Bonferroni-corrected) cut-off for statistical significance. No CpGs reached the Bonferroni-corrected cut-off for famine[32]. We

additionally appraised this overlap using the FDR<0.05 cut-off for all traits (n=8,696 birthweight-related CpGs, n=6,703 smoking-related CpGs, n=16,067 BMI-related CpGs, n=443 folate-related CpGs, n=7 famine-related CpGs). These FDR results were available from the publications for smoking, folate and famine, and we obtained them from the corresponding author for BMI.

Metastable epialleles and imprinted genes

We tested the birthweight-associated CpGs for enrichment of metastable epialleles and CpGs associated with imprinted genes. The metastable epialleles were derived from a recently published study that identified 1,936 putative metastable epialleles[34]. For imprinted genes, we first identified a set of CpGs falling within a curated set of imprinting control regions; differentially methylated regions controlling the parental-specific expression of one or more imprinted genes[36]. Second, we extracted the set of imprinting control region controlled genes from the above source and identified all 450k CpGs within +/- 10kbp of the gene transcription start site, including all known alternative TSS identified in grch37.ensembl.org using biomaRt[61,62].

Comparison with recent GWAS findings for newborn birthweight

We compared the birthweight-associated CpGs with the 60 SNPs from the most recent GWAS meta-analyses of fetal genotype associations with birthweight in >150,000 newborns[37] and with 10 SNPs from the most recent GWAS meta-analysis of maternal genotype associations with birthweight in >86,000 women[38]. With this comparison we checked if the EWAS top hits were located within a 4Mb window (+/- 2Mb) surrounding these SNPs. We additionally checked whether SNPs and CpGs were located in the same gene.

Functional analyses

To explore the association of methylation with gene expression, we compared birthweight-related CpGs with a recently published list of 18,881 cis-eQTMs from whole blood samples of 2,101 Dutch adult individuals[40]. With a hypergeometric test, we calculated enrichment of cis-eQTMs in the list of

birthweight-associated CpGs. We further explored methylation of birthweight-associated CpGs in relation to whole blood mRNA gene expression (transcript levels) within a 500kb region of the CpGs (+/- 250kb, FDR<0.05) in 112 Spanish four-year-olds[41] and 84 Gambian two-year-olds[42] (**Supplementary Methods**). To better understand the potential mechanisms linking DNA methylation and birthweight, we explored the potential functions of the birthweight-associated CpGs using GO and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses. We used the missMethyl R package[63], which enabled us to correct for the number of probes per gene on the 450k array, based on the November 2018 version of the GO and KEGG source databases. To filter out the large, general pathways we set the number of genes for each gene set between 15 and 1000, respectively. We calculated FDR at 5% corrected P-values for enrichment.

Mendelian randomization

MR uses genetic variants as instrumental variables to study the causal effect of exposures on outcomes[64,65]. We aimed to use two-sample MR[22,66] to explore (a) evidence of a causal association of methylation levels at the identified CpGs with birthweight and (b) evidence of a causal association of these CpGs with later-life health outcomes (i.e. to explore our hypothesised causal mechanisms shown in **Figure 1**). We did this by first searching a publicly available mQTL database[39] to identify cis-mQTLs within 1Mb of each of the Bonferroni-corrected, with I^2<50%, birthweight-related differentially methylated CpGs. These mQTLs could then be used as genetic instrumental variables for methylation levels of the birthweight-related CpGs. We then aimed to determine the association of these mQTLs with birthweight and later-life health outcomes from publicly available summary GWAS results[66].

Acknowledgements and competing financial interests

All authors declared no conflict of interest. For all studies, acknowledgements and funding statements can be found in **Supplementary Acknowledgements** and **Supplementary Funding**.

