

GENETICS

DNA methylation at a nutritionally sensitive region of the *PAX8* gene is associated with thyroid volume and function in Gambian children

Toby Candler¹, Noah Kessler², Chathura Gunasekara³, Kate Ward^{1,4}, Philip James⁵, Eleonora Laritsky³, Maria Baker³, Roger Dyer⁶, Rajavel Elango⁶, David Jeffries¹, Robert Waterland^{3,7}, Sophie Moore^{1,8}, Marian Ludgate⁹, Andrew Prentice¹, Matt Silver^{1*}

PAX8 is a key thyroid transcription factor implicated in thyroid gland differentiation and function, and *PAX8* gene methylation is reported to be sensitive to the periconceptual environment. Using a novel recall-by-epigenotype study in Gambian children, we found that *PAX8* hypomethylation at age 2 years is associated with a 21% increase in thyroid volume and an increase in free thyroxine (T4) at 5 to 8 years, the latter equivalent to 8.4% of the normal range. Free T4 was associated with a decrease in DXA-derived body fat and bone mineral density. Furthermore, offspring *PAX8* methylation was associated with periconceptual maternal nutrition, and methylation variability was influenced by genotype, suggesting that sensitivity to environmental exposures may be under partial genetic control. Together, our results demonstrate a possible link between early environment, *PAX8* gene methylation and thyroid gland development and function, with potential implications for early embryonic programming of thyroid-related health and disease.

INTRODUCTION

Thyroid hormones regulate a wide range of physiological processes and influence various outcomes related to cognition, growth, skeletal, cardiovascular, and metabolic health (1). While clinical sequelae of severe perturbations of thyroid hormone production are well documented (2), variation of thyroid function within the normal population reference range is also associated with a range of phenotypic traits including blood pressure, lipids, obesity, cardiovascular mortality, bone mineral density (BMD), and cancer risk (3).

The function of the hypothalamic-pituitary-thyroid axis is clinically assessed by measurement of pituitary-derived thyrotropin (TSH), free T4 (free thyroxine), and free T3 (free tri-iodothyronine). Higher TSH and lower free T4 are associated with adverse pregnancy outcomes (3) and increased body mass index (BMI) in adults (4) and children (5). Lower TSH and higher free T4 are associated with an increased risk of osteoporosis and fracture (3), and levels of free T4 are correlated with the concentration of various lipoproteins (6).

Heritability of thyroid function has been reported to be between 32 and 65% (free T4), 23 and 67% (free T3), and 32 and 65% (TSH) (7). However, current identified genetic variants associated with thyroid function contribute only a small amount to interindividual variation in thyroid hormone concentrations. For example, although a recent genome-wide association study (GWAS) identified 74 loci associated with TSH, together these explained just 13.3% of TSH variance, leaving much of the reported heritability unexplained (8).

Moreover, congenital hypothyroidism (CHT), the most common endocrinopathy of childhood, is generally not inherited [less than 5% of cases have an identifiable genetic cause (9)] with 98% of cases nonfamilial (10) and a high discordance rate (92%) in monozygotic twins (11). In addition, seasonal variation in CHT incidence has been reported in the United Kingdom and Japan (12, 13), and a recent study reported that prenatal famine exposure has been associated with higher TSH in adulthood (14). Together, these observations suggest that environmental factors and epigenetic or unknown genetic mechanisms may play a role in thyroid development and/or function.

PAX8 (paired-box 8) protein is one of four known thyroid transcription factors (TTFs) essential for thyroid development and function [others include NK2 homeobox 1 (NKX2-1), forkhead box E1 (FOXE1), and hematopoietically expressed homeobox (HHEX)] (15). These transcription factors regulate expression of thyroid-specific genes related to thyroid hormone production and storage such as TPO (thyroid peroxidase), Tg (thyroglobulin), and the sodium-iodide symporter (15, 16) and are important for development and differentiation of the thyroid gland (17). *PAX8* has been described as the “master regulator,” with a role in regulating the activity of other TTFs (16). *PAX8* knockout mice demonstrate that thyroid hypoplasia, low birth weight and growth retardation (18), and genetic mutations in *PAX8* can cause CHT in humans (19).

Epigenetic processes, including DNA methylation, histone modification, protein binding of DNA, chromatin remodeling, and RNA-based mechanisms, can affect gene expression (20). DNA methylation at CpG “islands” [cytosine (C) and guanine (G) separated by phosphate (p)] within promoter regions may regulate gene transcription in a tissue-specific manner (21). Promoter DNA methylation is usually associated with condensed heterochromatin and transcriptional down-regulation (22). Metastable epialleles (MEs) are epigenetic loci that demonstrate systemic (i.e., not tissue-specific) methylation with substantial variation between individuals (23). The methylation patterns are thought to be established early in embryonic development (24) and are not driven by genetic variation (23, 24). Methylation at MEs

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

¹MRC Unit The Gambia at the London School of Hygiene and Tropical Medicine, London, UK. ²Department of Genetics, University of Cambridge, Cambridge, UK. ³USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA. ⁴MRC Lifecourse Epidemiology, University of Southampton, Southampton, UK. ⁵Department of Population Health, London School of Hygiene and Tropical Medicine, London, UK. ⁶British Columbia Children's Hospital Research Institute, Vancouver, BC, Canada. ⁷Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. ⁸Department of Women and Children's Health, King's College London, London, UK. ⁹Thyroid Research Group, School of Medicine, Cardiff University, Cardiff, UK.

*Corresponding author. Email: matt.silver@lshtm.ac.uk

is influenced by maternal diet around conception (25–29), and putative human MEs with stable methylation levels have been linked to disease-related phenotypes (25, 26). These characteristics position MEs as potential epigenetic mediators of the effect of early environmental exposures in the developmental origins of health and disease (30).

