


RESEARCH ARTICLE

Identification of methylation changes associated with positive and negative growth deviance in Gambian infants using a targeted methyl sequencing approach of genomic DNA

Claire R. Quilter¹ | Kerry M. Harvey¹ | Julien Bauer¹ | Benjamin M. Skinner^{1,2} |
 Maria Gomez¹ | Manu Shrivastava¹ | Andrew M. Doel^{3,4} | Saikou Drammeh⁴ |
 David B. Dunger⁶ | Sophie E. Moore^{3,4} | Ken K. Ong^{6,7,8} | Andrew M. Prentice⁴ |
 Robin M. Bernstein^{5,9}  | Carole A. Sargent¹ | Nabeel A. Affara¹

¹Department of Pathology, University of Cambridge, Cambridge, UK

²School of Life Sciences, University of Essex, Colchester, UK

³Department of Women and Children's Health, King's College London, London, UK

⁴MRC Unit The Gambia at London School of Hygiene and Tropical Medicine, Banjul, The Gambia

⁵Growth and Development Lab, Department of Anthropology, University of Colorado, Boulder, CO, USA

⁶MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK

⁷Department of Paediatrics, University of Cambridge School of Clinical Medicine, Cambridge, UK

⁸Institute of Metabolic Science, Cambridge Biomedical Campus Cambridge, Cambridge, UK

⁹Institute of Behavioural Science, University of Colorado, Boulder, CO, USA

Abstract

Low birthweight and reduced height gain during infancy (stunting) may arise at least in part from adverse early life environments that trigger epigenetic reprogramming that may favor survival. We examined differential DNA methylation patterns using targeted methyl sequencing of regions regulating gene activity in groups of rural Gambian infants: (a) low and high birthweight (DNA from cord blood ($n = 16$ and $n = 20$, respectively), from placental trophoblast tissue ($n = 21$ and $n = 20$, respectively), and DNA from peripheral blood collected from infants at 12 months of age ($n = 23$ and $n = 17$, respectively)), and, (b) the top 10% showing rapid postnatal length gain (high, $n = 20$) and the bottom 10% showing slow postnatal length gain (low, $n = 20$) based on z score change between birth and 12 months of age (LAZ) (DNA from peripheral blood collected from infants at 12 months of age). Using BiSeq analysis to identify significant methylation marks, for birthweight, four differentially methylated regions (DMRs) were identified in trophoblast DNA, compared to 68 DMRs in cord blood DNA, and 54 DMRs in 12-month peripheral blood DNA. Twenty-five DMRs were observed to be associated with high and low length for age (LAZ) at 12 months. With the exception of five loci (associated with two different genes), there was no overlap between these groups of methylation marks. Of the 194 CpG methylation marks contained within DMRs, 106 were located to defined gene

Abbreviations: Cis-eQTM, Cis-acting quantitative trait methylation; Cis-meQTL, Cis-acting methylation quantitative trait locus; CTCF, CCCTC-binding factor; DAVID, Database for Annotation, Visualization, and Integrated Discovery; DMR, Differentially Methylated Region; EBI, European Bioinformatics Institute; EWAS, Epigenome-Wide Association Study; FDR, False Discovery Rate; GAD, Genetic Association Database; GWAS, Genome-Wide Association Study; LAZ, Length for Age Z score; MAF, Minor Allele Frequency; MRC, Medical Research Council; NCBI, National Center for Bioinformatics Technology; OMIM, Online Mendelian Inheritance in Man; PCA, Principal Components Analysis; SGA, Small for Gestation Age; SNP, Single Nucleotide Polymorphism; Trans-meQTL, Trans-acting methylation quantitative trait locus; TSS, Transcription Start Site; UGR, Uterine Growth Restriction.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *FASEB BioAdvances* published by the Federation of American Societies for Experimental Biology

Correspondence

Nabeel A. Affara, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK.

Email: na106@cam.ac.uk

Present address

Claire R. Quilter, East Midlands & East of England NHS Genomic Laboratory Hub, Genomics Laboratories, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

Maria Gomez, Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK

Manu Shrivastava, Oxford University Hospitals, Oxford, UK

Funding information

This work was funded by the Bill and Melinda Gates Foundation (OPP1066932) and by core funding to the MRC Unit The Gambia at LSHTM (MC-A760-5QX00) by the UK MRC and the UK Department for the International Development (DFID) under the MRC/DFID Concordat agreement.

regulatory elements (promoters, CTCF-binding sites, transcription factor-binding sites, and enhancers), 58 to gene bodies (introns or exons), and 30 to intergenic DNA. Distinct methylation patterns associated with birthweight between comparison groups were observed in DNA collected at birth (at the end of intrauterine growth window) compared to those established by 12 months (near the infancy/childhood growth transition). The longitudinal differences in methylation patterns may arise from methylation adjustments, changes in cellular composition of blood or both that continue during the critical postnatal growth period, and in response to early nutritional and infectious environmental exposures with impacts on growth and longer-term health outcomes.

KEY WORDS

birthweight, DNA methylation, environmental exposures, stunting

1 | INTRODUCTION

About 45% of global deaths in children under 5 years of age are thought to be related to undernutrition.¹ Children who survive early periods of undernutrition may suffer longer-term consequences, including stunting and other developmental deficits,² which are major contributors to long-term morbidity and mortality.^{3,4} Although the prevalence of stunting declined in sub-Saharan Africa from 42% in 1990 to 32% in 2015, the numbers of affected individuals increased from 47 million to 58 million.⁵ Studies estimate that 20% of growth retardation starts *in utero* where under-nutrition in pregnancy increases the risks of intrauterine growth retardation (IUGR) and small for gestation age (SGA) infants, preterm delivery⁶ and long-term impaired immunity. It is hypothesized that an adverse early life environment and nutrition induce phenotypic adaptations through developmental plasticity⁷ to favor survival in the short term, but at the expense of lifelong effects on health.^{8,9}

Nutritional interventions to improve child growth and adult health^{10,11} have had limited success, primarily for the lack of a clear understanding of optimal timing, target groups, and the composition of supplements. The period of growth and development from conception to a child's second birthday (coined the first 1000 days) is one of the most critical windows of opportunity for interventions.¹² There is a complex interplay between an individual's genetic constitution and the environment. Responses to extrinsic factors via

modifications to the epigenome (which may include to both chromatin-associated proteins and DNA bases) in the first 1000 days are believed to be important in establishing protective adaptations against the impact of under-nutrition and an adverse environment (thrifty phenotype).^{13,14} DNA methylation at CpG couplets is one of the most actively studied modifications to the epigenome.

A large meta-analysis of multiple epigenome-wide association studies (EWAS) by the Childhood Epigenetics Consortium found methylation at 914 CpG sites associated with birthweight in whole blood DNA from healthy neonates, but <1.3% persisted in children (2–13 years), <0.1% in adolescents (16–18 years), and none in adults (30–45 years).¹⁵ The current study uses samples and data from a cohort of Gambian mother–infant pairs exhibiting high rates of maternal and child under-nutrition. Rural Gambian infants are small at birth relative to international standards, show positive growth patterns during the first few months of life and then, enter a period of reduced growth marked by profound faltering until at least 24 months of age.^{16,17} Schoenbuchner et al.¹⁶ have suggested that stunting is an extreme adaptation to profound faltering episodes potentially arising from a complex interaction of malnutrition, infection, and disease. Despite four decades of nutrition-sensitive and nutrition-specific interventions halving under-nutrition for young children from rural Gambia, substantial (30%) growth faltering remains,¹⁷ indicating a gap in our understanding of its complex etiology.

Epigenetic studies carried out on Gambian populations have highlighted the importance of maternal nutrition and

exposures and the effects of maternal nutritional supplementation in this highly seasonal environment. Many aspects of health and behavior in rural Gambia are influenced by the annual seasonality with a single rainy “hungry” season (late June–October) followed by a dry “harvest” season (November–May/June).^{2,18} Specifically, there is evidence that seasonal variation in nutrition during the periconceptional period influences methylation status in postnatal infants at a number of loci,¹⁹ is related to methyl-donor nutrient content of the mother's diet^{20–23} and may be associated with an increase in both preterm and SGA infants.¹⁸ Periconceptional nutrition supplementation influences methylation changes in cord and postnatal infant blood DNA at CpG loci linked to genes associated with infection and immunity²⁴ and alters the methylation at imprinted loci.²⁵ Maternal exposure to aflatoxin B1 is also associated with DNA methylation changes at specific loci in Gambian infants.²⁶

The aim of the present study was to identify epigenetic marks that are established during the critical first 1000 days in a cohort of rural Gambian infants and explore how these may be associated with normal versus stunted growth outcomes in order to determine whether any targets for intervention are associated with prenatal and/or postnatal periods of epigenetic modification. We used a targeted methyl sequencing approach of genomic DNA from placental trophoblast tissue, cord, and infant (12 months of age) blood to identify the methylation changes. These changes may be useful as biomarkers, highlighting genes influenced by exposures during embryonic and fetal development and early infancy, and identifying potential pathways through which these may influence the growth outcomes at birth and in the first year of life.

2 | MATERIALS AND METHODS

2.1 | Samples

The study was conducted among pregnant women and their infants living in the rural West Kiang region of The Gambia. Participants were recruited as part of the HERO-G (Hormonal Regulators of Growth) study. The study cohort was 238 newborns whose growth had been assessed longitudinally to 24 months of age. Table 1 summarizes data associated with the samples from individuals used in this study. The full HERO-G protocol is described elsewhere.²⁷ Placentas from women who delivered at home were collected by trained field workers and immediately transported on ice to the nearby Medical Research Council (MRC) Unit The Gambia Keneba laboratory (within 20–30 minutes) and carefully processed to obtain trophoblast material following a standard protocol (see placenta sample collection protocol in Data S1). Placental samples each of 400 mg were taken at four different

evenly spaced locations, at least 2 cm from the edge, and at consistent relative positions in each placenta to mitigate placental tissue heterogeneity. Samples were cut into four pieces, placed in RNeasy lysis buffer at a volume of 5 x tissue weight (Cat No 76106, Qiagen), and transported frozen on dry-ice to the United Kingdom for DNA extraction. After extraction samples from each of the four placental regions were pooled equimolarly. Cord blood and infant blood samples were collected into EDTA-lined tubes (BD Vacutainer, pink top) for DNA extraction in the United Kingdom. Ethical approval for the study was given by the joint Gambia Government/Medical Research Council (MRC) Unit The Gambia Ethics Committee (SCC 1313v3), with additional approval from the University of Colorado Institutional Research Board (protocol number 13–0441). Community approval was obtained from each participating village, and written, informed consent was obtained from each participating family. Samples for analysis were selected retrospectively from the study cohort representing (a) the highest 20% and lowest 20% birthweights and (b) according to the top and bottom 10% change in length-for-age (LAZ) from birth to 12 months. For the 12-month samples the male average age = 376.4 days, SD 9 days (366–409 d) and females average age = 378.8 days, SD 10 days (367–413 d). Table 2 summarizes the number of samples analyzed after quality testing for each tissue and test group and those that are common between groups.

2.2 | Nucleic acid extraction

DNA for DNA methylation studies was extracted from tissues using the Quick-DNA Mini Prep Plus kit (Cat No. D4068, Zymo Research). DNA extracted from blood followed the Biological Fluids and Cells protocol and DNA extracted from placenta followed the Solid Tissue protocol. DNA abundance and quality were determined after extraction using a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Absorbance ratios (A260/A280 and A60/A230) were above the recommended 1.8. DNA from each sample was further quantified on a Qubit® Fluorometer using Qubit® dsDNA HS Assay kit (Cat. No. Q32854, Thermo Fisher Scientific, Waltham, Massachusetts, USA).

2.3 | Methyl-Seq library preparation

Methyl-Seq was performed using the SureSelect^{XT} Methyl-Seq kit (Cat. No. G9651B, Agilent, Santa Clara, California, USA) according to the manufacturer's protocol (SureSelect^{XT} Methyl-Seq Target Enrichment System for Illumina Multiplexed Sequencing protocol, version C.0, January 2015); this covers over 3.7 million individual CpG

TABLE 1 Summary of Individual-Specific Data of Those Included in the Study

Mat ID	Mat Age	GA (wks)	Parity (Cat)	S_MOC	S_MOB	BW (kg)	BW Cat	BL (cm)	LAZ Change V1–12 m	LAZ Cat	Tissue
MALES											
	20.27	36	Primiparous	W	W	1.7	low	45.00	−0.45	low	Pl, CB, 12mB, 12mH
	35.19	38.1	Multiparous	D	D	2.14	low	44.43	1.55	high	Pl, 12mB, 12mH,
	33.32	38.1	Multiparous	D	D	2.32	low	47.00	0.53	mid	Pl, CB, 12mB
	38.61	38.6	Multiparous	W	D	2.38	low	45.50	−0.39	low	Pl, CB, 12mB, 12mH
	23.6	37	Primiparous	W	D	2.44	low	47.80	0.42	mid	Pl, CB, 12mB
	23.32	37.8	Multiparous	W	D	2.48	low	45.90	0	mid	Pl
	26.33	40.7	Multiparous	D	D	2.51	low	49.00	2.14	high	CB,12mB, 12mH
	27	39.9	Multiparous	W	W	2.53	low	49.00	1.7	high	Pl, CB, 12mB, 12mH
	24.39	38.9	Multiparous	D	D	2.56	low	47.40	0.99	high	Pl, CB, 12mB
	18.65	39.4	Primiparous	D	W	2.59	low	50.40	0.5	mid	Pl, CB, 12mB
	31.49	41.2	Multiparous	W	D	2.59	low	42.30	−0.49	low	Pl, 12mB, 12mH
	20.67	38.9	Primiparous	D	D	2.67	low	46.70	1.8	high	12mB, 12mH
	31.16	40.7	Multiparous	D	D	2.7	low	48.40	1.21	high	12mB, 12mH
		45		W	D	2.91	mid		0.97	high	12mH
	38.5	41	Multiparous	D	W	2.92	mid	49.00	−0.91	low	12mH
	40.82	39.7	Multiparous	D	W	2.96	mid	50.10	1.96	high	12mH
	28.69	40.7	Multiparous	D	D	3.06	mid	49.33	−0.43	low	12mH
	41.27	40.7	Multiparous	D	D	3.07	mid	52.00	−0.75	low	12mH
	32.82	39.9	Multiparous	W	D	3.13	mid	53.00	−0.66	low	12mH
	27.11	41.2	Multiparous	D	D	3.26	high	51.50	−0.49	low	12mH
	23.04	39.4	Multiparous	W	D	3.26	high	51.23	−0.91	low	12mH
	29.15	38.6	Multiparous	D	D	3.26	high	48.00	1.52	high	12mB, 12mH
	37	38.1	Multiparous	D	D	3.27	high	48.10	1.37	high	12mB, 12mH
	37.53	38.1	Multiparous	W	D	3.28	high	49.30	1.08	high	Pl, CB, 12mB, 12mH
	25.64	40.2	Multiparous	D	W	3.34	high	51.00	0.95	mid	Pl, CB, 12mB
	34.46	40.4	Multiparous	W	W	3.34	high	48.30	0.19	mid	Pl, CB, 12mB
	20.29	40.2	Multiparous	W	W	3.36	high	53.47	−1.01	low	Pl, CB, 12mB, 12mH
	37.18	39.1	Multiparous	D	D	3.36	high	50.50	1.82	high	Pl, CB, 12mB, 12mH
	22.07	41	Multiparous	D	W	3.37	high	48.47	0	mid	Pl, CB
	28.91	40.2	Multiparous	D	D	3.39	high	50.50	0	mid	Pl, CB
	23.51	38.9	Multiparous	D	W	3.45	high	50.47	1.02	high	Pl
	31.48	40.7	Multiparous	D	D	3.5	high	50.00	0.08	mid	12mB
	37.96	41.2	Multiparous	D	D	3.52	high	50.40	1.7	high	12mB, 12mH
	40.36	39.1	Multiparous	W	D	3.55	high	49.50	0.6	mid	Pl
	41.88	41	Multiparous	D	W	3.59	high	51.00	−0.33	mid	Pl, CB, 12mB
	39.33	40.7	Multiparous	D	D	3.72	high	52.50	−0.93	low	Pl, CB
	31.19	41.8	Multiparous	D	D	3.79	high	50.77	−0.28	mid	Pl, CB, 12mB
	39.69	40.2	Multiparous	D	W	3.8	high	53.00	−1.35	low	Pl, CB, 12mB, 12mH
	39.65	41	Multiparous	D	W	3.9	high	50.00	0.82	mid	Pl, CB, 12mB
	36.61		Multiparous	D	W				1.31	high	12mH
FEMALES											
	18.72	39.4	Primiparous	D	W	2.42	low	46.00	1.12	high	Pl, 12mB, 12mH,

(Continues)

TABLE 1 (Continued)