Author contributions

LKK, DAL, CLR, HS and JFF conceived and designed the study. Study-specific analyses were completed by GCS (ALSPAC and GOYA), SKM (BAMSE), RR (CBC), PY (CHAMACOS), CVB (CHS), KMB (EARLI), AG and AN (EXPOsOMICS, The Gambia and MoBa3), SASL (FLEHS1) , LKK (GECKO), CA (Gen3G), CM (Generation R), JL (Glaku), APS (Healthy Start), LAS (INMA), FIR (IOW F1) , JWH (IOW F2) , DAVDP (Lifelines), CMP (MoBa1), SER (MoBa2), AJW (NCL), DDJ (NEST), MW (NFBC66 and NFBC86), TME (NHBCS and RICHS), JVD (NTR), CJX (PIAMA), DC (PREDO), ACJ (PRISM), ACJ (PROGRESS), SL (Project Viva), RCH (Raine), VU (STOPPA). LKK and CM meta-analysed the results. LKK, CM, GCS, PY, LAS, AG, AN and MJS performed follow-up analyses. LKK, DAL, CLR, HS and JFF interpreted the results. LKK, with input from DAL, CLR, HS and JFF, wrote the first draft of the manuscript, with all authors reading and suggesting critical revisions on subsequent drafts. Correspondence and material requests can be addressed to JFF, HS, DAL and CLR (j.felix@erasmusmc.nl, h.snieder@umcg.nl, d.a.lawlor@bristol.ac.uk, caroline.relton@bristol.ac.uk).

Data availability

All summary statistics from this EWAS meta-analysis are available via doi: 10.5281/zenodo.2222287.

Code availability

The code used for this EWAS meta-analysis is available from the authors upon request.

Figures

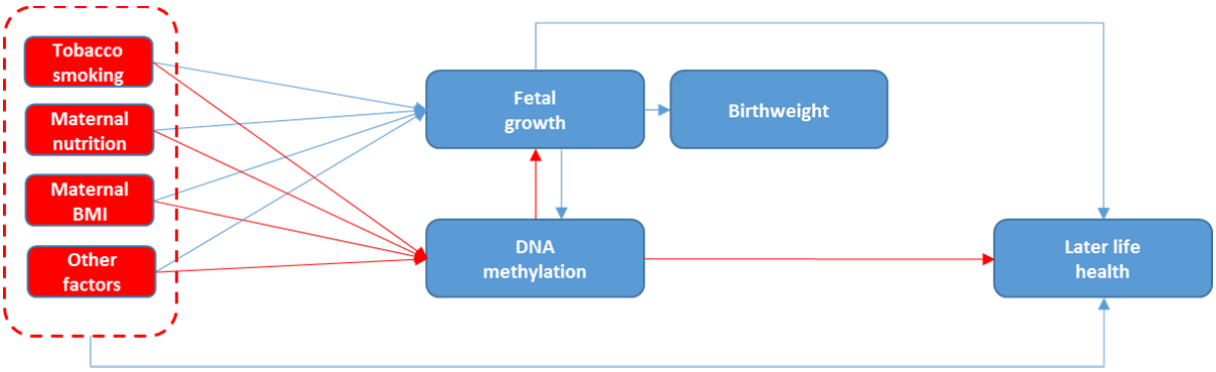

Figure 1. Hypothetical paths that might link intrauterine exposures to DNA methylation, birthweight and later-life health outcomes. Red arrows summarise the paths that have motivated the analyses undertaken in this study (i.e. that maternal environmental exposures influence DNA methylation that in turn influences fetal growth and hence birthweight). The EWAS meta-analysis undertaken sought to identify methylation associated with birthweight. Blue arrows summarise other plausible paths, including that maternal exposures influence fetal growth first and it then influences DNA methylation or that maternal exposures may influence fetal growth/birthweight and later-life health outcomes through other pathways than DNA methylation.