The *PAX8* gene contains a putative human ME with evidence of systemic interindividual variation (23, 25, 31) and sensitivity to periconceptional and prenatal environment (see table S1). There is evidence that *PAX8* promoter methylation patterns in leucocytes and thyroid are concordant in children (32). Data from The Gambia showed that leucocyte *PAX8* methylation was higher in children conceived in the annual rainy (or “hungry”) season (25, 31). In Bangladesh, higher methylation at the *PAX8* gene was reported in offspring following gestational famine exposure (27). A set of interlocking pathways, collectively known as one-carbon metabolism, provides methyl groups for DNA methylation and is dependent on multiple nutritional factors that act as substrates and essential cofactors (33). Seasonally driven variations in maternal circulating levels of one-carbon metabolites have been reported in The Gambia (34). Maternal supplementation with folic acid (a synthetic source of folate and a one-carbon metabolite) has been associated with differential *PAX8* methylation in adult offspring (35), and maternal preconception multiple micronutrient (MMN) supplementation was associated with differential methylation at *PAX8* in Gambian children (36).

Epigenetic influence on thyroid function or development has been little explored. Considering its key roles in thyroid development and regulation of the mature thyroid gland, and its epigenetic sensitivity to the early environment, *PAX8* is a prime candidate for study. Using a recall-by-epigenotype design, we examined links between *PAX8* DNA methylation measured at 2 years of age in peripheral blood and thyroid gland function and development in the same Gambian children aged 5 to 8 years. In addition, by examining body composition and BMD using dual energy x-ray absorptiometry (DXA) scans, we explored how *PAX8* methylation (via its putative effect on thyroid hormone production) may influence measures of adiposity and BMD. Using maternal biomarker data (including measures of one-carbon metabolites), we also explored how a child’s

PAX8 methylation status may be influenced by their mother’s nutritional status around the time of conception and assessed the influence of genetic variation in cis. Last, using data from The Cancer Genome Atlas (TCGA), we investigated the relationship between *PAX8* methylation and gene expression in thyroid tissue and assessed correlations between thyroid and whole blood methylation in samples from the Genotype-Tissue Expression (GTEx) Project.

RESULTS

PAX8 region of interest, participant selection, and baseline characteristics

Children from the “ENID” (Early Nutrition and Immune Development) trial (37) now aged between 5 and 8 years ($n = 493$) were recruited into “high” and “low” groups, according to their DNA methylation status in peripheral blood at age 2 years at a region of the *PAX8* gene (see Figs. 1 to 3). This region was selected on the basis of evidence of systemic interindividual variation and sensitivity to early environment from previous studies (23, 25, 31). DNA methylation was measured at four CpGs, which were highly correlated with adjacent CpGs and had sufficient coverage (see Materials and Methods and fig. S10 for further details).

One hundred eighteen children were recruited [high *PAX8* methylation group $n = 58$ (mean methylation (SD) = 0.96 (0.036)] and low *PAX8* methylation group $n = 60$ [0.50 (0.088)] with a median age [interquartile range (IQR)] of 7.18 years (1.67). Children in the low *PAX8* methylation group had a significantly lower BMI z score, but there were no significant differences in age, sex, weight-for-age z score (WAZ), height-for-age z score (HAZ), or pregnancy and infant supplementation received between the two groups (see Table 1).

Associations between *PAX8* methylation and thyroid outcomes

Thyroid volume

Thyroid ultrasound scans were performed blinded to *PAX8* methylation group on 114 children (four children did not attend for scans). Of these, two had technical difficulties in ascertaining accurate measurements and were removed from subsequent analysis (both were in the high methylation group). A single case of right thyroid lobe

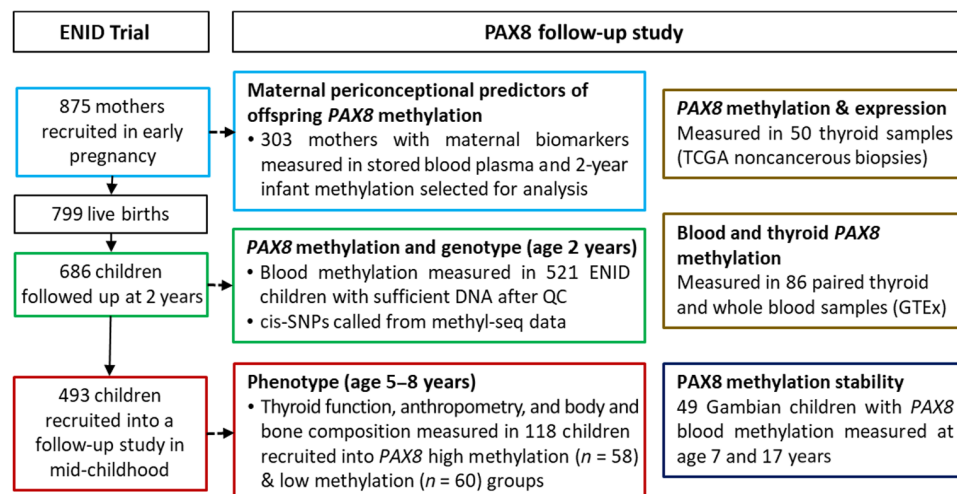


Fig. 1. Study overview. QC, quality control.