Mat ID	Mat Age	GA (wks)	Parity (Cat)	S_MOC	S_MOB	BW (kg)	BW Cat	BL (cm)	LAZ Change V1–12 m	LAZ Cat	Tissue
	19.04	39.9	Primiparous	W	W	2.46	low	46.00	0.34	mid	Pl, CB, 12mB
	29.89	39.1	Multiparous	D	D	2.48	low	50.00	0.54	mid	Pl, CB
	21.16	38.9	Multiparous	D	D	2.48	low	45.50	1.86	high	12mB, 12mH
	23.01	38.7	Multiparous	D	D	2.5	low	47.00	0.11	mid	12mB
	39.72	38.6	Multiparous	D	W	2.53	low	46.40	-0.28	mid	Pl, 12mB
	22.98	38.3	Multiparous	D	D	2.54	low	49.97	1.14	high	Pl, CB, 12mB, 12mH
	38.38	39.7	Multiparous	W	D	2.56	low	47.37	0.05	mid	Pl, CB, 12mB
	29.65	40.2	Multiparous	W	D	2.58	low	45.00	0	mid	Pl, CB
	26.92	38.1	Multiparous	W	D	2.61	low	46.50	0.65	mid	Pl, CB, 12mB
	41.69	38.1	Multiparous	D	D	2.63	low	48.50	0.9	mid	Pl, 12mB
	40.49	37	Multiparous	W	W	2.64	low	47.20	1.15	high	Pl, CB, 12mB, 12mH
	19.19	38.1	Primiparous	W	W	2.67	low	46.50	-4.44	low	Pl, CB, 12mB
	38.56	38.9	Multiparous	D	W	2.69	low	45.67	1.44	high	12mH
	32.99	40.7	Multiparous	D	D	2.73	low	46.27	-0.39	low	12mH
	23.95	39.9	Multiparous	W	W	2.96	mid	50.27	-0.47	low	12mH
	22.03	42	Multiparous	D	D	2.96	mid	49.40	1.67	high	12mH
	37.22	39.4	Multiparous	W	D	2.99	mid	49.50	-0.46	low	12mH
	26.08	38.3	Multiparous	D	D	3	mid	46.00	-0.46	low	12mH
	36.2	41.2	Multiparous	D	D	3.07	mid	50.07	-0.89	low	12mH
	38.07	40.4	Multiparous	D	D	3.21	mid	50.27	-0.76	low	12mH
	34.8	38.9	Multiparous	W	D	3.25	high	53.20	1.1	high	CB
	35.4	40.4	Multiparous	D	D	3.29	high	50.50	0.78	mid	CB
	36.78	40.2	Multiparous	D	W	3.31	high	49.00	0.07	mid	Pl
	32.43	39.7	Multiparous	W	D	3.33	high	51.00	0.81	mid	Pl, CB
	24.1	39.9	Multiparous	W	D	3.33	high	50.20	0	mid	CB
	34.3	39.4	Multiparous	W	D	3.33	high	50.00	-0.67	low	CB
	34.25	41	Multiparous	W	D	3.37	high	45.17	-1.39	low	Pl, CB, 12mB, 12mH
	39.54	40.2	Multiparous	W	D	3.42	high	49.80	-1.11	low	12mB, 12mH
	27.27	39.9	Multiparous	D	D	3.75	high	53.00	-0.17	mid	Pl, CB
	38.38	40.7	Multiparous	D	D	3.84	high	47.50	0.97	high	Pl, CB, 12mB
	27.12	41	Multiparous	D	D	3.97	high	52.00	-0.87	low	Pl, 12mB, 12mH
	27.07		Multiparous	D	W				1.1	high	12mH

Summary of individual-specific data for subjects contributing to study. Key: Mat Age, Maternal age; GA, Gestational age; S_MOC, Season of month of conception; S_MOB, Season of month of birth; BW, Birthweight; LAZ, Length for age Z score change birth to 12 months; BL, Birth length; D, Dry season; W, Wet season; pl, Placenta; CB, Cord blood; 12 mB, 12-month blood sample selected on birthweight; and 12 mH, 12-month blood sample selected on LAZ score. The samples categorized as high or low for both birthweight and length for age were those used in the analysis. Table 2 shows the numbers and sex for each tissue.

dinucleotide sequences covering CpG islands, CpG island shores, CpG island shelves, under-methylated regions, promoters, enhancers, transcription factors, CTCF-binding sites, DNase 1 hypersensitive sites, and DMRs. Three micrograms of DNA from each sample was initially sheared by Covaris sonication to 150–200 bp in size and used to prepare genomic DNA libraries with the SureSelect^{XT} Methyl-Seq Library Prep kit. After hybridization with the SureSelect^{XT} Methyl-Seq capture library, targeted regions

were isolated using complementary RNA baits. Isolated targets were bisulfite converted using the EZ DNA Methylation-GoldTM (Cat. No. D5005, Zymo Research) which converts unmethylated cytosines to Uracil, while methylated cytosines are unaltered. Subsequent PCR amplification creates an unmethylated CG→TA transition at unmethylated positions. Each sequence-modified, target-enriched library preparation was attached to a readable index (short DNA identifying code) by PCR. Libraries

were quantified on a Bioanalyzer 2100 (Agilent, Santa Clara, California, USA) using the Agilent High Sensitivity DNA kit (Cat. No. 507–4626, Agilent, Santa Clara, California, USA) or using a 2200 TapeStation (Agilent, Santa Clara, California, USA) with High Sensitivity DNA ScreenTapes (Cat. No. 5067–5593, Agilent, Santa Clara, California, USA). Equimolar indexed libraries were multiplexed (in groups of 6) to a final concentration of 4 nM in 20 μ L of nuclease-free dH₂O or 10 mM of Tris–Cl, pH 8.5 (Buffer EB, Cat. No. 19086, Qiagen) and run on a single flow cell on an Illumina NextSeq 500 according to the manufacturer's instructions using a 2x75 bp paired end read kit giving a total read length of 150 (TG NextSeq® 500 kit High Output Kit v2, Cat. No. TG-160–2002, Illumina). To overcome color imbalance inherent to low complexity in a bisulfite-converted genome, 10% of phiX genome was spiked into the reaction. Q30 scores of bases from NextSeq runs were within the threshold recommended by the manufacturer and depth of coverage was approximately 40x.

2.4 | DNA data mapping

FastQC v0.11.4²⁸ (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was used to visualize the sequencing quality of the raw reads which were then trimmed using Trim Galore! v0.4.0²⁹ (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). This removes low quality bases (Qscore <20) starting from the 3' end of the read. After trimming, short reads are removed (<20 bases). Figure S1 shows a typical example. The Bismark package uses Bowtie 2 alignment software v2.2.6^{30,31} to align sequences to the reference genome (GRCh38/hg19 assemblies) and then, methylation data were extracted employing default settings.³¹ Alignment mapping efficiency was in the region of 80% across all samples and is illustrated by Figure S2. The bisulfite error rate, estimated from the methylation status at cytosines outside a CpG context was in the region of 1.0% (Figure S3). Duplicated reads (removed using Bismark) were in the region of 20%. At each cytosine site, the methylation level was calculated as the ratio of the count of “C” (or the number of sequencing reads with methylated cytosine) to the count of “C” plus “T” (or the total number of reads covering that site). M-bias plots were generated after methylation data were extracted with Bismark to yield the percentage methylation across all reads in order to identify any bias (e.g., bias at the end of the read due to drop in quality) arising from the position in the read of the cytosine residue being called. Figure S4 illustrates an example (for infant bloods) of an M-bias plot illustrating the reduction of call quality at the 5' and 3' ends of the paired reads. This provides a guide to the extent of necessary sequence trimming (typically four

bases removed from the 5' end and one from the 3' end). Methylation information was then re-extracted and the output was processed and converted to a bedgraph.

2.5 | Methylation data analysis

The resulting bed files from Bismark were used for further statistical analysis. Three comparison groups based on different growth criteria were examined: (a) high versus low birthweight babies sampled at birth for placenta and cord blood, (b) high versus low birthweight groups sampled at 12 m for infant blood, and (c) high versus low length-for-age based on change in Z score (LAZ) between birth and 12 months sampled at 12 months for infant blood. Differential methylation between groups was examined using BiSeq.^{31,32} Only CpGs covered by at least 10 reads were included in the analysis.

2.6 | Detection of DMRs

Analysis was performed using R v 3.2.2³³ and BiSeq version 1.18.0 (32, see review 34). BiSeq is designed specifically for targeted bisulfite sequencing data and includes features such as limiting high coverage, removing low coverage, spatial correlation, a multiple testing correction, visualization, and genomic annotation. DMRs were detected by comparing birthweight categories (high vs. low) or length-for-age (LAZ) scores (high vs. low) and incorporated sex as a covariate. Briefly, sequences were grouped into clusters of adjacent CpG sites. CpG methylation often occurs in clusters and spatial correlation is a key characteristic of DNA methylation. As methylation is conserved across short distances, identification of these related regions reduces data dimensions and also increases detection power by borrowing nearby CpG information. BiSeq CpG clusters were defined as CpG sites covered in at least 25% of samples (defined as frequently covered CpG sites) with a maximum distance of 100 bp between CpG sites within a cluster and with clusters containing at least 5 of these CpGs. To mitigate sequence overrepresentation distorting the data, sequences with greater than 90% of maximum coverage were removed. The methylation data were smoothed within CpG clusters using the smoothing algorithm (“predictMeth”). This estimates the true methylation level of each site in each sample. The methylation data were tested for both the test groups and resampled datasets under the null hypothesis that differences in methylation are random. The data from both were modeled by beta regression, with the group as the independent variable and the methylation probability as the dependent variable. A Wald test was used to confirm the parameters used in the beta regression could be included in the model and associated *p*-values were transformed into Z scores to allow DMRs to be detected.³¹

TABLE 2 Summary of DNA Samples Analyzed in the Study

	DNA extraction		Total			Total
	Low BW	High BW		Low LAZ	High LAZ	
Placenta						
Male	10	14				24
Female	11	6				17
Both	21	20				41
Cord bloods						
Male	8	12				20
Female	8	8				16
Both	16	20				36
	DNA extraction					
	Low BW	High BW	Total	Low LAZ	High LAZ	Total
Infant blood (12 m)						
Male	12	13	25	11	13	24
Female	11	4	15	9	7	16
Both	23	17	40	20	20	40
Subjects in Common						
Placenta BW and cord blood BW	30					
Cord blood BW and Infant blood (12 m) BW	26					
Cord blood BW and Infant blood (12 m) LAZ	11					
Infant blood (12 m) BW and Infant blood (12 m) LAZ	22					

Numbers of DNA samples analyzed for each tissue according to sex and test group and the number of subjects in common between tissues and test groups. Key: BW, Birthweight; LAZ, Length for age Z score.

To account for multiple testing errors (multiple testing correction using the Benjamini–Hochberg method),³⁵ a two-step hierarchical procedure was employed. This first tests clusters, then individual CpG sites within those clusters. The two-step approach avoids loss of power by first testing at the cluster level and then, the CpG in the cluster that showed a change in methylation and hence the number of CpGs needing correction is greatly reduced. A variogram was created under the null hypothesis, which estimates the correlation in methylation between two CpG sites within a cluster. This was plotted and smoothed, with a sill of 1 for all our tests and was combined with the Z scores of the test results of interest to estimate the correlation of Z scores between two locations in a cluster. Clusters without differentially methylated CpG sites were removed (FDR >= 0.1), before the remaining

clusters were trimmed to indicate individual significant CpG sites (FDR <= 0.05). PCA analysis of methylation patterns determined from different DNA sequence runs did not reveal any batch effects.

2.7 | Pyrosequencing

Validation of differentially methylated cytosines as detected by Methyl-Seq was performed by bisulfite pyrosequencing on the *ZFHX3* gene. Initially, PCR primers were designed using the Pyromark assay design SW 2.0 (Cat. No.9019077, Qiagen, USA) and were supplied by Sigma-Aldrich, UK. One of the primers was biotinylated and purified by HPLC. The primers were; *ZFHX3*: forward PCR

primer GTTTAAATTTGATTGGGGGAAAG, reverse PCR primer CCTTAAACAACTAACCTCCTAACA, and forward biotinylated sequencing primer TTTTTTAAATGTAGATTTGAATT.

PCR amplification was performed with 10 ng of bisulfite converted DNA using EpiTaq HS (Cat. No. R110A, TaKaRa Bio Inc, Japan). PCR was set up according to the manufacturers' instructions but the concentration of $MgCl_2$ varied between 15 and 25 mM dependant on the primer set. Both methylated and unmethylated controls from the EpiTect PCR control DNA kit (Cat. No. 59695, Qiagen, USA) were run alongside. Thermal cycling conditions were performed using a touchdown program with an annealing temperature range of 53°C–62°C and cycle number range of 25–35, dependant on primer set. The PCR products were electrophoresed on a 3% of agarose gel to check for product specificity. Pyrosequencing was then performed on the PyroMark Q24 Vacuum Workstation (Qiagen, USA) as described in the manufacturer's instructions. PyroMark CpG software Design 2.0 (Cat. No. 9019067, Qiagen, USA) was used in this assay and primers with the best quality score were selected. Bisulfite conversion was shown to be efficient for all samples as the fluorescence signal by cytosine in a non-CpG context was $\leq 1\%$ of the signal produced by thymine.

2.8 | Cellular heterogeneity assessment between sample groups

For cord blood, cell composition was compared between low and high birthweight groups using overlaps with a cord blood cell type-specific reference panel of 215,000 CpGs derived from the Illumina EPIC 850 k array (ref: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6284779/>). The reference panel set of CpG loci was used to find overlaps with the processed Methyl-Seq capture dataset. Co-methylation patterns extend up to several 100 base pairs across CpG clusters.^{36,37} In order to obtain enough coverage for the regions covered by the EPIC reference set, we used intervals of 200 bases centered around the locations of the EPIC reference CpG set (updated in Human Genome—HG19). This yielded 14993 regions each containing CpG loci as present in the processed methyl capture sequence dataset. Methylation calls were extracted from the processed sequence data as described above and the mean values in these regions were used to generate PCA plots and heatmaps to calculate the correlation values between experimental groups (using Pearson correlation). In the absence of a 12-month blood reference panel, the adult blood reference panel based on the Illumina Infinium HM450 k and EPIC 850 K methylation chips^{38,39} was used and processed in the same way for coverage across the Methyl-seq capture dataset. An interval of 200

bases yielded 33 regions each containing CpG loci (providing coverage for CD4 and CD8 lymphocytes, NK cells, neutrophils, B-cells, and monocytes) that are present in the processed methyl capture sequence dataset to compare the 12-month groups.

2.9 | CpG and gene annotation

Ensembl was used to annotate differentially methylated CpGs (based on hg38 version GRCh38 human genome build) to determine their location with respect to regulatory features. Ontologies, mutational and functional data of those genes associated with significant differentially methylated CpGs were determined using the U.S. National Center for Biotechnology Information (NCBI; Bethesda, MD, USA; <http://www.ncbi.nlm.nih.gov/>) Gene, Online Mendelian Inheritance in Man (OMIM), PubMed databases and the Database for Annotation, Visualization, and Integrated Discovery v6.7 (DAVID - <http://david.abcc.ncifcrf.gov/>).⁴⁰ Disease associations were determined by interrogating the Genetic Association Database (GAD) for complex diseases and the EBI GWAS Catalogue. PANTHER v14.0 (<http://www.pantherdb.org/>)^{41,42} was used to provide an overview of gene ontology (GO Terms) defining protein classes, cellular components, biological processes, and molecular functions of genes implicated by methylation marks.

3 | RESULTS

3.1 | Quality Triage of Sample Cohorts

All samples underwent assessment to exclude maternal contamination and poor-quality samples. Maternal blood contamination of cord blood (for both sexes) was assessed using marker CpGs that are only methylated in adult blood DNA, and maternal contamination of placenta trophoblast samples from males was also flagged by examining the levels of Y DNA methylation dilution⁴³; see Figure S5a,b. Poor quality and/or obvious outlier samples were identified by plotting a heatmap of the methylation data for each experimental group (see example of the methylation data from the birthweight cohort at 12 months of age in Figure S6). The major component of variation was sex. Principal Component Analysis (PCA—done with and without inclusion of the sex chromosomes) matrices were also applied to a list of available variable information for the subjects contributing to each cohort and tissue sample (see Table S1) to determine whether they had a significant effect on the variation in the data. Examples for sex (male/female), birthweight category (high/low), and season (dry/wet) are illustrated in Figures S7a,b,c; only sex contributed significantly

to variation in the data. Sample sets emerging from these analyses were re-analyzed with sex as a covariate.

3.2 | Assessment of Confounding Cellular Heterogeneity

There is no available cell-type-specific reference set for cord and adult blood to assess cell composition changes based on DMRs detected by Methyl-Seq data. Potentially confounding differences in cell-type composition between cord blood groupings (high and low birthweight) and infant blood groupings (high and low birthweight and length for age) were, therefore, assessed using DMR regions within the capture DNA sequence dataset that overlap within a 200 base-pair interval with the EPIC 850 k cord blood and Infinium HM450 k adult blood cell-type-specific CpG panels (see methods). The adult overlaps were used for the 12-month infant blood data in the absence of an age-related reference panel for this time-point. The heatmap and PCA plots are shown in Figures S8 and S9. For both cord and infant blood data, the heatmaps indicate high correlation between samples and no clear clustering according to comparison groups. PCA analysis indicates inter-individual differences in cellular composition. For the small number of probes from the adult blood reference panel present within the Methyl-Seq capture DNA sequence dataset, the analysis shows separation of individual samples into two groups; this may reflect the small number of probes available and their disproportionate weighting or variation in the rate of loss of nucleated erythrocytes between individuals. However, in all three PCA plots the variation between samples captured in PC1 and PC2 is distributed fairly uniformly across both experimental groups (high or low birthweight or tall or short height for age) indicating little difference in cellular composition between comparison groups to confound the determination of differential methylation values at the same time-point.

3.3 | Differentially Methylated Loci Identified Through BiSeq Analysis

The total number of significant differentially methylated regions (DMRs) and the direction of median methylation change identified for each comparison group (placenta-birthweight, cord blood-birthweight, infant blood-birthweight at 12 months, and infant blood-LAZ) by BiSeq analysis and the total number of significant CpGs they contain is shown in Table S2a–d and summarized in Table 3; 194 CpG loci in total. Each significantly differentially methylated CpG in each DMR was examined for the presence of single-nucleotide polymorphism (SNP) directly in the CpG; these are shown in Table S2a–D. Apart from three SNP-containing CpGs, all the

MAFs (minor allele frequencies) were <0.01 . For the CpGs associated with implicated genes *RPS6KA2*, *PRSS3*, and *GARI*, the MAFs were <0.03 , <0.11 , and <0.02 , respectively. These MAFs are at a level that would not significantly alter the estimation of methylation differences between test groups.