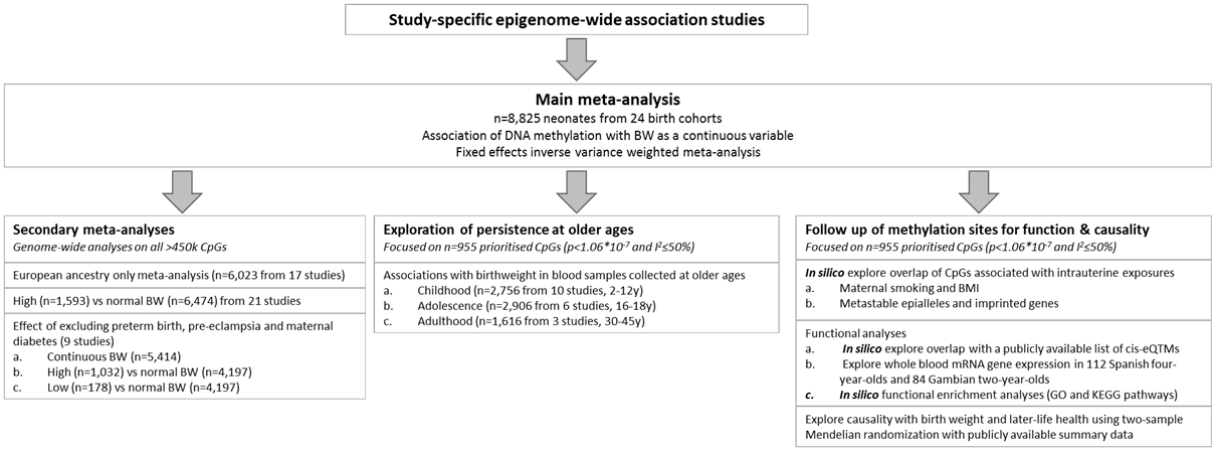

Figure 2. Study design specifying the main meta-analysis, secondary meta-analyses, follow-up analyses and exploration of persistence at older ages.