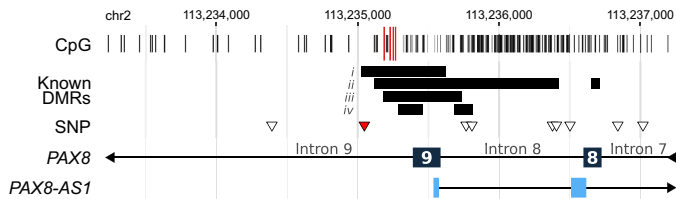


Fig. 2. PAX8 region of interest. The *PAX8* gene extends from chr2:113,215,997 to 113,278,921 (hg38) and contains 12 exons and 11 introns. Exons 8 and 9 of *PAX8* are shown, as well as the first two exons of an isoform of the *PAX8-AS1* antisense long noncoding RNA. CpGs are shown in the top track. The four CpGs highlighted in red were analyzed in this study (chr2:113,235,186; 113,235,228; 113,235,251; and 113,235,267). The “Known DMRs” (differentially methylated regions) track highlights regions identified in the following studies: putative MEs displaying systemic inter-individual variation identified in (i) Silver *et al.* (25) and (ii) Kessler *et al.* (23); (iii) DMR associated with gestational famine (27), Gambian SoC (35), and maternal folic acid supplementation (35); and (iv) DMRs associated with Gambian SoC in an additional study (31). The SNP track denotes variants within 2000 bp of the CpGs of interest that were called from the methyl-seq data. The SNP highlighted in red is close to our region of interest and tags an LD block encompassing it (see fig. S11). This is the SNP used for the genotype analyses (rs10193733; chr2:113,235,047 T > C).

hemigenesis in the high *PAX8* methylation group was the only abnormality identified. This participant was retained in the analysis as total thyroid volume was within normal range.

In crude unadjusted analyses, total thyroid volume was elevated in the low methylation group (3.24 versus 2.87 cm³ in the high *PAX8* methylation group, $P = 0.035$; see Table 2). In a multiple linear regression model adjusting for age, sex, BMI, and urinary iodine concentration (UIC), the association between *PAX8* methylation group and total thyroid volume was strengthened, yielding a 0.61 cm³ [SE = 0.15] or 21% higher thyroid volume in the low versus high group ($P = 0.0001$; see Table 3). This analysis also revealed significant associations between total thyroid volume and age, sex, BMI, and UIC (see table S2).

Thyroid function

In crude comparisons of thyroid function measures (free T4, free T3, TSH, and Tg) between the two groups, free T4 was observed to be significantly lower in the high *PAX8* methylation group (13.3 versus 13.9 pM, $P = 0.009$; see Table 2).

Iodine insufficiency was evaluated by two methods: plasma Tg level and UIC. On the basis of Tg level, there was no difference in iodine insufficiency between the groups, but significantly more children were classified with iodine insufficiency by UIC in the high *PAX8* methylation group compared to the low *PAX8* methylation group [21 of 56 (37.5%) versus 10 of 60 (16.7%), $P = 0.02$], and UIC was also higher in the low *PAX8* methylation group (170 versus 128 μg/liter, $P = 0.04$; see table S3). Since iodine concentration and/or insufficiency can affect thyroid function, UIC was included as an adjustment covariate in regression models to ensure appropriate adjustment was made for these group differences.

In a multiple linear regression model with adjustment for age, sex, and UIC, free T4 level was 0.85 pM [SE = 0.24] higher in the low *PAX8* methylation group ($P = 0.0007$; see Table 3 and table S4). The laboratory normal range for free T4 is between 9.0 and 19.1 pM, so this difference between the two groups—8.4% of the normal range—is substantial. No significant associations between *PAX8* methylation and TSH, free T3, or Tg were observed after adjustment for age, sex, and UIC (see Table 3 and table S4).

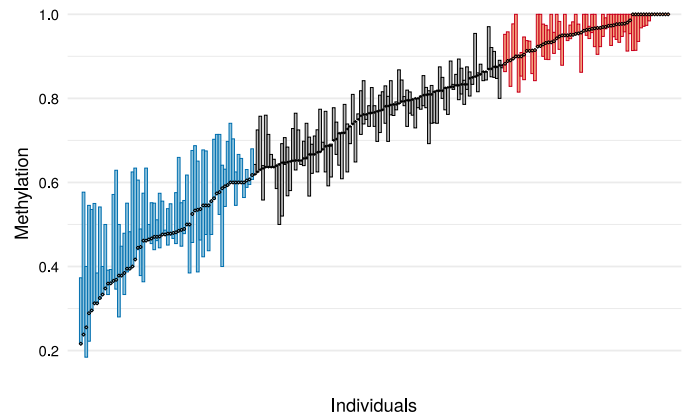


Fig. 3. DNA methylation in the study population at the PAX8 CpGs of interest. Individuals were sorted by methylation level at CpG chr2:113,235,228 (indicated by dots), which was used to select the high (red) and low (blue) *PAX8* methylation groups. Methylation range across the four CpGs of interest is indicated by the boxes.

Association between free T4 and body composition and bone mineral outcomes

We hypothesized that *PAX8* methylation (via its putative effect on thyroid hormone production) may influence measures of adiposity and BMD. To explore potential associations between free T4 and fat mass index (FMI), lean mass index (LMI), and BMD, whole-body DXA scans were performed on 113 of the 5- to 8-year-old children.

In multiple regression models adjusted for relevant covariates, log FMI was associated with free T4 ($\beta = -0.04$ [SE 0.02], $P = 0.033$; see Table 4 and table S5), so that for every picomolar increase in free T4, there was a 4.3% reduction in FMI. Free T4 was not associated with LMI (see table S6). Free T4 was inversely associated with log total-body-less-head (TBLH) BMD ($\beta = -0.008$ [0.004], $P = 0.044$), so that TBLH BMD was reduced by 0.8% for every picomolar increase in free T4 (see Table 4 and table S7).