The distribution of CpGs between gene regulatory elements, gene bodies, and intergenic regions is shown in Table 4 (see Table S2a–d for full details on all DMRs and CpG locations). Figure 1A–D provides an overview of the gene ontology (determined using PANTHER v14.0 available at <http://www.pantherdb.org>) characterizing genes implicated by differential methylation marks. The pie charts summarize the distribution of this gene set across GO terms defining molecular functions, biological processes, cellular components, and protein classes. It can be seen that certain GO categories predominate. For example, analysis of molecular function reveals that binding, catalytic activity, molecular function regulator, and transcriptional regulator activity are most prominent. Detailed information on the genes and proteins in each of the GO categories can be obtained by uploading the gene lists to <http://www.pantherdb.org> from Table S2a–b and interrogating each pie chart sector.

Very few of the differentially methylated CpGs found in DMRs identified from the cord blood comparisons are found in the 12-month infant blood comparisons; (a) of the four closely linked CpGs associated with *TNXB*, one (upstream intergenic) is differentially methylated in the cord blood birthweight group and the remaining three (within intron 1 of the gene) in the 12-month infant blood birthweight group and (b) the intergenic CpG upstream of the *HLX* gene is differentially methylated in both the 12-month birthweight and 12-month length for age groups.

3.4 | CpG Loci Showing 5% or Greater Methylation Change

We have chosen to focus on those marks that show 5% or greater methylation change. Figure 2 summarizes the CpGs that have been located to regulatory features (promoters, CTCF-binding sites, and transcription factor-binding sites—56 CpG loci in total) and figure 3 those located to gene bodies (introns and exons) and closely linked intergenic regions (64 CpG loci in total). Figures 2 and 3 also present the locations of CpGs (based on hg38 release 85 from ENSEMBL) with respect to the Transcription Start Site (TSS) of genes implicated by location (93 in total), the median *p*-value for the DMR corrected for multiple testing, direction and change in median methylation value, GWAS disease associations, and a short vignette summarizing any mutational data and functional studies of implicated genes culled from the various databases outlined in the materials and methods. Finally, Figures 2 and 3 flag whether any of the DMR-associated genes are also subject to Trans or Cis-meQTLs (genetic variation that influences methylation at

TABLE 3 Summary of Numbers of DMRs, CpG Loci, and Implicated Genes

Cohorts	Total Number of DMRs	Direction of Median Methylation Change for DMRs Relative to High Groupings for Birthweight and LAZ		Total Number of CpG sites in DMRs and implicated genes
		+ve	-ve	
Placenta BW	4	2	2	4 (4)
Cord blood BW	68	25	43	88 (78)
Infant blood (12 m) BW	54	29	25	71 (65)
Infant blood (12 m) LAZ	25	13	12	31 (26)

Summary of total number of DMRs, CpG loci, implicated genes (in brackets), and direction of median methylation change identified from the comparisons made at each time-point between groupings. Key: BW, Birthweight; LAZ, Length for age Z score. Median methylation change is expressed relative to the high birthweight and high length for age groupings.

CpG sites adjacent to implicated genes) and/or Cis-eQTMs (variation of methylation that influences expression of an adjacent gene) collated in the Bios QTLBrowser held at www.genenetwork.nl/biosqtlbrowser (85), these are marked in red in the first column). These data were based on the analysis of cohorts from the Dutch population and may only partially reflect genetic variation in the Gambian population with Trans or Cis-eQTL effects.

The implicated gene names in Figures 2 and 3 are color coded according to categories of gene function/disease revealed by functional and/or mutation analysis (see legends) and Figure 4 summarizes the numbers of implicated genes found in associated disease categories bearing the same color coding. It is immediately clear that neurological, growth and development, and oncological disorders are the most prominent among the implicated genes showing 5% or greater methylation change.

3.5 | Replication of Findings in Other Methylation Studies

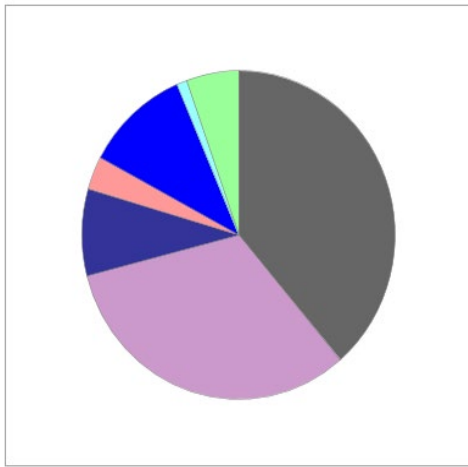
Identification of a significant proportion of the same implicated genes reported in related studies provides strong

validation of the findings reported here. Highlighted in red bold in Table S2a–d are the DMR-implicated genes that are also documented in the recent large meta-analysis of multiple EWAS by the Childhood Epigenetics Consortium examining DNA methylation associated with birthweight.¹⁵ When all genes (4848, representing about 19% of the estimated 25,000 genes in the human genome) from the consortium study associated with 8170 CpGs significant after FDR correction for multiple testing are screened, 62 DMR implicated genes from the current study show a match (of which 34 show >5% methylation change). If this is restricted to those genes (729; about 2.9% of human genes) associated with 914 CpG loci surviving Bonferroni correction ($p < 1.06E-7$), then 11 matches are found (marked with a red asterisk of which 10 show >5% methylation change in the current study). The vast majority of CpGs in the meta-analysis are located within or closely linked to genes. Thus taking the number of genes in the genome as 25,000, the approximate probability of a match by chance for any given DMR-implicated gene in the present study is 0.19 (4848/25000) for all genes and 0.029 (729/25000) for those associated with the 914 CpG loci. The probability that these matches have occurred by chance for 62 and 11 genes is 0.19^{-62} and 0.029^{-11} , respectively.

TABLE 4 The distribution of methylation marks between regulatory features, gene bodies, and intergenic regions

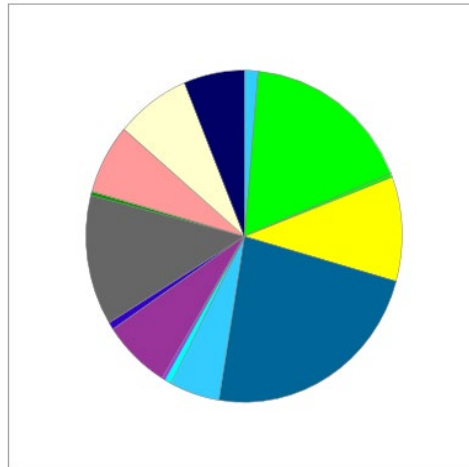
Genomic Feature	Number of CpGs in Feature	Number of CpGs in Feature	% of Total
	>5% median methylation change	<5% median methylation change	
Promoter	30	44	37.9
Promoter and CTCF-binding site	9	3	6.2
CTCF-binding site	11	3	7.2
Transcription Factor-binding site	4	1	2.5
Enhancer	0	1	0.5
Exon	16	5	10.8
Intron	28	9	19.5
Intergenic	17	13	15.4

(A) Panther GO-Slim Molecular Function



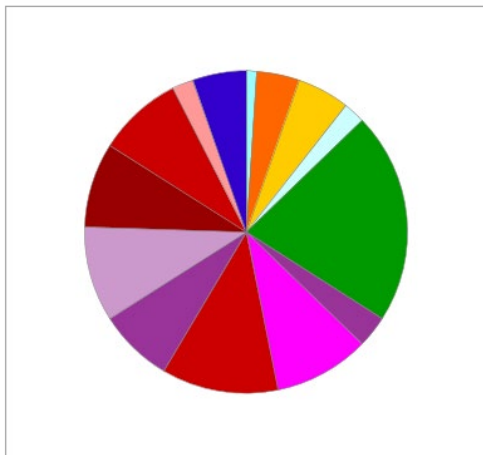
- Binding (GO:0005488)
- Catalytic Activity (GO:0003824)
- Molecular Function Regulator (GO:0098772)
- Molecular Transducer Activity (GO:0060089)
- Transcription Regulator Activity (GO:00140110)
- Translation Regulator Activity (GO:0045182)
- Transporter Activity (GO:0005215)

(B) Panther GO-Slim Biological Process



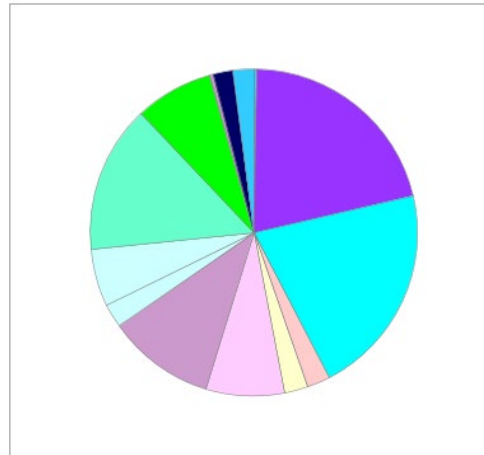
- Biological Adhesion (GO:0022610)
- Biological Regulation (GO:0065007)
- Cell Population Proliferation (GO:0008283)
- Cellular Component Organisation or Biogenesis (GO:0071840)
- Cellular Process (GO:0009987)
- Developmental Process (GO:0032502)
- Growth (GO:0040007)
- Immune System (GO:0002376)
- Localization (GO:0051179)
- Locomotion (GO:0040011)
- Metabolic Process (GO:0008152)
- Multi-organism Process (GO:0051704)
- Multicellular Organismal Process (GO:0032501)
- Response to Stimulus (GO:0050896)
- Signalling (GO:0023052)

(C) Panther GO-Slim Protein Class



- Cell Junction Protein (PC00070)
- Chromatin/Chromatin-Binding or Regulatory Protein (PC00077)
- Cytoskeletal Protein (PC00085)
- Extracellular Matrix (PC00102)
- Gene-specific Transcriptional Regulator (PC00264)
- Intracellular Signal Protein (PC00207)
- Membrane Traffic Protein (PC00150)
- Metabolite Interconversion Enzyme (PC00262)
- Nucleic Acid Binding Protein (PC00171)
- Protein Modifying Enzyme (PC00260)
- Protein-binding Activity Modulator (PC00095)
- Scaffold Adapter Protein (PC00226)
- Transmembrane Signal Receptor (PC00197)
- Transporter (PC00227)

(D) Panther GO-Slim Cellular Component



- Synapse (GO:0045202)
- Cell Part (GO:0044464)
- Cell (GO:0005623)
- Extracellular Region Part (GO:004421)
- Extracellular Region (GO:0005576)
- Membrane Part (GO:0044425)
- Membrane (GO:0016020)
- Membrane-enclosed Lumen (GO:0031974)
- Organelle Part (GO:0044422)
- Organelle (GO:0043226)
- Protein-containing Complex (GO:0065007)
- Supramolecular Complex (GO:0099080)
- Synapse Part (GO:0044456)

FIGURE 1 Gene Ontology Analysis of Genes Implicated by Associated Methylation Marks. Panther version 14 was used to provide an overview of the gene ontology characterizing genes implicated by methylation marks. Panther 14.0 identified 161 hits from the uploaded list of 173 genes. A. GO terms for Molecular Function found 93 molecular function hits. B. GO terms for Biological Process found 276 process hits. C. GO terms for Protein Class found 94 class hits. D. GO terms for Cellular Component found 300 cellular component hits. Color coding has been assigned starting at 12 o'clock and working clockwise on the pie chart

Comparisons have been made to two further studies examining the impact of gestational age⁸⁶ (where some of the data are subsumed in the large meta-analysis mentioned above) and smoking on birthweight⁸⁷; these share, respectively, 53 (marked with a green asterisk in Table S2a–d) and 11 (marked with a blue asterisk) genes associated with differential methylation identified in the current study. These two studies identify a further 22 DMR-associated genes that overlap with our findings, bringing the replication in other related studies to 84 (48%) of the 173 implicated genes we have documented (The asterisks in figure 2 and 3 mark which of those shared genes show 5% or greater methylation change—in total 49 of the 93 in tables 5 and 6 between the three birthweight-related studies).

The genes *ZNF678*, *VTRNA2-1*, *SCRIB*, and *TNXB* match those reported in other studies on maternal exposures and differential methylation of associated CpG loci in Gambian infants^{22,26} (marked with black triple asterisks in Table S2a–d) and *SEMA3B*, *ARID1B*, and *HOXA10* from the Cambridge Baby Growth Study⁸⁸ (marked with black double asterisks).

The high degree of replication observed in related studies provides robust validation of the findings reported here. Pyrosequencing analysis of the methylation mark associated with the *ZFH3* gene was performed to illustrate an example experimental confirmation of methyl-seq derived methylation data. Table 5 summarizes the data for several individuals selected from the high and low groups for the 12-month LAZ comparison. The results in Table 5 show good concordance between the two methods in both the quantum and direction of change when compared to the median methylation value change derived from the group comparisons by BiSeq analysis of methyl-seq data.

4 | DISCUSSION

It has been suggested that epigenetic changes may be involved in the mechanism of reprogramming induced by under-nutrition, infection, and adverse environmental exposures, although, it is not clear whether these are primary or

secondary events in the chain of causality. This paper has used extremes of variation in birthweight and subsequent gains in length to examine associated methylation changes in DNA from trophoblast and cord blood DNA from small and large babies, and blood DNA from 12-month old infants analyzed both according to their size at birth and their change in length from birth to 12 months (LAZ).

The methylation marks found at birth and those at 12 months in relation to birthweight show little longitudinal persistence (see figures 2 and 3 and Table S2a–d). This suggests ongoing epigenetic adjustments, significant changes in blood cellular composition (such as the loss of nucleated erythrocytes⁸⁹ or both in the critical postnatal growth period, and the subsequent infancy-childhood growth transition (ICT)).⁹⁰ Nonetheless, what does persist at 12 months is different, almost completely non-overlapping methylation patterns (not confounded by cellular composition differences) between the high and low birthweight comparison groups and the comparison groups showing rapid or slow postnatal height gain. These two distinct methylation patterns may reflect different interactions with nutritional, infectious, and other environmental exposures during the postnatal growth phase potentially associated with negative or positive growth trajectories or a combination of both. Thus, any continued challenges (such as those provoked by under nutrition and infection) to homeostasis during the development period may trigger epigenetic programming and shift the timing and duration of these periods of growth. The study reported by Bernstein et al.⁹¹ has revealed an accelerated transition to a childhood pattern of growth in Gambian compared to UK infants. A later transition, observed in U.K. infants, extends the high growth rate experienced during the infancy stage. This is reduced in Gambian infants, potentially impacting on growth outcomes in childhood while diverting energy into other processes critical for responses to acute infectious challenges; later developmental stages in this population offer an extended window for catch-up growth.

Over half (54.3%) of the identified methylation marks are located in gene regulatory elements, 30.3% in gene bodies, and the remaining 15.4% in intergenic regions closely linked to implicated genes. Alteration of gene activity by methylation of

FIGURE 2 Implicated Genes Associated with Methylation Changes of 5% or Greater in Regulatory Elements. This figure documents those genes where a methylation change of 5% or greater has occurred within a defined regulatory feature. It also provides a summary of function and any disease associations resulting from genome-wide association studies (GWAS—culled from the GAD and EMBL genetic association catalogue databases), mutation analysis, and functional investigations. The methylation change is expressed relative to the high birthweight and high length for age groups. All mapping of DMRs is based on human genome build hg38 version GRCh38 of the human genome. The positions of CpGs is given in relation to the Transcription Start Site (TSS) of the implicated gene. Also shown highlighted in red in the first column is whether the gene is associated with Trans and/or Cis-meQTLs and/or Cis-eQTLs. Where plural is shown, this indicates two or more Trans-meQTLs, Cis-meQTLs, or Cis-eQTLs associated with the gene (information obtained from the BIOS QTL Browser at www.genenetwork.nl/biosqtlbrowser). NI, No Information. Color key of gene disease and functional associations: brown, neurological; purple, fertility; light blue, growth and development; dark blue, oncological; and light green, immunological. The asterisks mark the implicated genes found associated with methylation marks in related studies.^{15,86,87}

implicated genes may occur by impacting the functionality of cis-transcriptional regulatory elements or changing chromatin conformation and accessibility to the transcriptional machinery. Several of the methylation marks are found in the binding

site (an estimated 326,000 in the human genome) for the multifunctional CTCF zinc finger protein. This protein plays a key regulatory role through a number of varied functions that include influencing chromatin architecture (binding at chromatin