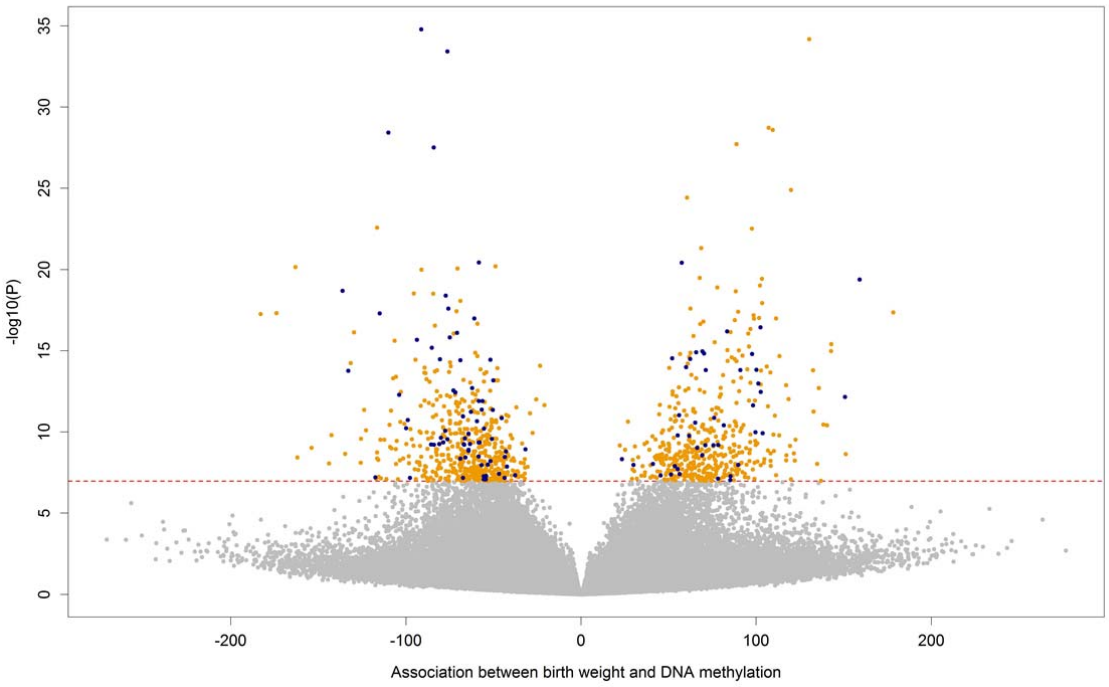

Figure 3. Volcano plot showing the direction of associations between DNA methylation and birthweight as a continuous measure in n=8,825 neonates from fixed effects inverse variance weighted meta-analysis of 24 birth cohorts. The X-axis represents the difference in birthweight in grams per 10% methylation difference, the Y-axis represents the $-\log_{10}(P)$. The red line shows the Bonferroni-corrected significance threshold for multiple testing ($p<1.06\times10^{-7}$). Highlighted in orange are the 914 CpGs with $p<1.06\times10^{-7}$ and $I^2\leq50\%$ and highlighted in blue are the 115 CpGs with $p<1.06\times10^{-7}$ and $I^2>50\%$.