Causal mediation analysis

Since we demonstrated significant associations between *PAX8* methylation group and both thyroid volume and free T4, we postulated that the methylation state set early in embryonic development (which we measured at the age of 2) could influence thyroid development and volume via regulation of *PAX8* expression and that this change in thyroid volume could in turn affect thyroid function and free T4 measured at 5 to 7 years of age. We therefore conducted a causal mediation analysis to test this, but this provided no evidence that the effect of differential methylation at *PAX8* on thyroid function was mediated by thyroid volume [average causal mediation effect (ACME) or mediated effect = 0.031, 95% confidence interval -0.15 to 0.22 , $P = 0.74$; see fig. S1]. As we did not find any evidence of a significant effect of *PAX8* methylation group on either BMD or any fat measure, we did not perform causal mediation analysis for these pathways.

Predictors of PAX8 methylation

We performed an analysis of the potential influence of sex and child and maternal BMI on *PAX8* methylation in $n = 521$ ENID-recruited children with 2-year *PAX8* methylation measurements (i.e., not restricted to the $n = 118$ high and low methylation groups analyzed above). In simple linear regression models, sex was significantly

Table 1. Baseline characteristics between high and low PAX8 methylation groups. Group differences in normally distributed variables (WAZ, HAZ, and BMI) assessed by Student's *t* test, nonnormally distributed variables (age) by Mann-Whitney *U* test, and categorical variables (sex and ENID supplementation group) by chi-square test. WAZ, weight-for-age z score; HAZ, height-for-age z score; PE, protein energy supplementation; FeFol, iron and folate supplementation; MMN, multiple micronutrient supplementation.

	All (n = 118)	High PAX8 methylation (n = 58)	Low PAX8 methylation (n = 60)	P value
Sex				
Male	71	30	41	0.10
Female	47	28	19	
Age (years)				
Median [IQR]	7.18 [1.67]	7.28 [1.49]	7.09 [1.61]	0.29
Range	5.22 to 8.71	5.22 to 8.68	5.23 to 8.71	
Body size measures*				
Birth weight (kg) [SD]	3.03 [0.43]	3.01 [0.46]	3.05 [0.41]	0.62
Mean WAZ [SD]	-1.25 [0.85]	-1.18 [0.87]	-1.32 [0.82]	0.36
Mean HAZ [SD]	-0.75 [0.83]	-0.81 [0.86]	-0.70 [0.79]	0.43
Mean BMI z score [SD]	-1.20 [0.87]	-1.01 [0.87]	-1.38 [0.84]	0.02
ENID pregnancy supplementation				
PE	31	16	15	0.90
FeFol	28	12	16	
MMN	26	13	13	
PE and MMN	33	17	16	
ENID infant supplementation				
MMN	54	27	27	1.00
No MMN	64	31	33	

*Body size measures reported were measured at recruitment for this current study (i.e., between 5 and 8 years of age).

Table 2. Thyroid volume and function comparison by PAX8 methylation group; crude (unadjusted) analyses. Group differences in normally distributed variables (thyroid volume, free T4, and free T3) were assessed by Student's *t* test and nonnormally distributed variables (TSH and Tg) by Mann-Whitney *U* test.

	All	High PAX8 methylation	Low PAX8 methylation	P value
Thyroid volume				
Mean total thyroid volume [SD] cm ³	3.06 [0.93] n = 112	2.87 [0.83] n = 53	3.24 [0.98] n = 59	0.035
Thyroid function				
Mean free T4 [SD], pM	13.6 [1.34] n = 114	13.3 [1.33] n = 55	13.9 [1.29] n = 59	0.009
Mean free T3 [SD], pM	6.09 [0.68] n = 114	6.08 [0.63] n = 55	6.09 [0.74] n = 59	0.92
Median TSH [IQR], mU/liter	1.82 [1.01] n = 116	1.86 [0.92] n = 56	1.73 [1.08] n = 60	0.34
Median Tg [IQR], μg/liter	19.2 [10.70] n = 105	18.5 [7.80] n = 49	19.4 [15.1] n = 56	0.46

associated with PAX8 methylation, with males having 0.24 [SE = 0.09] lower mean logit methylation compared to females ($P = 0.005$; see fig. S2 for comparison of equivalent untransformed mean % methylation difference). There were no significant associations

between PAX8 methylation and child BMI z score ($\beta = -0.02$ [SE = 0.04], $P = 0.64$), maternal BMI ($\beta = 0.0004$ [0.01], $P = 0.98$), infant WAZ ($\beta = -0.05$ [0.05], $P = 0.27$), or season of conception (SoC) ($\beta = 0.09$ [0.09], $P = 0.29$).

Table 3. PAX8 methylation group as a predictor of thyroid volume and function; multiple linear regression (adjusted) analyses. Low PAX8 methylation group coefficient gives the mean increase (decrease if negative) relative to the high PAX8 methylation group. Adjustment covariates for the multiple linear regression model for total thyroid volume are age, sex, BMI, urinary iodine (see table S2). Adjustment covariates for multiple linear regression model for free T4, free T3, logTSH, and logTg are age, sex, and urinary iodine (see table S3).

Outcome	Number of individuals	Low PAX8 methylation group coefficient [SE]	P value
Total Thyroid Volume (cm ³)	112	0.61 [0.15]	0.0001
Free T4 (pM)	113	0.85 [0.24]	0.0007
Free T3 (pM)	113	-0.02 [0.13]	0.88
Log TSH (mU/liter)	115	-0.11 [0.09]	0.26
Log Tg (µg/liter)	94	0.02 [0.12]	0.85

Three hundred three of the 521 children with PAX8 methylation data also had paired maternal biomarker data. In multiple linear regression models adjusted for sex, four one-carbon metabolites were associated with a decrease in PAX8 methylation (see Table 5): homocysteine (standardized $\beta = -0.11$ [0.05], $P = 0.048$), cysteine ($\beta = -0.16$ [0.05], $P = 0.003$), vitamin B12 ($\beta = -0.1$ [0.05], $P = 0.05$), and vitamin B6/pyridoxal-5'-phosphate (PLP, a vitamin B6 vitamer) ($\beta = -0.12$ [0.06], $P = 0.03$).