Implicated Gene and Regulatory Feature containing CpG	Location of Methylation Change	Median % Methylation Change and p-value	Disease Associations (GWAS)	Entrez Summary; Phenotype from Functional Studies and Mutation Analysis
BiSeq Analysis >5% Methylation Change				
Trophoblast				
AP2M1: Adaptor related protein complex 2 mu 1 subunit.* CpG in internal promoter 4.248 kb downstream of TSS	3q27.1: position 184178937	+9.3 p=5.16E-08	HIV Infections[X]Human immunodeficiency virus disease, kidney failure, chronic	A subunit of the heterotetrameric coat assembly protein complex 2 (AP2), which belongs to the adaptor complexes medium subunits family. The encoded protein is required for the activity of a vacuolar ATPase, which is responsible for proton pumping occurring in the acidification of endosomes and lysosomes. The encoded protein may also play an important role in regulating the intracellular trafficking and function of CTLA-4 protein. Mutations in <i>AP2M1</i> cause deficits in intellectual development and epileptic seizures (44). See OMIM entries 601024, 618587
Cis-meQTLs Cis-eQTM				
DLG3: Discs large MAGUK scaffold protein 3 CpG in internal promoter 10.961 kb downstream of TSS	Xq13.1: position 70455796	-19.3 p=6.84E-08	Several psychiatric disorders, Type 2 Diabetes] edema rosiglitazone	The encoded protein may play a role in clustering of NMDA receptors at excitatory synapses. It may also negatively regulate cell proliferation through interaction with the C-terminal region of the adenomatous polyposis coli tumor suppressor protein. Mutations in this gene have been associated with X-linked mental retardation. <i>MRX90</i> (45, 46)
Cord Blood Birth Weight				
MARK1: Microtubule affinity regulating Kinase 1 CpG in 5' promoter 559 bp upstream of TSS	1q41: position 220527577	+6.9 p=5.65E-06	Autism, Body Mass Index, Tobacco Use Disorder	A kinase involved in the phosphorylation of microtubule-associated proteins. Shown to be associated with autism and over-expressed in the prefrontal cortex. See OMIM entries 606511 and 610836
AC0108940.3: anti-sense to WIPF1 and CHRNA1 CpG in promoter within intron 136.296 kb downstream of TSS and 521 bp upstream of WIPF1 TSS	2q: position 174683437	+10.23 p=7.32E-07	NI	NI
VTRNA2-1: Vault RNA 2-1* CpGs in 5' promoter and CTCF binding site 113 bp downstream of TSS and and CTCF binding site 631 bp downstream of TSS in intergenic region	5q31.1: position 136079966 136080484	+30.9 p=4.56E-07	NI	Direct RNA inhibitor of EIF2AK2 and has important role in cell growth. Acts as a tumour suppressor and is hypermethylated in cancers (47, 48, 49, 50). It is an imprinted locus (51, 52). See OMIM entries 614938, 176871. Associated CpG loci show differential methylation in response to nutritional exposures during the periconceptual period (25)
RPS6KA2: Ribosomal protein S6 kinase * CpG in CpG island and transcription factor binding site 6.467 kb downstream of TSS in first intron	6q27: position 166899984	-8 p=4.01E-06	Colonc Neoplasms] Rectal Neoplasms, Diabetes Mellitus, Heart Failure, Maximal Expiratory Flow Rate, Platelet Count, Respiratory Function Tests, Tobacco Use Disorder	Member of the RSK (ribosomal S6 kinase) family of serine/threonine kinases. This kinase contains two non-identical kinase catalytic domains and phosphorylates various substrates, including members of the mitogen-activated kinase (MAPK) signalling pathway. The activity of this protein has been implicated in controlling cell growth and differentiation. See OMIM entry 601685
Cis-meQTLs Cis-eQTM				
AL078602.1: LncRNA CpG in CTCF binding site within intron 368.430 kb down stream of TSS	6q14.1: position 163834847	-12.4 p=2.26E-06	NI	NI
TTYH3: Tweety family member 3* CpG in CTCF binding site 20.122 kb downstream of TSS within intron of gene	7p22: position 2652108	-7.6 p=4.3E-08	NI	Member of the tweety family of proteins that function as chloride anion channels. The encoded protein functions as a calcium(2+)-activated large conductance chloride(-) channel. See OMIM entry 608919
Cis-meQTLs Cis-eQTM				
BHLHA15: Basic helix-loop-helix family member A15 * CpG in 5' promoter and CpG island 608 bp upstream of TSS	7q21.3: position 98210819	+5 p=1.06E-06	Body Fat Distribution, Body Weights and Measures, Chronic renal failure] Kidney Failure, Chronic, Echocardiography, null, prostate cancer, Prostatic Neoplasms	Belongs to the protein tyrosine kinase family and interacts with several other proteins, such as Inhibitor-2 (Inh2), protein phosphatase-1 (PP1C), p35, and myosin VI. It phosphorylates other proteins, and is itself also phosphorylated when interacting with cyclin-dependent kinase 5 (cdk5)/p35 complex. This protein is involved in nerve growth factor (NGF)-TrkA signalling, and also plays a critical role in endosomal membrane trafficking. Mouse studies suggested an essential role of this protein in spermatogenesis. See OMIM entry 608606
SH3GL2: SH3 domain containing GRB2 like 2, endophilin A1. CpGs in promoter and CpG island 686 and 693 bp downstream of TSS	9p22: positions 17579752 17579759	+5.4 p=5.05E-07	Apolipoproteins B, Cardiomegaly, Carotid artery stenosis] Carotid Stenosis, Cognitive performance, Creatinine, Echocardiography, Folic Acid Stimulating Hormone, Heart Failure, Iron, Leprosy, Multiple Sclerosis, Neoplasms, Parkinson Disease, Psychomotor Performance, Pulse, Respiratory Function Tests, smoking cessation, Tobacco Use Disorder, Walking	Interacts with the cytoplasmic domains of the metalloprotease disintegrins ADAM9 and ADAM15. Gene is essential for the formation of synaptic-like microvesicles from the plasma membrane. Interacts with parkin (PARK2) and involved in parkin-mediated synaptic ubiquitination. See OMIM entry 604465
DPH7: Diphthamide biosynthesis 7. CpG in promoter and CpG island 54 bp downstream of TSS	9q34.3: position 137578871	+5.3 p=3.82E-185	Tobacco Use Disorder	Diphthamide is a post-translationally modified histidine residue present in elongation factor 2, and is the target of diphtheria toxin. DPH7 contains a WD-40 domain, and is thought to be involved in diphthamide biosynthesis. See OMIM entry 613210
DRD4: Dopamine receptor D4 Two CpGs in 5' promoter and CpG island 812 and 816 bp upstream from TSS	11p15.5: positions 636713 636716	+11.8 p=2.47E-06	NI	G protein-coupled receptor that belongs to the dopamine D2-like receptor family. Functionally, the D2-like receptors are characterized by their ability to inhibit adenylyl cyclase. Allelic variants of the gene have been associated with ADHD (Attention Deficit Hyperactivity Disorder), Novelty-Seeking Personality Trait and Risk-Taking Behavior, Alcoholism, Mood Disorders. See Omim Entry 126452 See review (53)
BCKDK: Branched chain ketoacid dehydrogenase kinase* CpG in 5' internal promoter flank 3.15 kb downstream of TSS	16p11.2: position 31109257	-6.6 p=1.99E-07	Acquired Immunodeficiency Syndrome]Disease Progression, Type 2 Diabetes] edema rosiglitazone	The branched-chain alpha-ketoacid dehydrogenase complex (BCKD) is an important regulator of the valine, leucine, and isoleucine catabolic pathways. The protein encoded by this gene is found in the mitochondrion, where it phosphorylates and inactivates BCKD. Homozygous mutation found in individuals with autism, epilepsy and intellectual disability. Mouse knock-out model shows delayed growth and neurological abnormalities. See OMIM entries 614901 and 614923
DOC2A: Double C2-like domain-containing protein , alpha* CpG in CTCF binding site with exon of gene 16.817 kb downstream of TSS	16p11.2: position 30006453	-14 p=5.58E-07	Autism, Narcolepsy, Schizophrenia, schizophrenia autism	Mainly expressed in brain and involved in Ca(2+) dependent neurotransmitter release. Association with autism, narcolepsy and schizophrenia.

MYO1D: Myosin ID* CpG in CTCF binding site 218,139 kb downstream of TSS within intron of gene	17q11.2-q12: position 32659038	+19.3 p=9.81E-07	autism, Body Height, hypertension, Magnesium, Pancreatic Neoplasms, Type 2 Diabetes edema rosiglitazone	Chiral determinant through interaction with actin cytoskeleton. Function in right left patterning in development. See OMM entry 606539
SECTM1: Secreted and transmembrane 1* CpGs in CTCF binding site and CpG island in last exon of gene 13.13, 13.025, 13.027 kb downstream of TSS Cis-meQTLs Cis-eQTM	17q25: positions 82321061 82321072 82321074	-15.8 p=1.9E-06	Birth Weight Leukemia Leukemia, Myeloid, Acute Precursor Cell Lymphoblastic Leukemia-Lymphoma, Meningeal Neoplasms meningioma	A transmembrane and secreted protein with characteristics of a type 1a transmembrane protein. It is found in a perinuclear Golgi-like pattern and thought to be involved in hematopoietic and/or immune system processes.
MCOLN1: Mucopolip 1 CpG in 5' promoter 717 bp upstream from TSS Cis-meQTL Cis-eQTL	19p13.3-13.2: position 7521907	-17.4 p=3.65E-08	NI	A transmembrane protein localized in intracellular vesicular membranes. Functions in the late endocytic pathway and regulation of lysosomal exocytosis. Mutations cause mucopolipidosis type IV: ML4, a neurodegenerative lysosomal storage disorder with psychomotor retardation and ophthalmologic abnormalities. Mouse model has similar phenotype and neurological deficits. See OMM entries 605248 and 256250
NDUFA7: NADH:ubiquinone oxidoreductase subunit A7. CpG in 5' promoter 575 bp downstream of TSS	19p13.2: position 8320800	-7.1 p=9.9E-185	Acquired Immunodeficiency Syndrome Disease Progression Aging Telomere Length,cognitive trait,drug-related genes	This gene encodes a subunit of NADH:ubiquinone oxidoreductase (complex I), which is a multiprotein complex located in the inner mitochondrial membrane. Complex I functions in the transfer of electrons from NADH to the respiratory chain.
PDE4C: phosphodiesterase 4C* CpG in CTCF binding site and CpG island in last exon of gene 40.147 kb down stream of TSS AC005759.1: Anti-sense to PDE4C CpG in CTCF binding site and CpG island in last exon of gene 3.323 kb down stream of TSS	19p13.11: position 18208053	-12.6 p=7.2E-08	Type 2 Diabetes edema rosiglitazone	The protein encoded by this gene belongs to the cyclic nucleotide phosphodiesterase (PDE) family, and PDE4 subfamily. This PDE hydrolyzes the second messenger, cAMP, which is a regulator and mediator of a number of cellular responses to extracellular signals. Thus, by regulating the cellular concentration of cAMP, this protein plays a key role in many important physiological processes.
NPBWR2: neuropeptides B/W receptor 2* CpG in promoter and first exon of gene 165 bp downstream of TSS	20q13.3: position 64107006	-5 p=2.7E-05	NI	The protein encoded by this gene is an integral membrane protein and G protein-coupled receptor. The encoded protein is similar in sequence to another G protein-coupled receptor (GPR7), and it is structurally similar to opioid and somatostatin receptors. This protein binds neuropeptides B and W. This gene is intronless and is expressed primarily in the frontal cortex of the brain. See OMM entry 600731
ZNF295-AS1: ZNF295 antisense RNA 1* CpG in 5' internal promoter flank 13.173 kb downstream of TSS	21q22.3: positions 42022367 42022373	+8.5 p=1.15E-09	NI	Long non-coding RNA
Infant Blood Birth Weight				
ZNF678: Zinc finger protein 678 CpG in 5' promoter, CTCF binding site and CpG island 2.616 kb upstream of TSS	10q2.13: position 227560927	-10.6 p=3.04E-07	Body Height, Crohn Disease Crohn's disease Growth Disorders, height, Tobacco Use Disorder	Association with growth disorders and body height. Found to be differentially methylated in response to periconceptual maternal nutrition (17)
TCEANC2: Transcription elongation factor A N-terminal and central domain containing 2* CpG in 5' promoter 851, 844, 842 bp upstream of TSS	1p32.3: position 54054435 54054442 54054446	+6.9 p=1.36E-07	Tobacco Use Disorder	Located in the PARK10 locus which is strongly associated with Parkinson Disease (54)
COL25A1: Collagen type XXV alpha 1 chain* CpG in 5' promoter and CpG island 92 bp upstream of TSS	4q25: position 109302844	+9 p=3.96E-208	Alcoholism, Alzheimer's disease, Blood Pressure Determination, Cholesterol, Heart Rate, Hypertrophy, Left Ventricular, Tobacco Use Disorder	Brain-specific membrane associated collagen. Proteolytic product (CLAC) binds to amyloid beta-peptides found in Alzheimer amyloid plaques but CLAC inhibits rather than facilitates amyloid fibril elongation (55). Over-expression of this collagen in mice, however, found changes in pathology and behaviour suggesting that the encoded protein may promote amyloid plaque formation. Mutation causes congenital fibrosis of extraocular muscles-5 (CFEOM5) possibly because of atypical motor neuron development. See OMM entries 61004 and 616219
TRIO: Trio Rho guanine nucleotide exchange factor* CpG within internal promoter in intron of gene 275,274 kb downstream of TSS	5p15.2: position 14418616	-8.6 p=4.6E-09	Blood Vessels,Breath Tests,coronary spastic angina,Tobacco Use Disorder	This gene encodes a large protein that functions as a GDP to GTP exchange factor. This protein promotes the reorganization of the actin cytoskeleton, thereby playing a role in cell migration and growth. Mutations in this gene have been found in Autosomal Dominant Intellectual Developmental Disorder-44 with Microcephaly (MRD44) and Autosomal Dominant Intellectual Developmental Disorder-63 with Macrocephaly (MRD63). See OMM entries 601893, 617061, 618825
TCERG1L: Transcription elongation regulator 1 like* CpG in 5' promoter and CTCF binding site and CpG island 984 bp upstream of TSS	10q26.3: positions 131312705 131312725	-6.2 p=3.8E-07	ADHD, Alzheimer's disease, Crohn Disease, Diabetes Mellitus, Type 2,Hypertrophy, Left Ventricular, Stroke, Tobacco Use Disorder	NI
JAKMIP3: Janus kinase and microtubule interacting protein 3 CpG in 5' CTCF binding site 13.24 kb upstream from TSS	10q26.3: positions 132023096 132023101	-11.2 p=1.28E-6	Alzheimer's disease, Echocardiography, Stroke	Member of golgin family of proteins. Expressed highest in neural and endocrine tissue. Role in Golgi structure and function (56)
CDHR1: Cadherin related family member 1* CpG in CTCF binding site within last intron of gene 20.01 kb downstream of TSS Cis-eQTM	10q23.1: position 84214547	-28.6 p=5.58E-06	Alzheimer's disease, Cholesterol, retinitis pigmentosa; Leber congenital amaurosis; Usher Syndrome Type I	This gene belongs to the cadherin superfamily of calcium-dependent cell adhesion molecules. The encoded protein is a photoreceptor-specific cadherin that plays a role in outer segment disc morphogenesis. Mutations in this gene have been found in autosomal recessive cone-rod dystrophy (CORD15) and autosomal recessive retinitis pigmentosa (RP65). See OMM entries 609502, 613660,613862
CNNM2: Cyclin and CBS domain divalent metal cation transport mediator 2* CpG within internal promoter flank in intron of gene 103,164 kb downstream of TSS	10q24.32: position 103021458	+14 p=2.28E-09	Alzheimer's disease ,hypertension, Parkinson's Disease, systolic blood pressure, Tobacco Use Disorder	Involved in magnesium homeostasis through epithelial transport and renal reabsorption of Mg ²⁺ . Mutations cause Hypomagnesemia, Seizures and Mental Retardation 1 – HOMGSMR1. Seizure and delayed psychomotor development. Also defective renal reabsorption of Mg(2+). See OMM entries 607803, 616418, 63882
AP001267.2: lincRNA CpG within promoter and CTCF binding site 803 bp downstream of TSS	11: position 118434403	+9.9 p=1.56E-238	NI	NI
TOLLIP: Toll interacting protein* CpG in 5' promoter 613 bp downstream of TSS Cis-eQTM	11p15.5: position 1309041	-8.3 p=6.35E-09	Albumins, antibody response to pertussis vaccination, Apoplexy Cerebral Hemorrhage Cerebral Hemorrhages Intracranial Hemorrhages Stroke Subarachnoid Hemorrhage, Birth Weight Leukemia Leukemia, Myeloid, Acute Precursor Cell Lymphoblastic Leukemia-Lymphoma, dermatitis and eczema, hypertension, Meningeal Neoplasms meningioma	This gene encodes a ubiquitin-binding protein that interacts with several Toll-like receptor (TLR 2 and 4) signaling cascade components in innate leukocytes. The encoded protein regulates inflammatory signaling and is involved in interleukin-1 receptor trafficking and in the turnover of IL1R-associated kinase. It is implicated (from clinical studies and animal models) in inflammatory bowel diseases, pulmonary fibrosis and atherosclerosis (see review 57).