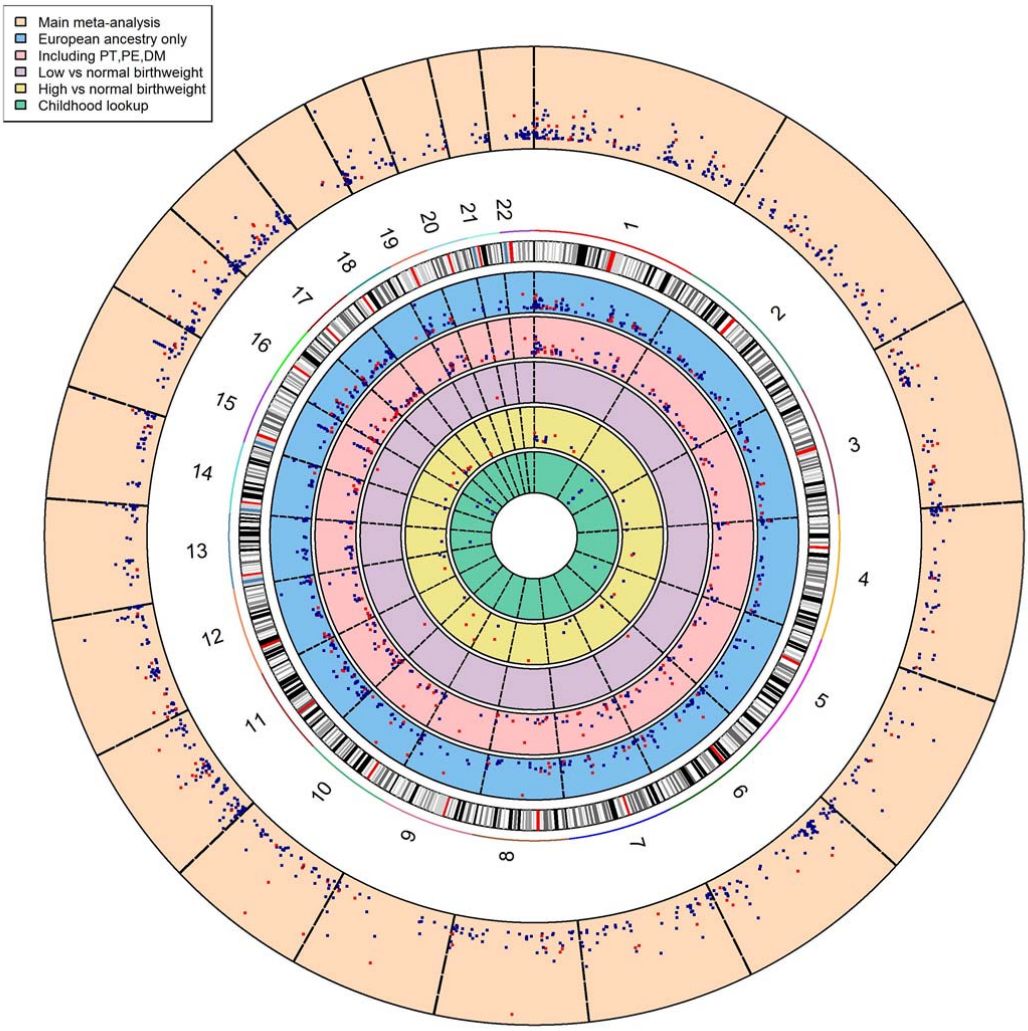

Figure 4. **Circos plot showing the (Bonferroni-corrected $p<1.06x10^{-7}$) meta-analysis results for associations between DNA methylation and birthweight.** Results are presented as CpG-specific associations ($-\log_{10}(P)$, each dot represents a CpG) by genomic position, per chromosome. From outer to inner track: [1, orange] Main analysis results for associations between DNA methylation and birthweight as a continuous measure (N=8,825), [2, blue] Results from participants from European ethnicity only, DNA methylation and birthweight as a continuous measure (N=6,023), [3, red] Results from analysis without exclusion for preterm births, pre-eclampsia and maternal diabetes, DNA methylation and birthweight as a continuous measure N=5,414), [4, purple] Results from logistic regression analysis without exclusion for preterm births, pre-eclampsia and maternal diabetes, for low (N=178) vs normal (N=4,197) birthweight, [5, yellow] Results from logistic regression analysis for associations between DNA methylation and high (N=1,590) vs normal (N=6,114) birthweight, [6, green] Results from look-up analysis in methylation samples taken during childhood and its association with birthweight as a continuous measure (N=2,756). Track 1: highlighted in red are 115 CpGs with I^2>50%. Tracks 2-6: highlighted in red are CpGs that were not found in the 914 main meta-analysis hits (though note differences in sample size and hence statistical power for different analyses presented in the different tracks).

Table 1. Characteristics for the participating studies in the main meta-analysis for the association between neonatal blood DNA methylation and birthweight

Study	N total	N normal birthweight (%)	N high birthweight (%)	Birthweight (g)	Gestational age (wk)	Ethnicity	N Boys (%)
ALSPAC	633	547 (86.4)	79 (12.5)	3512 ± 443	39.7 ± 1.3	European	301 (47.6)
CBC[a]: Hispanic	127	106 (83.5)	19 (15.0)	3445 ± 484	39.8 ± 1.3	Hispanic	74 (58.3)
CBC[a]: Caucasian	136	108 (79.4)	26 (19.1)	3625 ± 472	39.7 ± 1.5	European	79 (58.1)
CHAMACOS	283	236 (83.4)	44 (15.5)	3520 ± 446	39.3 ± 1.2	Hispanic	142 (50.1)
CHS[a]	199	168 (84.4)	28 (14.1)	3486 ± 476	40.2 ± 1.2	Mixed	79 (39.7)
EARLI	131	113 (86.3)	16 (12.2)	3507 ± 480	39.3 ± 1.0	Mixed	70 (53.4)
EXPOsOMICS: Rhea, Environage and Piccolipiu	324	297 (91.7)	22 (6.8)	3368 ± 437	39.4 ± 1.2	European	169 (52.1)
GECKO	255	206 (80.8)	46 (18.0)	3543 ± 533	39.7 ± 1.3	European	136 (53.3)
Gen3G	162	145 (89.5)	15 (9.3)	3408 ± 431	39.5 ± 1.1	European	74 (45.7)
Generation R	717	589 (82.1)	122 (17.0)	3572 ± 465	40.2 ± 1.1	European	372 (51.9)
GOYA[b]	947	649 (68.5)	294 (31.0)	3750 ± 501	40.4 ± 1.3	European	483 (51.0)
Healthy Start: African American	77	-	-	3059 ± 358	38.9 ± 1.3	African American	42 (54.5)
Healthy Start: Hispanic	115	-	-	3322 ± 395	39.1 ± 1.1	Hispanic	55 (47.8)
Healthy Start: Caucasian	240	220 (91.7)	14 (5.8)	3325 ± 425	39.3 ± 1.1	European	125 (52.1)
INMA	166	-	-	3297 ± 400	39.9 ± 1.2	European	82 (49.4)
IOW F2	118	97 (82.2)	17 (14.4)	3432 ± 525	39.7 ± 1.6	European	59 (50.0)
MoBa1	1066	795 (74.6)	251 (23.5)	3644 ± 544	39.5 ± 1.6	European	568 (53.3)
MoBa2	587	435 (74.1)	146 (24.9)	3701 ± 487	40.1 ± 1.2	European	329 (56.0)
MoBa3	205	153 (74.6)	51 (24.9)	3706 ± 491	39.8 ± 1.2	European	106 (51.7)
NCL[a]	792	592 (74.7)	192 (24.2)	3671 ± 506	40.0 ± 1.3	European	453 (57.2)
NEST: African American	99	-	-	3197 ± 534	39.3 ± 1.2	African American	47 (47.5)
NEST: Caucasian	111	94 (84.7)	13 (11.7)	3446 ± 471	39.5 ± 1.2	European	50 (45.0)
NHBCS	96	84 (87.5)	12 (12.5)	3509 ± 453	39.6 ± 1.1	European	53 (55.2)
PREDO	540	428 (79.3)	99 (18.3)	3572 ± 478	40.1 ± 1.2	European	264 (48.8)
PRISM	138	-	-	3385 ± 441	39.5 ± 1.1	Mixed	76 (55.1)
PROGRESS	143	-	-	3124 ± 387	38.6 ± 1.1	Hispanic	77 (53.8)
RICHS	89	52 (58.4)	23 (25.8)	3335 ± 734	38.9 ± 1.2	European	35 (39.3)
Project Viva	329	263 (79.9)	64 (19.5)	3623 ± 473	40.0 ± 1.2	European	168 (51.1)
N total	*8825*	*6377*	*1593*				