We also tested an extended panel of maternal biomarkers including amino acid levels (see table S8). Two of these biomarkers were associated with an increase in PAX8 methylation: uracil (standardized $\beta = 0.21$ [0.05], $P = 0.0001$) and arginine ($\beta = 0.16$ [0.06], $P = 0.006$). Valine was associated with a decrease in PAX8 methylation ($\beta = -0.13$ [0.05], $P = 0.02$).

PAX8 methylation, gene expression, and methylation tissue concordance

We next investigated the relationship between PAX8 methylation and gene expression in thyroid tissue by using data from TCGA. TCGA provides gene expression and DNA methylation data on a variety of noncancerous tissues including 50 thyroid samples. The methylation data are derived from the Illumina HM450 array, which includes two of the PAX8 CpGs of interest in this study. Since this region overlaps transcripts from sense (PAX8) and antisense (PAX8-AS1) genes, we considered the expression of both genes. Using fragments per kilobase per million mapped reads (FPKM) as an expression metric, PAX8 and PAX8-AS1 expression were positively correlated (Spearman $R = 0.477$, $P = 0.0006$; see fig. S3, left), but we found only weak evidence of an inverse relationship between PAX8 methylation and PAX8-AS1 or PAX8 expression (see fig. S4, top). However, with RNA sequencing by expectation maximization (RSEM) as the metric of gene expression, we found strong evidence that PAX8 methylation was negatively correlated with PAX8-AS1 expression (Spearman $R = -0.70$, $P = 6 \times 10^{-8}$; see fig. S4, bottom right), although there was

Table 4. Free T4 as a predictor of FMI and BMD; multiple linear regression (adjusted) analyses. Adjustment covariates for the multiple linear regression model for logFMI are age, sex, and weight (see table S4). Adjustment covariates for multiple linear regression models for log TBLH BMD and log TBLH BMC are age, sex, height, and weight (see table S6).

Outcome	Number of individuals	Free T4 (pM) coefficient [SE]	% change in outcome per one unit increase in free T4*	P value
Log FMI [fat mass (kg)/m ²]	113	-0.04 [0.02]	-4.30	0.033
Log TBLH BMD (g/cm ²)	113	-0.008 [0.004]	-0.80	0.044
		112		114

*Back transformation if outcome log-transformed using $[\exp(\beta) - 1] \times 100$.

no correlation between PAX8 and PAX8-AS1 expression (fig. S3, right). There was no association between PAX8 methylation and expression using RSEM (fig. S4, bottom left).

We analyzed mean DNA methylation across the four CpGs of interest in paired peripheral blood cell and thyroid tissue from 86 adult samples from the GTEx Project (38) and found no correlation (see fig. S5). Public GTEx expression data from 54 tissues shows that both PAX8 and PAX8-AS1 are predominantly expressed in thyroid (see fig. S6). Note that, in both the TGCA and GTEx dataset, thyroid methylation levels of the PAX8 region of interest fall only within the high methylation group from our Gambian analysis in peripheral blood. Together, our interpretation is that although individual variation in PAX8 methylation in the early embryo may affect thyroid development, at some stage of thyroid development, this region becomes uniformly highly methylated in thyroid of all individuals.

Stability of PAX8 methylation from 7 to 17 years of age

We measured PAX8 methylation in 49 Gambian children from stored peripheral blood DNA taken at age 7 and 17 years by pyrosequencing. Methylation was highly correlated between the two age groups ($R = 0.76$, $P < 2.2 \times 10^{-16}$; see fig. S7), indicating that blood cell methylation at this locus is highly stable across this age range.

Exploring the relationship between PAX8 methylation, genotype, and thyroid function

We focused on a single-nucleotide polymorphism (SNP) (rs10193733) proximal to our PAX8 region of interest for genetic analyses (see Fig. 2 and Materials and Methods for rationale for selecting this SNP). We observed higher methylation for homozygotes for the alternate allele (C/C, mean methylation = 0.97) compared to heterozygotes (T/C = 0.82 or wild-type alleles (T/T = 0.66; see Fig. 4). However, the most notable was that homozygotes for the alternate allele had a markedly limited methylation range (C/C, range = 0.93 to 1.00) compared to heterozygotes (C/T, range = 0.50 to 0.99) or homozygotes for the reference allele (T/T, range = 0.18 to 1.00; see Fig. 4).

This led us to postulate that the genotype with the greatest variability (T/T) could be more sensitive to periconceptual environment,

Table 5. Maternal one-carbon metabolite biomarkers as predictors of PAX8 Methylation. Multiple linear regression models have PAX8 mean logit methylation z score as the dependent variable with maternal one-carbon metabolites measured in maternal plasma and back-extrapolated to the time of conception as predictors, adjusted for sex. Hcy, homocysteine; DMG, dimethylglycine; PLP, pyridoxal 5-phosphate (B6 vitamer).

Covariate	Standardized coefficient	SE	t value	P value
Hcy (μM)	-0.11	0.05	-1.99	0.048
Methionine (μM)	-0.01	0.06	-0.13	0.89
Cysteine (μM)	-0.16	0.05	-2.95	0.003
Choline (μM)	-0.01	0.06	-0.20	0.84
Betaine (μM)	-0.07	0.05	-1.29	0.20
DMG (μM)	-0.03	0.06	0.52	0.61
B12 (pM)	-0.10	0.05	-1.96	0.050
Folate (nM)	-0.07	0.05	-1.37	0.17
PLP (nM)	-0.12	0.06	-2.17	0.031
Riboflavin (nM)	-0.11	0.06	-1.93	0.055

offering a potential explanation for the lack of evidence for a SoC effect at this locus. Methylation distributions stratified by genotype suggested a potential interaction between SoC and genotype (see fig. S8), although this was not statistically significant, possibly due to a lack of power. We also found no evidence for an interaction between maternal nutritional biomarkers and genotype.