FIGURE 2 (Continued)

LINC00654: Long intergenic non-protein coding RNAs CpG in internal promoter and CTCF binding site within intron 21.993 kb downstream of TSS	20p12.3: position 5504716	-16.4 p=2.81E-07	Alzheimer's disease	LINC00654 is a regulator of TGF-beta induced proliferation and migration of glioma cells via LINC00645/miR205-3p/ZEB1 axis. Association with Alzheimer Disease.
MX2: MX dynamin like GTPase 2* CpG in internal promoter and CTCF binding site 7.695 kb downstream of TSS Trans-meQTLs Cis-meQTLs Cis-eQTM	21q22.3: position 41369694	-12.8 p=3.6E-08	Cataplexy Narcolepsy, Hepatitis C Remission, Spontaneous, Magnesium, melanoma, Type 2 Diabetes edema rosigitazone	21q22.3: Member of dynamin and GTPase gene families. Upregulated by interferon alpha. MX2 is potent inhibitor of HIV-1 infection. See OMM entry 147890
Infant Blood LAZ				
PRDM16: PR/SET domain 16* CpG in promoter flank in first intron of gene 66.722 kb downstream of TSS	1p36.23-p33: 3135890	-10.8 p=4.61E-9	NI	The reciprocal translocation t(1;3)(p36;q21) occurs in a subset of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). This gene is located near the 1p36.3 breakpoint and has been shown to be specifically expressed in the t(1;3)(p36;q21)-positive MDS/AML. The protein encoded by this gene is a zinc finger transcription factor and contains an N-terminal PR domain. The translocation results in the overexpression of a truncated version of this protein that lacks the PR domain, which may play an important role in the pathogenesis of MDS and AML. Mutations in this gene found in the cardiomyopathy left ventricular noncompaction (LVNCB). Mouse gene inactivation model shows role in cleft palate formation. Gene regulates TGF beta and shows expression pattern consistent with cranio-facial development. See OMM entries 605557, 607872, 615373
RUSC1-AS1: RUSC1 antisense RNA 1 CpG in promoter and CTCF binding site in intron 3.299 kb downstream of TSS	1q22: position 155320877	+5.7 p=7.17E-07	NI	RUSC1-AS1 is an antisense transcript of the RUSC1 gene, a signalling adapter in neurotrophic signal transduction. RUSC1 interacts with the nerve growth factor/MAPK signalling cascade to promote neurite outgrowth (58). RUSC1-AS1 contains an open reading frame (function unknown), but may regulate the level of RUSC1 mRNA through anti-sense degradation/repression of the RUSC1 transcript
NFASC: Neurofascin. CpG in 5' internal promoter flank 153.482 kb downstream from TSS	1q32.1: position 204982133	-9 p=4.32E-193	Dengue Hemorrhagic Fever, Erythrocyte Count, Hemoglobins, Iron, Tobacco Use Disorder	Li family immunoglobulin adhesion molecule with role in neurite outgrowth and axon initial segment (AIS) organisation and nodes of Ranvier in early development. Links AIS extra cellular matrix to actin cytoskeleton. Mouse model and mutations in patients show the gene is key in neurodevelopmental disorder – Neurodevelopmental Disorder with Central and Peripheral Motor Dysfunction – NEDCPMD. See OMM entry 609145 (59, 60, 61, 62)
EIF2AK3: Eukaryotic translation initiation factor 2 alpha kinase 3. CpG in 5' promoter 2.382 kb downstream of TSS	2p12: position 88625082	+14.8 p=2.26E-09	Body Height, Longevity, Supranuclear Palsy, Progressive, Tobacco Use Disorder, Type 2 Diabetes edema rosigitazone	Inhibits eukaryotic translation initiation factor 2 by phosphorylation. Induced by ER stress. Reduces protein synthesis. Mutations in Wolcott-Rallison syndrome - neonatal/early infancy insulin-dependent diabetes and later epiphyseal dysplasia, osteoporosis, growth retardation, hepatic and renal dysfunction and mental retardation. See OMM entries 604032, 226980
ABLIM3: Actin binding LIM protein family member 3* CpG in promoter and CTCF binding site 191 bp upstream of TSS	5q31.3 position 149141292	+16.4 p=1.4E-06	Echocardiography, Socioeconomic Factors	This gene encodes a member of the actin-binding LIM (ablIM) family of proteins. These proteins are characterized by an N-terminal LIM domain and a C-terminal dermatin-like domain. The encoded protein interacts with actin filaments and may be a component of adherens junctions in several cell types. A variant of this gene may be associated with pain sensitivity in male human patients. See OMM entry 611305 (63)
ARID1B: AT-rich interaction domain 1B* CpG in transcription factor binding site in last exon 431.577 kb downstream of TSS	6q25.1: position 157207597	-6.6 p=6.92E-233	Angiography, Blood Pressure, Blood Proteins, Body Mass Index, Calcium, Cardiovascular Diseases, Cholesterol, Cholesterol, LDL, Chronic renal failure Kidney Failure, Chronic, Creatinine, Diabetes Mellitus, Exercise Test, Heart Rate, Menopause, Respiratory Function Tests, Schizophrenia, Sleep, Tobacco Use Disorder, Tunica Media, Type 2 Diabetes edema rosigitazone, Waist-Hip Ratio	This locus encodes an AT-rich DNA interacting domain-containing protein. The encoded protein is a component of the SWI/SNF chromatin remodeling complex and may play a role in cell-cycle activation. The protein encoded by this locus is similar to AT-rich interactive domain-containing protein 1A. These two proteins function as alternative, mutually exclusive ARID-subunits of the SWI/SNF complex. The associated complexes play opposing roles. Mutations in this gene have been found in the neurodevelopmental disorder Coffin-Siris syndrome. See OMM entries 614556, 135900
AL354707.1: LncRNA CpG in promoter 273 bp upstream of TSS	9p23: position 6715884	+9.2 p=1.11E-6	NI	
OPCML: Opioid binding protein/cell adhesion molecule like. CpG in internal promoter flank in first intron of gene 303.827 kb downstream from TSS	11q25: position 133228692	+9 p=8.38E-209	Angiography, Blood Pressure, Blood Proteins, Body Mass Index, Calcium, Cardiovascular Diseases, Cholesterol, Cholesterol, LDL, Chronic renal failure Kidney Failure, Chronic, Creatinine, Diabetes Mellitus, Exercise Test, Heart Rate, Menopause, Respiratory Function Tests, Schizophrenia, Sleep, Tobacco Use Disorder, Tunica Media, Type 2 Diabetes edema rosigitazone, Waist-Hip Ratio	Member IgLON subfamily and localized in plasma membrane. GPI-anchored cell adhesion molecule. Believed to have role in opioid receptor function. Mutation analysis shows involved as tumour suppressor gene in Ovarian Cancer. OPCML CpG island is hypermethylated in ovarian cancer. See OMM entries 600632, 167000
CCDC177: Coiled-coil domain containing 177* CpG in 5' promoter and CpG island 1.237 kb downstream of TSS	14q24.1: position 69573634	+12.4 p=3.46E-06	NI	NI
C17orf97: Chromosome 17 open reading frame 97 CpG in promoter 2.857 kb downstream of TSS	17p13.1: position 413182	+10 p=3.87E-07	NI	NI
ARHGAP27P1-BPTFP1-KPNA2P3: readthrough transcribed pseudogene CpG in promoter within intron 3.145 kb downstream of TSS	17: position 64778562	+7.9 p=8.1E-8	NI	NI
RRP7A: Ribosomal RNA processing 7 homolog A. CpG within transcription factor binding site in last exon 8.083 kb downstream of TSS Cis-meQTLs Cis-eQTLs	22q13.2: position 42511713	-14.8 p=3.8E-08	NI	NI
SERHL: serine hydrolase-like (pseudogene)* CpG in transcription factor binding site in intron of gene 11.145 kb downstream of TSS Cis-meQTLs Cis-eQTLs	22q13.2: position 42511713	-14.8 p=3.8E-08	NI	NI

FIGURE 2 (Continued)

domain boundaries and the formation of chromatin loops), binding to promoters, enhancers and within gene bodies, and recruitment of transcription factors. The protein can also act as an insulator, blocking long-range promoter-enhancer interactions for review see (92). Of particular relevance is the observation that methylation at CTCF-binding sites in imprinted regions can disrupt the binding of the CTCF protein and its insulator activity⁹³ and, more generally, at many other methylated sites outside imprinted regions.⁹⁴ From the annotation associated with each of the CpG loci covered by this methyl-seq capture set, almost all the methylation marks described in this study are in regions containing DNase 1 hypersensitive Sites (DHS—markers of DNA regulatory regions and transcriptionally active open chromatin) described by the ENCODE (Encyclopedia of DNA Elements)⁹⁵ project. The ENCODE project has shown that a small proportion (~ 5%) of DHSs are found in TSS (Transcription Start Site) regions, that most are located in introns and intergenic DNA and that there is cell-type specificity in the distribution of DHSs. This indicates that the majority of methylation marks reported in our study are potentially in areas of remodeled open chromatin associated with transcriptional activity and may influence target gene activity possibly by altering chromatin architecture. Figure 2 and 3 also indicate that a number of the DMR-associated genes showing 5% or greater methylation change are subject to trans and/or cis genetic variation (Trans-meQTL and Cis-meQTL) that impacts the level of methylation of closely linked CpG loci; in some cases these methylation changes affect gene expression (Cis-eQTM). One consequence of this polymorphism in the genetic modulation of methylation marks is that it is likely to lead to a diversity of methylation responses to environmental exposures in different populations. Thus interaction between environmental exposures, genetic background and modulation of methylation patterns will have to be assessed for each study population.

Distribution of implicated genes across GO term categories demonstrates that they encompass biochemical and biological functions that include signaling or interaction with signaling pathways; interacting with or acting as receptors; constituents of or interacting with the extracellular matrix; deposition of connective tissue; structure and function of the actin cytoskeleton;

trafficking across cellular membranes; cell cycle control and cellular growth; transcription regulation; and metabolic regulation (see Figure 1). Biological functions revealed by functional studies, animal models, and mutation analysis primarily highlight roles in neurological, growth and developmental, neoplastic, and immunological dysfunction (see figures 2, 3, and 4 and Table S2a–d for details). The precise impact of the methylation changes on the expression of implicated genes, however, is unknown and awaits more detailed functional analysis. Nevertheless, the location of these methylation marks within appropriately positioned regulatory elements and gene bodies or in close intergenic linkage to implicated genes, encourages their consideration as biomarkers associated with and the genetic pathways within which they are active in as potential contributors to variation in prenatal and postnatal growth, subsequent outcomes in later life and as possible intervention targets.

In total, 84 genes implicated by DMRs (shown in red bold and flagged by green and blue asterisks—see Table S2a–d) are shared with DMR-associated genes reported in the large array-based meta-analysis of multiple EWAS by the Childhood Epigenetics Consortium and two further related studies.^{15,86,87} This demonstrates concordance with a substantial proportion (48%) of the genes documented in the current study and provides robust validation of the BiSeq analysis of methyl-seq data. Eleven matched genes are associated with CpG loci surviving stringent Bonferroni correction in the Koppers et al. study¹⁵ (marked with a red asterisk in Table S2a–d). Differences in genetic background, environmental exposures and nutrition between populations contributing to different studies could lead to methylation changes at different CpG loci but still affect DMRs associated with the same implicated genes. In the case of *MAD1L1* and *NFIX*, differential methylation has been detected at the same Bonferroni significant CpG sites that are reported in the meta-analysis¹⁵). *MAD1L1* (a component of the mitotic spindle-assembly checkpoint) has a role in cell cycle control and tumor suppression and methylation levels have been strongly correlated with hepatocellular carcinoma.⁹⁶ It is also a susceptibility gene for bipolar disorder and schizophrenia with a risk allele linked to reward systems in healthy adults.⁹⁷ *NFIX* is most

FIGURE 3 Implicated Genes Associated with Methylation Changes of 5% or Greater in Gene Bodies and Intergenic Regions. This figure documents those genes where a methylation change of 5% or greater has occurred within a gene body or intergenic region. It also provides a summary of function and any disease associations resulting from genome-wide association studies (GWAS—culled from the GAD and EMBL genetic association catalogue databases), mutation analysis and functional investigations. The methylation change is expressed relative to the high birthweight and high length for age groups. All mapping of DMRs is based on human genome build hg38 version GRCh38 of the human genome. The positions of CpGs is given in relation to the Transcription Start Site (TSS) of the implicated gene. Also shown highlighted in red in the first column is whether the gene is associated with Trans and/or Cis-meQTLs and/or Cis-eQTMs. Where plural is shown, this indicates two or more Trans-meQTLs, Cis-meQTLs, or Cis-eQTMs associated with the gene (information obtained from the BIOS QTL Browser at www.genenetwork.nl/biosqtlbrowser). NI, No Information. Color key of gene disease and functional associations: brown, neurological; purple, fertility; light blue, growth and development; dark blue, oncological; light green, immunological; dark green, connective tissue; orange, metabolic; vermillion, cardiovascular; and gray, hearing. The asterisks mark the implicated genes found associated with methylation marks in related studies.^{15,86,87}

highly expressed in brain, fat, and prostate, is linked to cancer (DNA hypermethylation associated with lung adenocarcinoma—LUAD),⁹⁸ muscle development and dystrophies.⁹⁹

Interestingly, 19p13 microduplications encompassing *NFIX* are responsible for intellectual disability, short stature, and small head circumference.¹⁰⁰

Implicated Gene Name Closest to CpG	Location of Methylation Change	Median % Methylation Change and p-value	Disease Associations from GAWAS Studies	Entrez Summary; Phenotype from Functional Studies and Mutation Analysis
BiSeq Analysis >5% Methylation Change				
Trophoblast				
AL109910.1: Long intergenic noncoding RNA CpG intergenic 22.152 kb of TSS	6q27: position 170240417	-10 p=4.2E-17	NI	NI
IKBK: inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta* CpG within intron of gene 44.847 kb downstream from TSS	8p11.2: position 42316149	+5.8 p=3.45E-08	Arthritis, Rheumatoid [Arthritis, Rheumatoid Rheumatoid Arthritis Anti-TNF Response, Asthma Bronchiolitis, Viral Respiratory Syncytial Virus Infections, benzene haematotoxicity Bone Mineral Density, Bronchiolitis, Viral Respiratory Syncytial Virus Infections, Colorectal Cancer, diabetes, type 2,Hepatitis C] Remission, Spontaneous, HIV, Hodgkin Disease Leukemia, Lymphocytic, Chronic, B-Cell Lymphoproliferative Disorders Waldenstrom Macroglobulinemia, Lymphoma, Non-Hodgkin, multiple myeloma, respiratory syncytial virus bronchiolitis, hyroid cancer	The protein encoded by this gene phosphorylates the inhibitor in the inhibitor/NF-kappa-B complex, causing dissociation of the inhibitor and activation of NF-kappa-B. The encoded protein itself is found in a complex of proteins. Mutations in this gene have been found in Immunodeficiency 15A (IMD15A) and 15B (IMD15B). See OMIM entries 603258, 618204, 615592
Cord Blood Birth Weight				
LGR6: Leucine rich repeat containing G protein-coupled receptor* CpG within 3' exon of gene 124.641 kb downstream of TSS	1q32.1: position 202318542	-12.4 P=5.65E-07	Erythrocyte Count	This gene encodes a member of the leucine-rich repeat-containing subgroup of the G protein-coupled 7-transmembrane protein superfamily. The encoded protein is a glycoprotein hormone receptor with a large N-terminal extracellular domain that contains leucine-rich repeats important for the formation of a horseshoe-shaped interaction motif for ligand binding.
CRACD: Cancer-Related Regulator of Actin Dynamics CpG within intron of gene 120.614 downstream of TSS	4q12: position 56169687	-14.8 p=5.77E-08	Erythrocyte Count, Prostatic Neoplasms, Respiratory Function Tests	Cytoskeletal regulator that stabilizes cadherin-catenin-actin complex. Mutated in colorectal cancer. See OMIM entry 618327
SEMA3B: Semaphorin 3B* CpGs in CpG island in 3' exon of gene	3p21.3: positions 50276465 50276472 50276474	+5.1 p=9.76E-07	Lung cancer prostate cancer, Schizophrenia	Belongs to the class-3 semaphorin/collapsin family, whose members function in growth cone guidance during neuronal development. This family member inhibits axonal extension and has been shown to act as a tumor suppressor by inducing apoptosis. See OMIM entry 601261
PHYKPL: 5-phosphohydroxy-L-lysine phospho-lyase CpG within exon of gene 10.339 kb downstream of TSS	5q35.3: position 178222463	-5 p=1.47E-08	Acquired Immunodeficiency Syndrome Disease Progression	This is a nuclear gene encoding a mitochondrial enzyme that catalyzes the conversion of 5-phospho-L-lysine to ammonia, inorganic phosphate, and 2-aminoadipate semialdehyde. Mutations in this gene may cause phosphohydroxylysineuria. See OMIM entries 614683, 615011
ANKS1A: ankyrin repeat and sterile alpha motif domain containing 1A* CpGs within last exon of gene 198.866, 198.869 downstream of TSS	6q21.31: positions 35088121 35088124	-8 p=1.03E-07	Alcoholism, Coronary Artery Disease, height, Lupus Erythematosus, Systemic Systemic lupus erythematosus, Tobacco Use Disorder, Type 2 Diabetes edema rosiglitazone	May have a negative role in growth factor receptor signalling pathways. See OMIM entry 608994
TNXB: tenascin XB* CpG intergenic 35.762 kb upstream of TSS	6q21.3: position 32096415	-18.2 p=1.08E-5	Abortion, Spontaneous, Alzheimer's disease, Arthritis, Rheumatoid Chronic renal failure Kidney Failure, Chronic, Diabetes Mellitus, Type 1, diabetes, type 1, Ehlers-Danlos syndrome,height,HIV-1 control, Lupus Erythematosus, Systemic, nasal polyposis, Schizophrenia, smoking behavior, Type 2 Diabetes edema rosiglitazone	Member of the tenascin family of extracellular matrix glycoproteins. The tenascins have anti-adhesive effects, as opposed to fibronectin which is adhesive. This protein is thought to function in matrix maturation during wound healing. Mutations give rise to Ehlers-Danlos Syndrome (64) and a form of chronic kidney failure, Vesicoureteral Reflux—VUR (65), both of which involve alterations to the ECM related to collagen deposition. See OMIM entries 600985, 606408, 615963
AL078602.1: LncRNA CpG within intron of transcript 163.270 kb downstream of TSS	6q14.1: position 163834847	-8.1 p=3.06E-06	NI	NI
C7orf50: Chromosome 7 open reading frame 50* CpGs in CpG island within intron	7p22.3: positions 1006309 1006312	-10.6 p=1.44E-07	Longevity	Associated with Longevity.
MAD1L1: MAD1 mitotic arrest deficient like 1* CpG in last intron of gene 411.326 downstream of TSS	7p22: position 1821917	-6.9 p=7.64E-06	Bipolar Disorder, Carcinoma, Hepatocellular Liver Neoplasms Neoplasm Recurrence, Local, Chronic renal failure Kidney Failure, Chronic, Iron, longevity, Lung Cancer, Mental Disorders, Myocardial Infarction, Narclepsy, Neutrophils, Schizophrenia, Tobacco Use Disorder	MAD1L1 is a component of the mitotic spindle-assembly checkpoint that prevents the onset of anaphase until all chromosome are properly aligned at the metaphase plate. MAD1L1 functions as a homodimer and interacts with MAD2L1. MAD1L1 may play a role in cell cycle control and tumor suppression. See OMIM entry 602686
LMTK2: Lemur tyrosine kinase 2 CpG in CpG island 311 base pairs 3' to the gene	7q21.3: position 98210819	+5 p=1.06E-06	Body Fat Distribution, Body Weights and Measures, Chronic renal failure Kidney Failure, Chronic, Echocardiography, null, prostate cancer, Prostatic Neoplasms	Belongs to the protein tyrosine kinase family and interacts with several other proteins, such as Inhibitor-2 (Inh2), protein phosphatase-1 (PP1C), p35, and myosin VI. It phosphorylates other proteins, and is itself also phosphorylated when interacting with cyclin-dependent kinase 5 (cdk5)/p35 complex. This protein is involved in nerve growth factor (NGF)-TrkA signalling, and also plays a critical role in endosomal membrane trafficking. Mouse studies suggested an essential role of this protein in spermatogenesis. See OMIM entry 610989
ZNF783: Zinc finger family member 783 CpGs in CpG island 32.8 and 32.817 kb downstream of TSS within intron of gene	7: positions 149294971 149294988	+7.1 p=5.6E-07	NI	NI
ZNF395: Zinc finger protein 395* CpG within intron of gene 49.401 kb downstream of TSS	8p21.1: position 28353300	-15.5 p=7.44E-09		8p21.1: Binding factor for the papilloma virus promoter and the Huntington gene promoter. See OMIM entry 606494