Results are presented as mean ± SD or N (%). Normal birthweight: 2500 - 4000 g, high birthweight: > 4000 g, low birthweight: <2500 g. Studies with mixed ethnicities analysed all participants together with adjustment for ethnicities. g: grams, wk: weeks, y: years. Full study names can be found in study-specific Supplementary Methods. For some studies the sample size for defining normal/high BW was too small.

a) CBC, CHS and NCL used heel prick blood spot samples instead of cord blood.

b) GOYA is a case-cohort study (cases are mothers with BMI>32 and controls are mothers randomly sampled from the underlying study population in which the cases were identified), in analyses where we included a random sample with a normal BMI distribution results were essentially the same as in the main analyses.

References

1. Tyrrell, J. *et al.* Genetic Evidence for Causal Relationships Between Maternal Obesity-Related Traits and Birth Weight. *JAMA* **315,** 1129 (2016).
2. Tyrrell, J. *et al.* Genetic variation in the 15q25 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) interacts with maternal self-reported smoking status during pregnancy to influence birth weight. *Hum. Mol. Genet.* **21,** 5344–58 (2012).
3. Bakker, R., Steegers, E. A. P., Hofman, A. & Jaddoe, V. W. V. Blood pressure in different gestational trimesters, fetal growth, and the risk of adverse birth outcomes: the generation R study. *Am. J. Epidemiol.* **174,** 797–806 (2011).
4. Lawlor, D. A. *et al.* Association of existing diabetes, gestational diabetes and glycosuria in pregnancy with macrosomia and offspring body mass index, waist and fat mass in later childhood: Findings from a prospective pregnancy cohort. *Diabetologia* **53,** 89–97 (2010).
5. van Uitert, E. M. & Steegers-Theunissen, R. P. M. Influence of maternal folate status on human fetal growth parameters. *Mol. Nutr. Food Res.* **57,** 582–595 (2013).
6. Painter, R. C., Roseboom, T. J. & Bleker, O. P. Prenatal exposure to the Dutch famine and disease in later life: An overview. *Reproductive Toxicology* **20,** 345–352 (2005).
7. Whincup, P. H. *et al.* Birth Weight and Risk of Type 2 Diabetes A Systematic Review. *JAMA* **300,** 2886–2897 (2008).
8. Lawlor, D. A., Ronalds, G., Clark, H., Davey Smith, G. & Leon, D. A. Birth weight is inversely associated with incident coronary heart disease and stroke among individuals born in the 1950s: Findings from the Aberdeen children of the 1950s prospective cohort study. *Circulation* **112,** 1414–1418 (2005).
9. O'Donnell, K. J. & Meaney, M. J. Fetal Origins of Mental Health: The Developmental Origins of Health and Disease Hypothesis. *Am. J. Psychiatry* **174,** 319–328 (2017).
10. McCormack, V. A., Silva, I. D. S., Koupil, I., Leon, D. A. & Lithell, H. O. Birth characteristics and adult cancer incidence: Swedish cohort of over 11,000 men and women. *Int. J. Cancer* **115,** 611–617 (2005).
11. Risnes, K. R. *et al.* Birthweight and mortality in adulthood: A systematic review and meta-analysis. *Int. J. Epidemiol.* **40,** 647–661 (2011).
12. Freathy, R. M. Can genetic evidence help us to understand the fetal origins of type 2 diabetes? *Diabetologia* **59,** 1850–1854 (2016).