We next assessed the direct effect of rs10193733 genotype on free T4 and total thyroid volume in multiple linear regression models with free T4 and total thyroid volume as the dependent variables regressed against the number of rs10193733 C alleles and adjusted for age, sex, and UIC. The C allele had a significant effect on both free T4 (β (per C allele) = -0.61 [SE = 0.16], $P = 0.0002$) and thyroid volume [β (per C allele) = -0.25 [0.11], $P = 0.02$]. As expected, (see Figs. 3 and 4), rs10193733 genotype was also significantly associated with PAX8 methylation group (chi-square test $P = 4.24 \times 10^{-13}$). However, the effect of PAX8 methylation group on thyroid volume and free T4 did not appear to be purely driven by genotype, as T/C and T/T individuals were well represented in both the high and low PAX8 methylation groups (see fig. S9).

DISCUSSION

Previous studies in The Gambia and elsewhere have shown a consistent association between DNA methylation at a genomic region in the PAX8 gene and the maternal periconceptional environment (25, 27, 29, 31). Here, we have demonstrated an association between PAX8 methylation and thyroid function in Gambian children, specifically free T4 and thyroid volume, with a relatively large effect size. Thyroid volume differed by >20% between low and high PAX8 methylation groups after adjusting for covariates. Free T4 demonstrated a difference of 0.85 pM between groups or 8.4% of the normal range.

A recent GWAS of thyroid function found four SNPs [with minor allele frequency (MAF) > 1%] associated with free T4. The largest reported SNP effect was 0.22 pM per variant allele, and overall, common genetic variants together explained just more than 20% of the variance in free T4 (39). The difference in free T4 between high

and low PAX8 methylation groups in our study is therefore much greater than the largest individual genetic effect previously observed (0.85 versus 0.22 pM). Similarly, a recent GWAS found four variants associated with thyroid volume explaining just more than 3% of thyroid volume variance (40). Here, the largest SNP effect size was 0.093 cm³ per allele, again far smaller than the effect size seen in our study (0.61 cm³ after adjustment for covariates).

Few studies have analyzed associations between DNA methylation and thyroid function. A recent epigenome-wide association study (EWAS) did not find any associations with free T4 but did identify two differentially methylated positions (DMPs) associated with TSH and six DMPs associated with free T3 (41). There were no DMPs from the PAX8 gene associated with thyroid function in this EWAS, although the study analyzed data from older European children (aged 14 to 17 years), and they may not have analyzed CpGs from our region of interest that has limited coverage on the Illumina HM450k array used in the study.

Our previous work in The Gambia demonstrated that PAX8 methylation is sensitive to the periconceptional environment, with rainy season conceptions having higher methylation compared to conceptions in the dry season (25, 31). There is evidence to suggest that the prenatal window where environmental exposures could influence PAX8 gene methylation may not be limited to the periconceptional period. Famine exposure (for at least 7 months prenatally) (27) and pregnancy folic acid supplementation (35) have been associated with PAX8 gene methylation in offspring. We did not find a significant SoC effect in this cohort. Our CpGs of interest only partially overlap the SoC-sensitive PAX8 differentially methylated region highlighted by Silver *et al.* (25), and they are adjacent to (but do not include) the SoC-sensitive region reported by Waterland *et al.* (31). Thus, while they are correlated with regions previously identified as being sensitive to SoC effects (see fig. S10), the PAX8 CpGs in this study were different and may be less sensitive to SoC effects. In addition, seasonal effects vary from year to year [see figure S12 by Waterland *et al.* (31)], so that it is possible that there may have been a diminished seasonal effect during the period of ENID trial recruitment compared to previous years. However, we did find

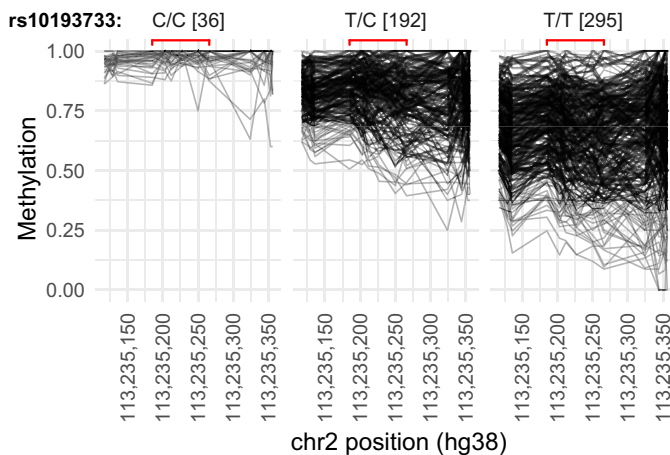


Fig. 4. C alleles at rs10193733 are associated with decreased DNA methylation variation at the *PAX8* region of interest. DNA methylation at CpGs is shown for the *PAX8* region of interest (delimited by red brackets) and flanking 200-bp region, split by rs10193733 genotype. Numbers in brackets denote the counts of individuals with each genotype at the SNP (total $N = 523$).

associations with circulating levels of several one-carbon metabolites and amino acids measured in maternal plasma and back-extrapolated to the time of conception, and these have previously been shown to vary seasonally in this population (34). One-carbon metabolites covary through complex interactions so that, as previously, we have presented their nominal associations, unadjusted for multiple comparisons (26, 33, 42).