NACC2: NACC family member 2* CpG 5' 96 bp upstream from CTCF binding site in 3'exon of gene <i>Cis-meQTLs</i> <i>Cis-eQTLs</i>	9q34.3: position 136009696	+7.2 p=4.88E-08	Body Fat Distribution, Insulin, Stroke, Tobacco Use Disorder	Acts as a transcriptional repressor, has histone deacetylase activity and coprecipitates with subunits of the nucleosome remodelling and deacetylase (NURD) complex. See OMM entry 615786
XPNPEP1: X-prolyl aminopeptidase 1* CpG within intron of gene 30.583 kb down stream of TSS <i>Trans-meQTL</i> <i>Cis-eQTM</i>	10q25.3: position 109892970	-10.9 p=1.85E-08	Alzheimer's disease, biliary atresia, Tobacco Use Disorder	Cytosolic form of a metalloaminopeptidase that catalyzes the cleavage of the N-terminal amino acid adjacent to a proline residue and may play a role in degradation and maturation of tachykinins, neuropeptides, and peptide hormones. See OMM entry 602443
OR52I1: Olfactory receptor family 52 subfamily I member 1 CpG within exon 600 bp downstream of TSS	11p15.4: position 4594214	-5.4 p=7.63E-07	NI	The olfactory receptor proteins are members of a large family of G-protein-coupled receptors (GPCR) arising from single coding-exon genes. Olfactory receptors share a 7-transmembrane domain structure with many neurotransmitter and hormone receptors and are responsible for the recognition and G protein-mediated transduction of odorant signals.
EPS8L2: EPS8 like 2* CpG in intron 4.532 kb downstream of TSS <i>Cis-meQTL</i> <i>Cis-eQTM</i>	11p15.5: position 698970	-10.6 p=4.84E-07	EPS8L2:Body Weight,Inflammatory Bowel Diseases	EPS8L2 protein, like other members of the family, is thought to link growth factor stimulation to actin organization, generating functional redundancy in the pathways that regulate actin cytoskeletal remodelling. Mutations in humans (DFNB106--Deafness, autosomal recessive) and a mouse knock-out model lead to hearing loss. See OMM entries 614988, 617637
TMEM80: Trans-membrane protein 80 CpG in intron 3.542 kb downstream of TSS <i>Cis-meQTL</i> <i>Cis-eQTM</i>	11p15.5: position 380196	-5 p=5.38E-06	NI	B4GALNT4 modifies N- or O-linked oligosaccharide structures by transferring beta-1,4-linked GalNAc to the terminal acceptors. It is most highly expressed in ovary, followed by fetal brain and various adult brain regions. B4GALNT4 is also highly expressed in foetal kidney and lung. See OMM entry 618560
AL512484.1: LncRNA CpG intergenic 142.139 kb upstream from TSS	13q: position 22736010	-16.6 p=6.7E-6	NI	NI
ADPRHL1: ADP-ribosylhydrolase like 1* CpG within last exon of gene 47.628 kb downstream of TSS <i>Cis-meQTLs</i> <i>Cis-eQTLs</i>	13q34: position 113405896	-6.8 p=3.79E-07	NI	ADP-ribosyltransferases (see ART1: OMM 601625) transfer ADP-ribose from NAD+ to the target protein, and ADP-ribosylhydrolases, such as ADPRHL1, reverse the reaction.
RASA3: RAS p21 protein activator 3* CpG within intron of gene 73.375 kb downstream of TSS <i>Trans-meQTLs</i> <i>Cis-meQTLs</i> <i>Cis-eQTLs</i>	13q34: position 114055314	-13.38 p=1.07E-06	NI	Negative regulator of the RAS signalling pathway.
SERPINA4: Serpin family A member 4 CpG within intron of gene 8.289 kb downstream of TSS	14q31-q32: position 94569731	-14.4 p=3.25E-07		Inhibitors of serine proteases and play and play a role in haemostasis and thrombosis. See OMM entry 147935
AL355102.2: novel protein CpG within intron 47.695 kb downstream of TSS	14: position 96252539	-7 P=1.66E-6	NI	
DLK1: Delta like non-canonical Notch ligand* CpG within 3'UTR 11.855 kb downstream of TSS	14q32: position 100737560	+5.8 p=1.59E-06	Bone Mineral Density, diabetes, type 1, Disease Models, Animal Obesity, null, Optic Disk, type 1 diabetes	This gene encodes a transmembrane protein that contains multiple epidermal growth factor repeats that functions as a regulator of cell growth. The encoded protein is involved in the differentiation of several cell types including adipocytes. This gene is located in a region of chromosome 14 frequently showing uniparental disomy, and is imprinted and expressed from the paternal allele. This imprinted region is important for growth and development and leads to Temple Syndrome when uniparental disomy is present. A single nucleotide variant in this gene is associated with child and adolescent obesity and shows polar overdominance, where heterozygotes carrying an active paternal allele express the phenotype, while mutant homozygotes are normal. See OMM entry 176290
GABRG3: gamma-aminobutyric acid type A receptor gamma3 subunit CpGs within intron of gene 527.896, 527.887 kb downstream of TSS	15q12: positions 27499077 27499086	+7.1 p=1.22E-07	alcohol consumption, alcohol dependence, autism, Bipolar Disorder, bipolar schizoaffective disorder, Blood Pressure Determination, Bulimia, Cholesterol, LDL, Dyskinesia, Drug-Induced], Interleukin-6, null, Parkinson Disease, Psychiatric Disorders, schizophrenia autism, several psychiatric disorders, Tobacco Use Disorder	GABA is the major inhibitory neurotransmitter in the mammalian brain where it acts at GABA-A receptors, which are ligand-gated chloride channels. Chloride conductance of these channels can be modulated by agents such as benzodiazepines that bind to the GABA-A receptor, GABA-A receptors are pentameric, consisting of proteins from several subunit classes: alpha, beta, gamma, delta and rho. The protein encoded by this gene is a gamma subunit, which contains the benzodiazepine binding site.
MARVELD3: MARVEL domain containing 3 CpGs in intron 14.366 kb downstream of TSS	16q22.2: positions 71640523 71640527	-9.9 p=2.04E-07	NI	Protein found colocalised with occludin at junctional complexes. Possible association with resistance to malaria. See OMM entries 6140494, 611162.

FIGURE 3 (Continued)

Three implicated genes match those flagged by methylation changes found in DNA from babies in the Cambridge Baby Growth Study investigating the effects of maternal gestational diabetes or intrauterine growth retardation.⁸⁸

ARID1B and *SEMA3B* are potential tumor suppressor genes. *ARID1B* is a chromatin remodeling factor and individuals with *ARID1B*-related disorder have many phenotypic features including slow growth.¹⁰¹ The third gene is *HOXA10*

MBTPS1: Membrane bound transcription factor peptidase, site 1 CpG within intron 57.788 kb downstream of TSS	16q23.3-q24.1: position 84059154	-5 p=6.64E-07	plasma HDL cholesterol (HDL-C) levels, Type 2 Diabetes edema rosigitazone	This gene encodes a type 1 membrane bound protease which is ubiquitously expressed and regulates cholesterol or lipid homeostasis via cleavage of substrates at non-basic residues. Mutations in this gene may be associated with lysosomal dysfunction. See OMM entries 603355, 618392
PLIN4: Perilipin 4 CpG within 3'UTR of gene 15.910 kb downstream of TSS	19p13.3: position 4502564	-15.8 p=7.54E-09	Obesity	19p13.3: Coats intracellular lipid storage droplets. Associated with obesity.
NFIX: Nuclear factor I X* CpG within intron of gene 92.356 kb downstream of TSS <i>Cis-eQTM</i>	19p13.3: position 13087964	-9.18 p=4.61E-07	NI	The protein encoded by this gene is a transcription factor that binds the palindromic sequence 5-TTGGCNNNNNGCCAA-3 in viral and cellular promoters. The encoded protein can also stimulate adenovirus replication in vitro. Mutations in this gene have been found in Sotos Syndrome 2 and Marshall-Smith Syndrome, both associated with overgrowth, skeletal malformation and impaired neural function. See OMM entries 164005, 614753, 602535
NANOS2: Nanos C2HC-type zinc finger CpG within exon of gene 204 bp downstream of TSS	19q13.32: position 45914574	-19.8 p=1.87E-05	NI	Zinc finger protein expressed in testis. Inactivation of the genes in mouse model causes ablation of germ line and infertility. (66) See OMM entry 608228
BCOR: BCL6 corepressor CpG in CpG island in first intron 74.386 kb downstream of TSS	Xp21-p11.4: position 40102943	+11.6 p=1.25E-05	NI	Interacting corepressor of BCL6, a POZ/zinc finger transcription repressor that is required for germinal center formation and may influence apoptosis. This protein selectively interacts with the POZ domain of BCL6, but not with eight other POZ proteins. Specific class I and II histone deacetylases (HDACs) have been shown to interact with this protein, which suggests a possible link between the two classes of HDACs. Mutations in this gene have been shown to cause developmental syndromic microphthalmia-2 (MCOPS2) and oculofaciocardiodental syndrome (OFCD). See OMM entries 300485 and 300166
Infant Blood Birth Weight				
PYGO2: Pygopus family PHD finger 2* CpG in 3' UTR of gene 5.964 kb downstream of TSS	1q21.3: position 154957889	-6.2 p=8.65E-08	NI	The gene product is required for WNT signal transduction at the level of nuclear beta-catenin in Drosophila. It is suggested that the recruitment of human PYGO permits beta-catenin to transcriptionally activate WNT target genes and that deregulation of this interaction may play a causal role in the development of B-cell malignancies. See OMM entry 606903
CFAP74: Cilia and flagella associated protein 74 CpG within intron 18.37, 18.360 kb downstream of TSS	1p36.33: positions 1985466 1985477	+10 p=4.2E-07	NI	Cilia and flagella associated protein with biased expression in the testis.
KLHL23: Kelch like family member 23 CpG within intron of gene 21.502 kb downstream of TSS	2q31.3: position 169715990	+5.1 p=2.94E-07	NI	NI
CPLEX1: Complexin 1* CpG intergenic 41.512 kb downstream of TSS	4p16.3: position 784617	+12.7 p=2.9E-07	Alcoholism, Behcet Syndrome, Schizophrenia	4p16.3: Cytosolic protein that functions in synaptic vesicle exocytosis. Binds to the SNAP receptor complex and disrupts it, allowing
				transmitter release. Mutations cause Early Infantile Epileptic Encephalopathy - EIEE63. See OMM entry 605032 (67, 68)
TNXB: Tenascin XB* All three CpGs are located within intron of the gene 7.159, 7.146, 9.024 kb downstream of TSS <i>Trans-meQTLs</i> <i>Cis-meQTLs</i> <i>Cis-eQTM</i>	6p21.3 positions 32053494 32053507 position 32069677	+10.1 p=1.07E-08 +5.8 p=3.33E-06	Abortion, Spontaneous, Alzheimer's disease, Arthritis, Rheumatoid Chronic renal failure Kidney Failure, Chronic, Diabetes Mellitus, Type 1 diabetes, type 1, Ehlers-Danlos syndrome,height,HIV-1 control, Lupus Erythematosus, Systemic, nasal polyposis, Schizophrenia, smoking behavior, Type 2 Diabetes edema rosigitazone	Member of the tenascin family of extracellular matrix glycoproteins. The tenascins have anti-adhesive effects, as opposed to fibronectin which is adhesive. This protein is thought to function in matrix maturation during wound healing. Mutations give rise to Ehlers-Danlos Syndrome (64) and a form of chronic kidney failure. Vesicoureteral Reflux -VUR (65), both of which involve alterations to the ECM related to collagen deposition. See OMM entries 600985, 606408, 615963
DMTN: Dematin actin binding protein* CpG within first intron of gene 6.539 kb downstream of TSS	8p22.1: position 22055354	-5.9 p=1.39E-07	NI	The protein encoded by this gene is an actin binding and bundling protein that plays a structural role in erythrocytes, by stabilizing and attaching the spectrin/actin cytoskeleton to the erythrocyte membrane in a phosphorylation-dependent manner. Disruption of this gene in a mouse model causes severe instability of the erythrocyte membrane and hypermethylation metastasis of colorectal cancer (69)
AC016816.1: novel transcript LncRNA CpG intergenic 306.103 kb downstream of TSS and between AC016816.1 and MIR378C	10: positions 130834812, 1300834815	-9.5 p=4.2E-9	NI	NI
MEG9: maternally expressed 9 (LINC00584)* CpG within last exon of gene 4.468 kb downstream of TSS	14q32: position 101072751	+22.6	Body mass and age of menarche (ref)	14q32: Maternally expressed imprinted non-coding RNA (LINC00584). Functional studies indicate role in megakaryocyte differentiation and angiogenesis (70). Located in the imprinted 14q32 region involved in Temple Syndrome
KIF26A: Kinesin family member 26A* CpGs intergenic 32.529 and 32.545 kb 3' to gene	14q32.33: positions 104213423 104213437	+16.7 p=5.77E-08	Pancreatic Neoplasms, Stroke, Waist Circumference	Microtubule associated kinesin. Mouse KO model shows growth retardation with defective development of the bowel. See OMM entry 613231
SGTA: Small glutamine rich tetratricopeptide repeat containing alpha CpG within intron of gene 11.165 kb downstream of TSS <i>Cis-meQTLs</i> <i>Cis-eQTM</i>	19p13.3: position 2772117	+5.7 p=9.42E-07	Insulin Resistance Polycystic Ovary Syndrome, POLYCYSTIC OVARIAN SYNDROME Polycystic Ovary Syndrome	SGTA protein interacts with the growth hormone (69) and steroid hormone (both androgen and progesterone) receptor signalling pathways (72, 73), potentially regulating growth and development of polycystic ovary syndrome, prostate and breast cancers (73). SGTA also promotes the proteasomal degradation of mislocalized proteins and protects Amyloid Precursor Protein from such a fate, possibly implicating it in Alzheimer's Disease (73)
MYT1: myelin transcription factor 1. CpG in CpG island within first intron of gene 31.563 kb downstream of TSS	20q13.33: position 64186999	+5 p=4.47E-06	NI	The protein encoded by this gene is a member of a family of neural specific, zinc finger-containing DNA-binding proteins. The protein binds to the promoter regions of proteolipid proteins of the central nervous system and plays a role in the developing nervous system. Inactivation of the gene in a mouse model is embryonic lethal with poor innervation of the diaphragm. Conditional pancreatic inactivation is associated with a glucose intolerance phenotype. Mutation of MYT1 has been found in a patient with Hemifacial macrosomia (HFM). See OMM entries 600379, 164210

FIGURE 3 (Continued)

(homeobox A10), whose expression is downregulated in endometriosis but in late gestation is required for proper placental differentiation and function.^{100,102}

Implicated genes *ZNF678* (a zinc finger gene), *VTRNA2-1*, *SCRIB*, and *TNXB* have been reported in other Gambian-based studies investigating periconceptual