13. Hanson, M. Birth Weight and the Fetal Origins of Adult Disease. *Pediatr. Res.* **52,** 473–474 (2002).

14. 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).

15. 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. *Hum. Mol. Genet.* **26,** 4067–4085 (2017).

16. Gruzieva, O. *et al.* Epigenome-Wide Meta-Analysis of Methylation in Children Related to Prenatal NO2 Air Pollution Exposure. *Environ. Health Perspect.* **125,** 104–110 (2017).

17. Lawlor, D. A., Relton, C., Sattar, N. & Nelson, S. M. Maternal adiposity—a determinant of perinatal and offspring outcomes? *Nat. Rev. Endocrinol.* **8,** 679–688 (2012).

18. Adkins, R. M., Tylavsky, F. A. & Krushkal, J. Newborn umbilical cord blood DNA methylation and gene expression levels exhibit limited association with birth weight. *Chem Biodivers* **9,** 888–899 (2012).

19. Engel, S. M. *et al.* Neonatal genome-wide methylation patterns in relation to birth weight in the Norwegian Mother and Child Cohort. *Am J Epidemiol* **179,** 834–842 (2014).

20. Simpkin, A. J. *et al.* Longitudinal analysis of DNA methylation associated with birth weight and gestational age. *Hum. Mol. Genet.* **24,** 3752–3763 (2015).

21. Agha, G. *et al.* Birth weight-for-gestational age is associated with DNA methylation at birth and in childhood. *Clin. Epigenetics* **8,** 1–12 (2016).

22. Lawlor, D. A. Commentary: Two-sample Mendelian randomization: opportunities and challenges. *Int. J. Epidemiol.* **45,** 908–915 (2016).

23. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13,** 1–16 (2012).

24. 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).

25. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30,** 1363–1369 (2014).

26. Chen, Y. A. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics* **8,** 203–209 (2013).

27. Naeem, H. *et al.* Reducing the risk of false discovery enabling identification of biologically significant

genome-wide methylation status using the HumanMethylation450 array. *BMC Genomics* **15,** 51 (2014).

28. Hartigan, J. & Hartigan, P. The Dip Test of Unimodality. *Ann. Stat.* **13,** 70–84 (1985).

29. Maechler, M. Hartigan's Dip Test Statistic for Unimodality - Corrected. R package. (2015).

30. Relton, C. L. *et al.* Data resource profile: Accessible resource for integrated epigenomic studies (ARIES). *Int. J. Epidemiol.* **44,** 1181–1190 (2015).

31. Joubert, B. R. *et al.* Maternal plasma folate impacts differential DNA methylation in an epigenome-wide meta-analysis of newborns. *Nat. Commun.* **7,** 10577 (2016).

32. Tobi, E. W. *et al.* Early gestation as the critical time-window for changes in the prenatal environment to affect the adult human blood methylome. *Int. J. Epidemiol.* **44,** 1211–1223 (2015).

33. Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I. & Whitelaw, E. Metastable epialleles in mammals. *Trends Genet.* **18,** 348–51 (2002).

34. Van Baak, T. E. *et al.* Epigenetic supersimilarity of monozygotic twin pairs. *Genome Biol.* **19,** 2 (2018).

35. Moore, G. E. *et al.* The role and interaction of imprinted genes in human fetal growth. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370,** 20140074 (2015).