There is evidence that the *PAX8* region of interest is an ME (23, 25, 27), and circulating levels of one-carbon metabolites at the time of conception have previously been associated with offspring DNA methylation at several MEs in our population (25, 26, 33). In our study, increased cysteine, homocysteine, and PLP (a B6 vitamers) were associated with lower DNA methylation at the *PAX8* region of interest, as previously observed at other MEs (29, 42). Furthermore, we identified the amino acids arginine and valine as potential predictors of *PAX8* methylation in a wider panel of periconceptional amino acid data. Amino acids and one-carbon metabolites are bidirectionally linked: Amino acids can supply cells with one-carbon metabolites (43), and levels of one-carbon metabolites can influence amino acid metabolism (44).

We observed a strong effect of genotype on DNA methylation variability, raising the possibility of a genotype-early environment interaction effect on methylation, as has been observed in a number of studies (45, 46). We did not find evidence for a genotype-SoC interaction at the *PAX8* region studied, although we may have had limited power to detect this. We noted the presence of an indel polymorphism within the *PAX8* variably methylated region and in very strong linkage disequilibrium (LD) with the cis-SNP associated with *PAX8* methylation variability. It is plausible that this polymorphism influences the binding of a protein that affects methylation variability. Future work should explore this and other mechanisms underpinning the strong variability effect that we observed and its potential link to differential sensitivity to environmental factors.

In this study, *PAX8* methylation was measured at 2 years of age in peripheral blood, with phenotypic measures of thyroid function measured at 5 to 8 years. There is evidence that methylation in this region is systemic, with similar methylation patterns across tissues

derived from endoderm, ectoderm, and mesoderm lineages, suggesting that methylation is established before gastrulation (23, 29, 31, 47). We therefore speculate that thyroid *PAX8* methylation, influenced by environmental factors and established early in gestation, sets a trajectory for thyroid gland development that is reflected in differential thyroid morphology and function in mid-childhood (see Fig. 5).

PAX8 is expressed early in embryological development [from days 20 to 22 in humans (15)]. Therefore, methylation patterns established in the cleavage-stage embryo could influence *PAX8* gene expression from the beginning of thyroid growth and development. Furthermore, methylation-driven differential *PAX8* expression in the differentiated gland could explain the difference in free T4 between the groups if high *PAX8* methylation down-regulates thyroid-specific genes associated with free T4 production. Further work is required to elucidate the cellular and molecular mechanisms underpinning these relationships.

Although our finding that this region is highly methylated in adult thyroid tissue from GTEx and TCGA appears to contradict this model, it is possible that thyroid methylation follows the systemic pattern in gestation and early life, followed by aging-associated hypermethylation specifically in thyroid. This is consistent with a previous study in children that compared DNA methylation in the *PAX8* promoter between leucocytes and thyroid tissue and found them to be correlated (32). A similar lineage-specific effect was observed at the murine *AxinFused* ME (48).

Our work also highlights a potential epigenetic contribution to thyroid gland development that has implications for understanding the etiology of nonheritable thyroid pathologies such as CHT. CHT is one of the top treatable causes of neurodevelopmental delay and is the commonest endocrinopathy of childhood (incidence between 1:2000 and 1:4000 newborns) (49); however, only a small proportion of CHT cases are attributed to known genetic mutations. Previous studies in Europe and Asia have found that the incidence of CHT displays a seasonal pattern (12, 13), in support of an environmental component in the etiology of CHT. Further characterization of environmentally sensitive epigenetic regions (such as *PAX8*) associated with thyroid development could thus have public health implications in identifying environmental drivers for CHT. Sexually dimorphic methylation effects such as those we observed at *PAX8* [and in a previous Gambian study (29)] may also help to explain the preponderance for CHT in females (2:1) (49).

We identified one child from the high methylation group with thyroid hemiagenesis, a rare developmental thyroid anomaly (prevalence reported to be between 0.05 and 0.5%) that has been associated with *PAX8* gene mutations (50). While we cannot attribute *PAX8* hypermethylation as the direct cause of this anomaly, the possible contribution of epigenetic alterations to abnormal thyroid development warrants further research.

In this study, we have also demonstrated associations between free T4 and body fat and BMD in Gambian children. In common with our findings, there is growing evidence that free T4 levels within the normal range are inversely correlated with measures of body fat (51). Data from Korean children found that free T4 was inversely correlated with waist circumference, BMI, and markers of insulin resistance (51). A study in U.K. children demonstrated that free T4 was negatively associated with FMI (from DXA) and BMI. To explore the direction of causation, the authors used Mendelian randomization and found that while BMI and body fat caused an increase in free T3, they did not seem to influence levels of free T4 (5). Thyroid

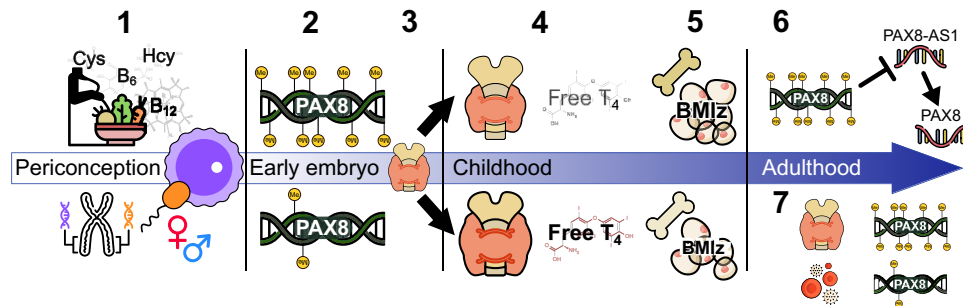


Fig. 5. Proposed model linking *PAX8* methylation and expression, thyroid function and development, and phenotype. (1) *PAX8* methylation is influenced by periconceptional nutritional factors, sex, and genotype. (2) *PAX8* methylation is set early in embryonic development. (3) *PAX8* methylation sets a trajectory for thyroid gland development: *PAX8* is expressed from gestational days 20 to 22 in humans, around the same time that thyroid progenitor cells begin specification in the endoderm (15). *PAX8* methylation state (or related epigenetic factors) alters *PAX8* expression that influences thyroid gland development. (4) *PAX8* methylation is inversely associated with thyroid size and free T₄ in mid-childhood. (5) Free T₄ is inversely associated with fat and BMD in mid-childhood. (6) *PAX8* methylation is inversely associated with *PAX8-AS1* expression. (7) There is a divergence between blood and thyroid methylation in adults.