TAB1: TGF-beta activated kinase 1 (MAP3K7) binding protein 1* CpGs in CpG island 10.653, 10.646, 10.642, 10.636 kb upstream of TSS <i>Trans-meQTLs</i> <i>Cis-meQTLs</i> <i>Cis-eQTM</i>	22q13.1: position 39389125 39389131 39389134 39389141	-35.6 p=3.39E-06	Arthritis, Rheumatoid [Rheumatoid Arthritis, Arthritis, Rheumatoid] [Rheumatoid Arthritis] Anti-TNF Response, Crohn Disease] Crohn's disease	This gene was identified as a regulator of the MAP kinase kinase kinase MAP3K7/TAK1, which is known to mediate various intracellular signaling pathways, such as those induced by TGF beta, interleukin 1, and WNT-1. This protein interacts and thus activates TAK1 kinase. It has been shown that the C-terminal portion of this protein is sufficient for binding and activation of TAK1, while a portion of the N-terminus acts as a dominant-negative inhibitor of TGF beta, suggesting that this protein may function as a mediator between TGF beta receptors and TAK1. This protein can also interact with and activate the mitogen-activated protein kinase 14 (MAPK14/p38alpha), and thus represents an alternative activation pathway, in addition to the MAPKK pathways, which contributes to the biological responses of MAPK14 to various stimuli. See OMIM entry 602615
Infant Blood LAZ				
HLX: H2.0 like homeobox* CpG intergenic 10.8 kb 5' to gene	1q41: position 220890231	+21.3 p=2.27E-07	Aorta, Asthma, Asthma], Blood Pressure, Body Height, Cholesterol, Cholesterol, LDL, Dupuytren Contracture, Echocardiography, Hip, Leukemia, Myeloid, Mortality	Over expression and insertional ablation in transgenic mice indicate a role in CD4+ T lymphocyte development (74) and liver and gut organogenesis (75) respectively. It is interesting to note that the gene has also been found to be expressed in cytotrophoblast cell types in early pregnancy human placentas (76), to be a regulator of trophoblast proliferation (77) and shows reduced expression associated with foetal growth restriction (78).
GAR1: ribonucleoprotein homolog (yeast) pseudogene CpG 781 bp upstream of TSS	4q28: position 184560564	-7.3 p=2.89E-10	NI	NI
IRF4: Interferon regulatory factor 4 CpG intergenic 10.49 kb 5' to gene <i>Trans-meQTL</i> <i>Cis-meQTLs</i> <i>Cis-eQTM</i>	6p25.3: position 390703	-15.3 p=2.7E-07	Abortion, Spontaneous, benzene haematotoxicity, Black vs blond hair color, Black vs red hair color, Carcinoma, Basal Cell, Celiac disease Chromosome Aberrations] Chromosome abnormality] Chronic Lymphocytic Leukemia Leukemia, Lymphocytic, Chronic, B-Cell, Chromosome Aberrations] Chromosome abnormality] Lymphocytosis, Chronic lymphocytic leukemia, Chronic Lymphocytic Leukemia [Hodgkin Disease [Leukemia, Lymphocytic, Chronic, B-Cell] Multiple Myeloma, Chronic Lymphocytic Leukemia] Leukemia, Lymphocytic, Chronic, B-Cell, Chronic Lymphocytic Leukemia] Leukemia, Lymphocytic, Chronic, B-Cell [Lymphoma] Syndrome, Chronic renal failure] Kidney Failure, Chronic, Eye Color, freckles, Hair Color, HIV, Hodgkin Disease [Leukemia, Lymphocytic, Chronic, B-Cell] Lymphoproliferative Disorders] Waldenstrom Macroglobulinemia, Leukemia, Leukemia, Lymphocytic, Chronic, B-Cell, Leukemia, Lymphoid, lymphoma, melanoma, melanoma] Nevus] Skin Neoplasms, melanoma] Nevus] Skin Neoplasms] Sunburn, Melanosis, multiple myeloma, Multiple Sclerosis, Neuroblastoma, Skin Neoplasms, Suntan, Supranuclear Palsy, Progressive, tanning phenotype	Belongs to the IRF (interferon regulatory factor) family of transcription factors, characterized by a unique tryptophan pentad repeat DNA-binding domain. The IRFs are important in the regulation of interferons in response to infection by virus, and in the regulation of interferon-inducible genes. This family member is lymphocyte specific and negatively regulates Toll-like-receptor (TLR) signaling that is central to the activation of innate and adaptive immune systems. A chromosomal translocation involving this gene and the IgH locus, t(6;14)(p25;q32), may be a cause of multiple myeloma. See OMIM entries 601900, 612558
ZFX3: Zinc finger* homeobox 3 CpG within intron of gene 858.287 kb downstream of TSS	16q22.2-q22.3 position 73033584	-14.8 p=6.9E-08	Alcoholism, Atrial Fibrillation, Atrial Fibrillation] Atrial Fibrillation] Brain Ischemia] Stroke, Body Mass Index, Cardiovascular Diseases, Coronary Disease, Kawasaki disease ,Mucocutaneous Lymph Node Syndrome, Myocardial Infarction] Ventricular Fibrillation, prostate cancer, Tobacco Use Disorder, Waist Circumference	ZFX3 (a zinc finger homeobox transcription factor - also known as <i>ATFB1</i> , Atrial fibrillation, family 1) can both repress and activate genes in cooperation with TGF beta signaling (79, 80, 81). Genetic evidence implicates the gene in a number of disease phenotypes; breast cancer through mutation and interaction with the oestrogen receptor (82, 83), gastric cancer (80) and prostate cancer (82). The important role the gene plays in myogenic differentiation (83) correlates with the genetic association to atrial fibrillation and other cardiac phenotypes. See OMIM entries 104155, 613055
LINC00511: long intergenic non-protein coding RNA 511. CpG in intron of gene 260,980 kb downstream of TSS	17q24.3: position 72379492	-11 p=3.3E-09	NI	Long intergenic non-coding RNA involved in promoting several cancers through regulation of a variety of microRNAs. Acts as an oncogenic LINC RNA (84)

FIGURE 3 (Continued)

nutritional exposures associated with differential methylation.^{11,22,26} *VTRNA2a-1* is a noncoding RNA gene that functions as a tumor suppressor⁴⁷⁻⁵⁰ and is an imprinted locus.^{51,52} *SCRIB* (a scaffold protein found at epithelial adherens junctions and neuronal presynaptic compartments) can act as a tumor suppressor gene and has been shown to be mutated in severe neural tube defects (see OMIM entry 607733). *TNXB* is an extracellular matrix glycoprotein thought to function in matrix maturation during wound healing. Different pathogenic alleles give rise to Ehlers–Danlos Syndrome⁶⁴ and a form of chronic kidney failure, Vesicoureteral Reflux –VUR,⁶⁵ both of which involve alterations to collagen deposition in the extracellular matrix.

The genes *DLK1* and *MEG9* (LINC00584—long intergenic noncoding RNA) are worthy of further comment given their location within an important imprinted region on chromosome 14 at 14q32. As revealed by maternal and paternal Uniparental Disomy (UPDm and UDPp, respectively), and genetic and functional studies of individual genes encompassed within the locus (for review see OMIM entries 601038, 60563, 611896,

172690, 613648, and ref (103), the region has a major impact on growth and development. The 14q23 locus is complex with a cluster of maternally and paternally imprinted genes, non-coding snoRNAs (small nucleolar organizer RNA), miRNAs (microRNAs), LncRNAs (long noncoding RNAs), and LINC RNAs under the control of an intergenic differentially methylated region (IG-DMR).¹⁰⁴ Three genes (*DLK1*, *RTL1*, and *DIO3*) are all expressed from paternal alleles. *DLK1*, containing six epidermal growth factor repeats, has reduced plasma levels in women bearing small for gestational age babies,¹⁰⁵ is an inhibitor of adipocyte differentiation¹⁰⁶ and shows genetic association with age of menarche.^{107,107,108,109,110} *RTL1* is essential for maintenance of fetal capillaries and potentially involved in formation of the chorioallantoic placenta,¹¹¹ while *DIO3* (Thyroxine Deiodinase Type III) is essential for the maturation and function of the thyroid axis.¹¹² A further four genes (*MEG3*, *RTL1as*, *MEG8*, and *MEG9*) are all expressed from maternal alleles. *MEG3* is a LncRNA affecting growth and development in *MEG3* knock-out mice¹¹³; *RTL1as* is an antisense transcript to the paternally expressed gene *RTL1* and encodes a number of microRNAs that may regulate the expression of

DMR Associated Genes	Neur	Repro	Growth and Dev	Onco	Imm	ConnTiss	Cardio-vasc	Vision	Hearing	Metabolic
Trophoblast	2				1					
Cord Blood Birthweight	8	3	4	4	1	1			1	3
Infant Blood Birthweight	7		4	3	2			1		
Infant Blood Height for Age	9		3	3	1		1			
Total	20	3	11	10	5	1	1	1	1	3

FIGURE 4 Number of Implicated Genes from Figures 2 and 3 Associated with Different Disease Categories. The color key allows cross-reference to the gene lists in Figures 2 and 3. Neur, Neurological; Repro, Reproductive; Growth and Dev, Growth and Development; Onco, Oncological; Imm, Immunological; Conn Tiss, Connective Tissue; and Cardio-vasc, Cardio-vascular

TABLE 5 Methylation Analysis of the DMR Associated with the *ZFH3* Gene by Pyrosequencing

<i>ZFH3</i> (% methylation)							
Pyrosequencing				Methyl-seq			
Sample	High	Sample	Low	Sample	High	Sample	Low
1	20	1	31	1	22	1	40
2	22	2	31	2	22	2	42
3	23	3	31	3	23	3	42
4	23	4	33	4	23	4	41
5	30	5	35	5	24	5	40
		6	35			6	46
Mean:							
23.6 sd+/-3.38		32.6 sd+/-1.79		22.8 sd+/- 0.74		41.3 sd+/- 1.1	

Individual samples sourced from the 12-month high and low LAZ (length for age) comparison groups.

Median methyl-seq determined methylation change between groups = -14.8 referenced to high LAZ value.

This table compares the methylation levels determine for the DMR associated with the *ZFH3* gene by pyrosequencing and methyl-seq analysis. Individuals from the high (*n* = 5) and low (*n* = 6) 12-month length for age comparison groups were selected for analysis. The table shows the mean % methylation values and standard deviation for the two types of analysis. The quantum and direction of change is close to that observed for median methylation change from the comparison of the high and low groups determined by BiSeq analysis of methyl-seq data.

*RTL1*¹⁰⁵; *MEG8* is a LncRNA involved in the regulation of trophoblast proliferation and invasion, and implicated in spontaneous early abortion¹¹⁴ and *MEG9*, a LINC RNA involved in megakaryocyte differentiation and angiogenesis⁷⁰ shows genetic association with body mass index and age of menarche.¹⁰⁷

The UPDm (no paternal transcripts: Temple Syndrome) phenotype is characterized by prenatal and postnatal growth retardation, neonatal hypotonia, precocious puberty, and facial dysmorphism. The UPDp (no maternal transcripts: Kagami-Ogata Syndrome) phenotype is characterized by severe growth retardation, skeletal abnormalities, facial anomalies, and abdominal muscular defects. Trans-regulation by maternally expressed small noncoding

RNAs from the 14q32 region on the activity of other genes in the genome is likely to contribute to these complex phenotypes.¹⁰⁵ On the maternal chromosome *DLK1* is silenced. The current study shows a 6% methylation difference of a *DLK1* DMR (higher in high birthweight than low birthweight babies). In contrast, *MEG9* is silenced on the paternal chromosome and shows a 22% methylation difference of a *MEG9* DMR (higher in high birthweight than in low birthweight babies). It is not clear what the impact of these methylation marks is on expression levels as they lie outside the immediate promoter within the gene body. Nevertheless, given that both methylation marks are in DMRs containing DHSs marking potentially open chromatin, it is reasonable to suggest that alteration of

the methylation landscape in this region of chromosome 14 could impact chromatin architecture and gene activity with a bearing on growth and development outcomes. It is interesting to note that a study examining the effect of maternal periconceptional micronutrient supplementation of Gambian mothers found increased methylation of a *DLKI*-associated CpG in cord blood DNA from offspring of mothers who had received the supplements.²⁴

A number of limitations should be noted. An accessible tissue such as blood as a proxy for methylation changes in other key target tissues will not capture all the relevant alterations in methylation status. However, there is sufficient concordance between tissues to yield a subset of potentially relevant loci.¹¹⁵⁻¹¹⁷ Analysis has been performed with males and females combined; hence sex differences in the methylation patterns have not been determined. The sample size is small, nevertheless, as outlined in the methods, BiSeq is designed for the analysis of targeted methyl sequence data and takes advantage of the conservation of methylation across short distances, co-assessing methylation changes at several individual cytosine residues within intervals of 100 base pairs. This reduces data dimensions and increases detection power by borrowing nearby CpG information and provides a more detailed and statistically significant evaluation of the methylation status across any given genomic region; this has allowed identification of statistically valid differentially methylated CpGs from this small study. Greater coverage (3.7 million CpGs as compared to the Illumina HM450 k and Epic 850 k chips) of the SureSelect targeted sequencing approach of key gene regulatory elements (adjacent and distant, proximal or distal) to genes they control, offers the opportunity to identify additional methylation marks not necessarily scored by the array-based platforms.

Studies, such as the one reported here, provide associations and not cause and effect relationships between genes and phenotypes. Mutational evidence is helpful in establishing the likelihood that a gene contributes to a complex phenotype. Identification of methylation marks can be useful in that (a) they might act as biomarkers of early life adverse exposures that impact on early growth and may potentially indicate those individuals with higher future disease risks and (b) potentially flag genes that may be useful intervention targets to ameliorate the consequences of stunting. An integrated large scale analysis of inter-individual variation of methylation marks in relation to genotype (Trans and Cis-meQTLs), eQTLs (expression quantitative trait loci including Cis-eQTLs), disease susceptibility, developmental phenotypes, nutrition, and environmental exposures provides a means of potentially unpicking causal relationships and the relevance of implicated genes. Clearly, the most effective approach to mitigate stunting and associated disease susceptibilities would be to ensure healthy nutrition, adequate sanitation, and living conditions early in the life course.

ACKNOWLEDGMENTS

The authors thank the families of West Kiang who patiently participated in this study. The authors acknowledge the enthusiastic work of the whole HERO-G Working Group, especially the fieldworkers, village assistants, midwives, clinical staff, data office staff, and laboratory technicians who tirelessly collected the data and samples. The raw data supporting the conclusions of this article are stored on the Open Science Framework (OSF), <https://doi.org/10.17605/OSF.IO/5ND3Y>, and at the time of article submission are available on request and subject to review. These data will be made publicly available no later than 1 July 2021. Requests to access the datasets should be directed to the corresponding author.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

CRQ developed and managed the experimental procedures and KMH performed them; MS and MG devised the initial analysis pipeline and JB and BS finalized the pipelines and provided the bioinformatic support; CRQ performed the methylation data analysis; CAS and NAA provided supervisory support; AMD and SD managed the collection and extraction of DNA from samples; CRQ and NAA wrote the first draft of the manuscript; RMB, NA, DBD, KKO, AMP, and SEM conceived of and designed the HERO-G study. All authors contributed to manuscript revision, read, and approved the submitted version.

ORCID

Robin M. Bernstein  <https://orcid.org/0000-0001-8805-7527>

REFERENCES

1. Black RE, Victora CG, Walker SP, et al. (2013) Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet*. 382, 427-451. Erratum in (2013): *Lancet*. 382, 396.
2. Moore SE. Early life nutritional programming of health and disease in The Gambia. *J Dev Orig Health Dis*. 2016;7:123-131.
3. Olofin I, McDonald CM, Ezzati M, et al. Associations of sub-optimal growth with all-cause and cause-specific mortality in children under five years: a pooled analysis of ten prospective studies. *PLoS One*. 2013;8:e64636.
4. Ong KK, Hardy R, Shah I, Kuh D. Childhood stunting and mortality between 36 and 64 years: the British 1946 Birth Cohort Study. National Survey of Health and Development Scientific and Data Collection Teams. *J Clin Endocrinol Metab*. 2013;98(5):2070-2077.
5. UNICEF. WHO. World Bank Levels and trends in child malnutrition, UNICEF–WHO–World Bank joint child malnutrition estimates. (2015). www.who.int/entity/nutrition/publications/joint_childmalnutrition_2015_estimates/en/. Accessed February 16, 2020.

6. Christian P, Lee SE, Angel MD, et al. Risk of childhood undernutrition related to small-for-gestational age and preterm birth in low- and middle-income countries. *Int J Epidemiol*. 2013;42:1340-1355.
7. Gluckman PD. Epigenetics and metabolism in 2011: Epigenetics, the life-course and metabolic disease. *Nat Rev Endocrinol*. 2011;8:74-76.
8. Barker DJP. Developmental origins of adult health and disease. *J Epidemiol Community Health*. 2004;58:114-115.
9. Gluckman PD, Hanson MA, Spencer HG, Bateson P. Environmental influences during development and their later consequences for health and disease: implications for the interpretation of empirical studies. *Proc Biol Sci*. 2005;272:671-677.
10. Stein AD, Wang M, Ramirez-Zea M, et al. Exposure to a nutrition supplementation intervention in early childhood and risk factors for cardiovascular disease in adulthood: Evidence from Guatemala. *Am J Epidemiol*. 2006;164:1160-1170.
11. Adu-Afarwah S, Young RT, Lartey A, et al. Randomized comparison of 3 types of micronutrient supplements for home fortification of complementary foods in Ghana: Effects on growth and motor development. *Am J Clin Nutr*. 2007;86:412-420.
12. Bhutta ZA, Ahmed T, Black RE, et al. What works? Interventions for maternal and child undernutrition and survival. *Lancet*. 2008;371:417-440.
13. Fleming TP, Watkins AJ, Velazquez MA, et al. Origins of lifetime health around the time of conception: causes and consequences. *Lancet*. 2018;391(10132):1842-1852.
14. Lillycrop KA, Burdge GC. Epigenetic mechanisms linking early nutrition to long term health. *Best Pract Res Clin Endocrinol Metab*. 2012;26:667-676.
15. Küpers LK, Monnereau C, Sharp GC, et al. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birth-weight. *Nat Commun*. 2019;10:1893.
16. Schoenbuchner SM, Dolan C, Mwangome M, et al. The relationship between wasting and stunting: a retrospective cohort analysis of longitudinal data in Gambian children from 1976 to 2016. *Am J Clin Nutr*. 2019;110(2):498-507.
17. Nabwera HM, Bernstein RM, Agbla SC, et al. Hormonal Correlates and Predictors of Nutritional Recovery in Malnourished African Children. *J Trop Pediatr*. 2018;64:364-372.
18. Rayco-Solon P, Fulford AJ, Prentice AM. Differential effects of seasonality on preterm birth and intrauterine growth restriction in rural Africans. *Am J Clin Nutr*. 2005;81(1):134-139.
19. Waterland RA, Kellermayer R, Laritsky E, et al. Season of conception in rural Gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet*. 2010;6:e1001252.
20. Dominguez-Salas P, Moore SE, Cole D, et al. DNA methylation potential: dietary intake and blood concentrations of one-carbon metabolites and cofactors in rural African women. *Am J Clin Nutr*. 2013;97:1217-1227.
21. Dominguez-Salas P, Moore SE, Baker MS, et al. Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat Commun*. 2014;5:3746.
22. Silver MJ, Kessler NJ, Hennig BJ, et al. Independent genome-wide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. *Genome Biol*. 2015;16(1):118.
23. Kessler NJ, Waterland RA, Prentice AM, Silver MJ. Establishment of environmentally sensitive DNA methylation states in the very early human embryo. *Sci Adv*. 2018;4(7):eaat2624.
24. Khulan B, Cooper WN, Skinner BM, et al. Periconceptional maternal micronutrient supplementation is associated with widespread gender related changes in the epigenome: A study of a unique resource in the Gambia. *Hum Mol Genet*. 2012;21:2086-2101.
25. Cooper WN, Khulan B, Owens S, et al. DNA methylation profiling at imprinted loci after periconceptional micronutrient supplementation in humans: results of a pilot randomized controlled trial. *FASEB J*. 2012;26:1782-1790.
26. Hernandez-Vargas H, Castelino J, Silver MJ, et al. Exposure to aflatoxin B1 in utero is associated with DNA methylation in white blood cells of infants in The Gambia. *Int J Epidemiol*. 2015;44(4):1238-1248.
27. Moore SE, Doel AM, Ong KK, et al. Identification of nutritionally modifiable hormonal and epigenetic drivers of positive and negative growth deviance in rural African fetuses and infants: Project protocol and cohort description [version 1; peer review: awaiting peer review]. *Gates Open Res*. 2020;2020(4):25.
28. Andrews S. FastQC A Quality Control Tool for High Throughput Sequence Data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
29. Krueger F, Kreck B, Franke A, Andrews SR. DNA methylome analysis using short bisulfite sequencing data. *Nat Methods*. 2012;9:145-151.
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357-359.
31. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27:1571-1572.
32. Klein HU, Hebestreit K. An evaluation of methods to test predefined genomic regions for differential methylation in bisulfite sequencing data. *Brief Bioinform*. 2016;17:796-807.
33. R Core Team. R. *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018. <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing>
34. Yu X, Sun S. Comparing five statistical methods of differential methylation identification using bisulfite sequencing data. *Stat Appl Genet Mol Biol*. 2016;15:173-191.
35. Benjamini Y, Heller R. False discovery rates for spatial signals. *J Am Stat Assoc*. 2007;102:1271-1281.
36. Lövkvist C, Dodd IB, Sneppen K, Haerter JO. DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res*. 2016;44:5123-5132.
37. Affinito O, Palumbo D, Fierro A, et al. Nucleotide distance influences co-methylation between nearby CpG sites. *Genomics*. 2020;112:144-150.
38. Salas LA, Koestler DC, Butler RA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. *Genome Biol*. 2018;19:64. <https://doi.org/10.1186/s13059-018-1448-7>
39. Reinius LE, Acevedo N, Joerink M, et al. Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. *PLoS One*. 2012;7:e41361.

40. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4:44-57.
41. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with PANTHER Classification System. *Nat Protoc.* 2013;8:1551-1566.
42. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucl Acids Res.* 2019. <https://doi.org/10.1093/nar/gky1038>
43. Morin AM, Gatev E, McEwen LM, et al. Maternal blood contamination of collected cord blood can be identified using DNA methylation at three CpGs. *Clin Epigenetics.* 2017;9:75.
44. Helbig I, Lopez-Hernandez T, Shor O, et al. A Recurrent Missense Variant in AP2M1 Impairs Clathrin-Mediated Endocytosis and Causes Developmental and Epileptic Encephalopathy. *Am J Hum Genet.* 2019;104:1060-1072.
45. Tarpey P, Parnau J, Blow M, et al. Mutations in the DLG3 gene cause nonsyndromic X-linked mental retardation. *Am J Hum Genet.* 2004;75:318-324.
46. Phillips AK, Sirén A, Avela K, et al. X-exome sequencing in Finnish families with intellectual disability—four novel mutations and two novel syndromic phenotypes. *Orphanet J Rare Dis.* 2014;9:49.
47. Lee K, Kunkeaw N, Jeon SH, et al. Precursor miR-886, a Novel Noncoding RNA Repressed in Cancer, Associates With PKR and Modulates Its Activity. *RNA.* 2011;17(1076-1089):14.
48. Treppendahl MB, Qiu X, Sogaard A, et al. Allelic methylation levels of the noncoding VTRNA2-1 located on chromosome 5q31.1 predict outcome in AML. *Blood.* 2012;119:206-216.
49. Cao J, Song Y, Bi N, et al. DNA methylation-mediated repression of miR-886-3p predicts poor outcome of human small cell lung cancer. *Cancer Res.* 2013;73:3326-3335.
50. Lee HS, Lee K, Jang HJ, et al. Epigenetic silencing of the non-coding RNA nc886 provokes oncogenes during human esophageal tumorigenesis. *Oncotarget.* 2014;5:3472-3481.
51. Paliwal A, Temkin AM, Kerker K, et al. Comparative anatomy of chromosomal domains with imprinted and non-imprinted allele-specific DNA methylation. *PLoS Genet.* 2013;9: <https://doi.org/10.1371/journal.pgen.1003622>
52. Romanelli V, Nakabayashi K, Vizoso M, et al. Variable maternal methylation overlapping the nc886/vtRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. *Epigenetics.* 2014;9:783-790.
53. Ptáček R, Kuzelová H, Stefano GB. Dopamine D4 receptor gene DRD4 and its association with psychiatric disorders. *Med Sci Monit.* 2011;17:RA215-20.
54. Beecham GW, Dickson DW, Scott WK, et al. PARK10 is a major locus for sporadic neuropathologically confirmed Parkinson disease. *Neurology.* 2015;84:972-980.
55. Osada Y, Hashimoto T, Nishimura A, Matsuo Y, Wakabayashi T, Iwatsubo T. CLAC binds to amyloid beta peptides through the positively charged amino acid cluster within collagenous domain 1 and inhibits formation of amyloid fibrils. *J Biol Chem.* 2005; 280: 8596-8605. Note: Erratum: *J. Biol. Chem.* 280, 15484.
56. Cruz-Garcia D, Vazquez-Martinez R, Peinado JR, et al. Identification and characterization of two novel (neuro)endocrine long coiled-coil proteins. *FEBS Lett.* 2007;581:3149-3156.
57. Kowalski EJA, Li L. Toll-Interacting Protein in Resolving and Non-Resolving Inflammation. *Front Immunol.* 2017;8:511.
58. MacDonald JI, Kubu CJ, Meakin SO. Nesca, a novel adapter, translocates to the nuclear envelope and regulates neurotrophin-induced neurite outgrowth. *J Cell Biol.* 2004;164:851-862.
59. Anazi S, Maddirevula S, Salpietro V, et al. Expanding the genetic heterogeneity of intellectual disability. *Hum. Genet.* 136: 1419-1429, Note: Erratum (2018): *Hum. Genet.* 2017;137:105-109.
60. Smigiel R, Sherman DL, Rydzanicz M, et al. Homozygous mutation in the neurofascin gene affecting the glial form of neurofascin causes severe neurodevelopment disorder with hypotonia, amimia and areflexia. *Hum Molec Genet.* 2018;27:3669-3674.
61. Monfrini E, Straniero L, Bonato S, et al. Neurofascin (NFASC) gene mutation causes autosomal recessive ataxia with demyelinating neuropathy. *Parkinsonism Relat. Disord.* 2019;63:66-72.
62. Pillai AM, Thaxton C, Pribisko AL, Cheng J-G, Dupree JL, Bhat MA. Spatiotemporal ablation of myelinating glia-specific neurofascin (Nfasc-NF155) in mice reveals gradual loss of paranodal axoglial junctions and concomitant disorganization of axonal domains. *J Neurosci Res.* 2009;87:1773-1793.
63. Krupp M, Weinmann A, Galle PR, Teufel A. Actin binding LIM protein 3 (abLIM3). *Int J Mol Med.* 2006;17(1):129-133.
64. Burch GH, Gong Y, Liu W, et al. Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. *Nat Genet.* 1997;17:104-108.
65. Gbadegesin RA, Brophy PD, Adeyemo A, et al. TNXB mutations can cause vesicoureteral reflux. *J Am Soc Nephrol.* 2013;24:1313-1322.
66. Saba R, Kato Y, Saga Y. Nanos2 promotes male germ cell development independent of meiosis suppression. *Dev Biol.* 2014;385:32-40.
67. Karaca E, Harel T, Pehlivan D, et al. Genes that affect brain structure and function identified by rare variant analyses of mendelian neurologic disease. *Neuron.* 2015;88:499-513.
68. Redler S, Strom TM, Wieland T, et al. Variants in CPLX1 in two families with autosomal-recessive severe infantile myoclonic epilepsy and ID. *Europ J Hum Genet.* 2017;25:889-893.
69. Ye Y-P, Jiao H-L, Wang S-Y, et al. Hypermethylation of DMTN promotes the metastasis of colorectal cancer cells by regulating the actin cytoskeleton through Rac1 signaling activation. *J Exp Clin Cancer Res.* 2018;37:299.
70. Espinosa Diez C, Wilson R, Mukherjee R, et al. DNA damage dependent hypomethylation regulates the pro-angiogenic lncRNA MEG9. *BioRxiv.* 2018. the preprint server for biology.
71. Roberts JD, Thapaliya A, Martínez-Lumbreras S, Krysztowska EM, Isaacson RL. Structural and Functional Insights into Small, Glutamine-Rich, Tetratricopeptide Repeat Protein Alpha. *Front Mol Biosci.* 2015;2:71.
72. Paul A, Garcia YA, Zierer B, et al. The cochaperone SGTA (small glutamine-rich tetratricopeptide repeat-containing protein alpha) demonstrates regulatory specificity for the androgen, glucocorticoid, and progesterone receptors. *J Biol Chem.* 2014;289:15297-15308.
73. Leznicki P, High S. SGTA antagonizes BAG6-mediated protein triage. *Proc Natl Acad Sci U S A.* 2012;109:19214-19219.
74. Deguchi Y, Agus D, Kehrl JH. A human homeobox gene, HB24, inhibits development of CD4+ T cells and impairs thymic involution in transgenic mice. *J Biol Chem.* 1993;268:3646-3653.
75. Hentsch B, Lyons I, Li R, et al. Hlx homeobox gene is essential for an inductive tissue interaction that drives expansion of embryonic liver and gut. *Genes Dev.* 1996;10:70-79.

76. Rajaraman G, Murthi P, Quinn L, Brennecke SP, Kalionis B. Homeodomain protein HLX is expressed primarily in cytotrophoblast cell types in the early pregnancy human placenta. *Reprod Fertil Dev*. 2008;20(3):357-367.
77. Rajaraman G, Murthi P, Leo B, Brennecke SP, Kalionis B. Homeobox gene HLX1 is a regulator of colony stimulating factor-1 dependent trophoblast cell proliferation. *Placenta*. 2007;28(10):991-998.
78. Murthi P, Doherty V, Said J, Donath S, Brennecke SP, Kalionis B. Homeobox gene HLX1 expression is decreased in idiopathic human fetal growth restriction. *Am J Pathol*. 2006;168(2):511-518.
79. Sakata N, Kaneko S, Ikeno S, et al. TGF- β Signaling Cooperates with AT Motif-Binding Factor-1 for Repression of the α -Fetoprotein Promoter. *J Signal Transduct*. 2014;2014:970346.
80. Mori Y, Kataoka H, Miura Y, et al. Subcellular localization of ATBF1 regulates MUC5AC transcription in gastric cancer. *Int J Cancer*. 2007;121:241-247.
81. Berry FB, Miura Y, Mihara K, et al. Positive and negative regulation of myogenic differentiation of C2C12 cells by isoforms of the multiple homeodomain zinc finger transcription factor ATBF1. *J Biol Chem*. 2001;276:25057-25065.
82. Dong XY, Sun X, Guo P, et al. ATBF1 inhibits estrogen receptor (ER) function by selectively competing with AIB1 for binding to the ER in ER-positive breast cancer cells. *J Biol Chem*. 2010;285:32801-32809.
83. Sun X, Frierson HF, Chen C, et al. Frequent somatic mutations of the transcription factor ATBF1 in human prostate cancer. *Nat Genet*. 2005;37(4):407-412. Erratum. In: *Nat Genet*. 37, 652. Cantarel, Brandi M [corrected to Cantarel, Brandi L].
84. Yu C-L, Xu X-L, Yuan F. LINC00511 is associated with the malignant status and promotes cell proliferation and motility in cervical cancer. *Biosci Rep*. 2019;39:BSR20190903.
85. Bonder MJ, Luijk R, Zhernakova DV. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet*. 2017;49:131-138.
86. Merid SK, Novoloaca A, Sharp GC, et al. Epigenome-wide meta-analysis of blood DNA methylation in newborns and children identifies numerous loci related to gestational age. *Genome Med*. 2020;12:25.
87. Hannon E, Schendel D, Ladd-Acosta C, et al. Variable DNA methylation in neonates mediates the association between prenatal smoking and birth weight. *Philos Trans R Soc Lond B Biol Sci*. 2019;374(1770):20180120.
88. Quilter CR, Cooper WN, Cliffe KM, et al. Impact on offspring methylation patterns of maternal gestational diabetes mellitus and intrauterine growth restraint suggest common genes and pathways linked to subsequent type 2 diabetes risk. *FASEB J*. 2014;28(11):4868-4879.
89. Bakulski KM, Feinberg JL, Andrews SV, et al. DNA methylation of cord blood cell types: Applications for mixed cell birth studies. *Epigenetics*. 2016;11:354-362.
90. Karlberg J, Fryer JG, Engström I, Karlberg P. Analysis of linear growth using a mathematical model. II. From 3 to 21 years of age. *Acta Paediatr Scand Suppl*. 1987;337:12-29.
91. Bernstein RM, O'Connor GK, Vance EA, et al. Timing of the infancy-childhood transition in rural Gambia. *Frontiers in Endocrinology*. 2020. <https://doi.org/10.3389/fendo.2020.00142>
92. Holwerda SJB, de Laat W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci*. 2013;368:20120369.
93. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature*. 2000;405:482-485.
94. Wang H, Maurano MT, Qu H, et al. Widespread Plasticity in CTCF Occupancy Linked to DNA Methylation. *Genome Res*. 2012;22:1680-1688.
95. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57-74.
96. Cui C, Lu Z, Yang L, et al. Genome-wide identification of differential methylation between primary and recurrent hepatocellular carcinomas. *Mol Carcinog*. 2016;55:1163-1174.
97. Trost S, Diekhof EK, Mohr H, et al. Investigating the Impact of a Genome-Wide Supported Bipolar Risk Variant of MAD1L1 on the Human Reward System. *Neuropsychopharmacology*. 2016;41:2679-2687.
98. Ge J, Dong H, Yang Y, et al. NFIX downregulation independently predicts poor prognosis in lung adenocarcinoma, but not in squamous cell carcinoma. *Future Oncol*. 2018;14(30):3135-3144.
99. Piper M, Gronostajski R, Messina G. Nuclear Factor One X in Development and Disease. *Trends Cell Biol*. 2019;29(1):20-30.
100. Trimouille A, Houcinat N, Vuillaume ML, et al. 19p13 microduplications encompassing NFIX are responsible for intellectual disability, short stature and small head circumference. *Eur J Hum Genet*. 2018;26:85-93.
101. Vergano SA, van der Sluijs PJ, Santen G. (2019) ARID1B-Related Disorder. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A, editors. *GeneReviews® [Internet]*. Seattle (WA): University of Washington, Seattle; 1993-2020. <http://www.ncbi.nlm.nih.gov/books/NBK541502/>
102. Özcan C, Özdamar Ö, Gökbayrak ME, Doğer E, Çakıroğlu Y, Çine N. HOXA-10 gene expression in ectopic and eutopic endometrium tissues: Does it differ between fertile and infertile women with endometriosis? *Eur J Obstet Gynecol Reprod Biol*. 2019;233:43-48.
103. Kagami M, Sekita Y, Nishimura G, et al. Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes. *Nat Genet*. 2008;40(2):237-242.
104. Royo H, Cavaillé J. Non-coding RNAs in imprinted gene clusters. *Biol Cell*. 2008;100:149-166.
105. Cleaton MAM, Dent CL, Howard M, et al. Fetus-derived DLK1 is required for maternal metabolic adaptations to pregnancy and is associated with fetal growth restriction. *Nature Genet*. 2016;48:1473-1480.
106. Mei B, Zhao L, Chen L, Sul HS. Only the large soluble form of preadipocyte factor-1 (Pref-1), but not the small soluble and membrane forms, inhibits adipocyte differentiation: role of alternative splicing. *Biochem J*. 2002;364:137-144.
107. Perry JRB, Day F, Elks CE, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*. 2014;514:92-97.
108. Day FR, Thompson DJ, Helgason H, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet*. 2017;49:834-841.
109. Dauber A, Cunha-Silva M, Macedo DB, et al. Paternally inherited DLK1 deletion associated with familial central precocious puberty. *J Clin Endocr Metab*. 2017;102:1557-1567.

110. Warrington NM, Beaumont RN, Horikoshi M. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet.* 2019;51:804-814.
111. Sekita Y, Wagatsuma H, Nakamura K, et al. Role of retrotransposon-derived imprinted gene, Rtl1, in the fetomaternal interface of mouse placenta. *Nature Genet.* 2008;40:243-248.
112. Hernandez A, Martinez ME, Croteau W, St. Germain DL. Complex organization and structure of sense and antisense transcripts expressed from the DIO3 gene imprinted locus. *Genomics.* 2004;83:413-424.
113. Takahashi N, Okamoto A, Kobayashi R, et al. Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. *Hum Molec Genet.* 2009;18:1879-1888.
114. Sheng F, Sun N, Ji Y, et al. Aberrant expression of imprinted lncRNA MEG8 causes trophoblast dysfunction and abortion. *J Cell Biochem.* 2019;120:17378-17390.
115. Huang YT, Chu S, Loucks EB, et al. Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood. *Epigenetics.* 2016;11:227-236.
116. Walton E, Hass J, Liu J, et al. Correspondence of DNA Methylation Between Blood and Brain Tissue and Its Application to Schizophrenia Research. *Schizophr Bull.* 2016;42:406-414.
117. Ma B, Wilker EH, Willis-Owen SA, et al. Predicting DNA methylation level across human tissues. *Nucleic Acids Res.* 2014;42:3515-3528.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Quilter CR, Harvey KM, Bauer J, et al. Identification of methylation changes associated with positive and negative growth deviance in Gambian infants using a targeted methyl sequencing approach of genomic DNA. *FASEB BioAdvances.* 2021;3:205–230. <https://doi.org/10.1096/fba.2020-00101>