36. Monk, D. *et al.* Recommendations for a nomenclature system for reporting methylation aberrations in imprinted domains. *Epigenetics* 1–5 (2016). doi:10.1080/15592294.2016.1264561

37. Horikoshi, M. *et al.* Genome-wide associations for birth weight and correlations with adult disease. *Nature* **4,** 1–20 (2016).

38. Beaumont, R. N. *et al.* Genome-wide association study of offspring birth weight in 86 577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Hum. Mol. Genet.* **27,** 742–756 (2018).

39. Gaunt, T. R. *et al.* Systematic identification of genetic influences on methylation across the human life course. *Genome Biol.* **17,** 1–14 (2016).

40. Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat. Genet.* **49,** 131–138 (2016).

41. Guxens, M. *et al.* Cohort profile: The INMA-INfancia y Medio Ambiente-(environment and childhood) project. *Int. J. Epidemiol.* **41,** 930–940 (2012).

42. Moore, S. E. *et al.* A randomized trial to investigate the effects of pre-natal and infant nutritional supplementation on infant immune development in rural Gambia: the ENID trial: Early Nutrition and

Immune Development. *BMC Pregnancy Childbirth* **12,** 1–8 (2012).

43. Hemani, G., Bowden, J. & Davey Smith, G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum. Mol. Genet.* **27,** R195–R208 (2018).

44. Richmond, R. C., Timpson, N. J. & Sørensen, T. I. A. Exploring possible epigenetic mediation of early-life environmental exposures on adiposity and obesity development. *Int. J. Epidemiol.* **44,** 1191–1198 (2015).

45. Küpers, L. K. *et al.* DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int. J. Epidemiol.* **44,** 1224–1237 (2015).

46. Valeri, L. *et al.* Misclassified exposure in epigenetic mediation analyses. Does DNA methylation mediate effects of smoking on birthweight? *Epigenomics* **9,** 253–265 (2017).

47. Reese, S. E. *et al.* DNA Methylation Score as a Biomarker in Newborns for Sustained Maternal Smoking during Pregnancy. *Env. Heal. Perspect* (2016). doi:10.1289/EHP333

48. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl,** 245–254 (2003).

49. Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol.* **15,** R31 (2014).

50. Bakulski, K. M. *et al.* DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics* **11,** 354–362 (2016).

51. Gervin, K. *et al.* Cell type specific DNA methylation in cord blood: A 450K-reference data set and cell count-based validation of estimated cell type composition. *Epigenetics* **11,** 690–698 (2016).

52. Dedeurwaerder, S. *et al.* Evaluation of the Infinium Methylation 450K technology. *Epigenomics* (2011). doi:10.2217/epi.11.105

53. Felix, J. F. *et al.* Cohort Profile: Pregnancy And Childhood Epigenetics (PACE) Consortium. *Int. J. Epidemiol.* **16,** 10–14 (2017).

54. van der Most, P. J., Küpers, L. K., Snieder, H. & Nolte, I. QCEWAS: automated quality control of results of epigenome-wide association studies. *Bioinformatics* **33,** 1243–1245 (2017).

55. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).

56. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to

multiple testing. *J R Stat. Soc B* **57,** 289–300 (1995).

57. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21,** 1539–1558 (2002).

58. Triche, T. FDb.InfiniumMethylation.hg19: Annotation package for Illumina Infinium DNA methylation probes. R package version 2.2.0. (2014).

59. Carlson, M. & Maintainer, B. TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s). R package version 3.2.2. (2015).

60. R Core Team & Team, R. D. C. R: A language and environment for statistical computing.

61. Durinck, S. *et al.* BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **21,** 3439–3440 (2005).

62. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/ Bioconductor package biomaRt. *Nat. Protoc.* **4,** 1184–1191 (2009).

63. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* **32,** 286–288 (2016).

64. Davey Smith, G. & Ebrahim, S. 'Mendelian randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* **32,** 1–22 (2003).

65. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27,** 1133–63 (2008).

66. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7,** e34408 (2018).