volume and TSH have been shown to change in response to weight loss, although free T₄ remains unchanged, further supporting the idea that free T₄ is not influenced by body composition (52). Data from African children are limited, although a positive correlation between free T₃/T₄ ratio and BMI has been reported in Nigerian children (53). We found no association between free T₄ and measures of lean mass, although an inverse relationship between free T₄ and lean mass has been reported previously (54).

Free T₄ has been reported to be negatively associated with BMD in adults (3), and higher free T₄ (within the population reference range) is associated with an increased incidence of fractures in older people (55). Chondrocytes, osteoblasts, and osteoclasts express TSH receptor and thyroid hormone receptor, and thus, there is a biological target for thyroid hormone to affect bone modeling (56). Our novel finding that BMD is associated with free T₄ in mid-childhood is of great interest as this reflects an important period of bone accrual and attainment of peak bone mass, which is related to future fracture risk. We found no overall effect of *PAX8* methylation on BMD and so did not investigate the possibility that the effect of methylation on BMD is mediated by free T₄. However, it is possible that *PAX8* methylation may exert countervailing effects on BMD such that the net effect is nonsignificant despite their being a significant causal pathway between *PAX8* methylation, free T₄, and BMD. *PAX8* has been reported to be hypomethylated in human cartilage in individuals with osteoarthritis (57), which is of interest considering our finding that low *PAX8* methylation is associated with higher free T₄, which, in turn, is associated with lower BMD.

The associations between *PAX8* methylation, thyroid phenotype and BMI, fat, and bone measures could reflect an adaptive process. For example, if higher *PAX8* methylation [associated with a negative maternal energy balance in the Gambian rainy season (34)] is associated with a reduced level of thinness (as shown by a higher BMI in the high methylation group), then this could be an example of the organism developing greater fat stores in response to a predicted nutritionally adverse postnatal environment.

We found an inverse relationship between *PAX8* methylation and *PAX8-AS1* expression (measured by RSEM) in normal thyroid biopsies using data from the TCGA. However, we obtained contradictory results when *PAX8-AS1* expression was measured using the FPKM metric (both expression measures are provided in the TCGA dataset). In general, these metrics are well correlated except when

the exons are short (as for *PAX8-AS1*). There may be difficulty in accurately assessing the expression with short transcripts due to limitations with the sequencing technology and normalization methods (58); therefore, we present both metrics for consideration. DNA methylation in gene promoters is generally associated with condensed heterochromatin and reduced gene expression (22), while methylation within genes, downstream from the transcription start site (TSS), does not have as clear a correlation with gene expression (59). Our region of interest is intragenic to *PAX8* but is located within a promoter region of the antisense long noncoding RNA (lncRNA) *PAX8-AS1* (also referred to as *LOC654433*). At loci such as this, it is possible for complex regulatory interactions to exist between DNA methylation, coding gene expression, and noncoding RNA expression (59). A study in fibroblasts demonstrated that the relationship between methylation and gene expression at *PAX8* is region specific, with a positive correlation between methylation and expression reported at CpGs close to the *PAX8* TSS and a strong negative correlation reported at intragenic CpGs located near the *PAX8-AS1* TSS, including one of the CpGs in our region of interest (60). Furthermore, this study demonstrated a positive correlation between *PAX8* and *PAX8-AS1* gene expression and postulated a potential chromatin-activation linked role of *PAX8-AS1* in regulating *PAX8* expression. Previous studies have demonstrated that polymorphisms in *PAX8-AS1* are potentially expression quantitative trait loci for *PAX8* (61), and therefore, it appears that our region could be a putative intragenic regulatory region. The specific biological consequence of the *PAX8-AS1* lncRNA is still yet to be fully elucidated, but variants in *PAX8-AS1* are associated with cancer risk (62).

A previous study identified moderate iodine deficiency in Gambian children aged 8 to 12 years (63), and mothers recruited for the ENID trial whose offspring were followed up for this study were found to have moderate iodine deficiency (64). We therefore attempted to account for this in our analysis as iodine levels can influence thyroid volume and function. We found no difference between the *PAX8* methylation groups when a Tg cutoff was used to define iodine insufficiency, but we did find a significant difference between the groups when using UIC and therefore adjusted for this in the analysis. We recognize that a measure of urinary iodine can be a useful tool to understand iodine insufficiency in a population but may be less useful for characterizing an individual's iodine status (65). However, we note that iodine levels were all taken at the same time

of year, all in the morning from fasted individuals and where the local diet day to day is relatively consistent, suggesting that UIC may be informative. It is possible that differential *PAX8* methylation could be contributing to the observed differences in urinary iodine between the two groups possibly by influencing expression of the sodium-iodide symporter.

Prior evidence that *PAX8* methylation is a putative ME, with systemic methylation established in the early embryo, supports the notion that interindividual variation in *PAX8* methylation or related epigenetic marks may drive the phenotypic differences observed in this study. This highlights the potential benefits of studying links between regions of systemic interindividual variation and risk of disease (66). In this study, we cannot rule out the possibility of a reverse causation effect of thyroid function measures on methylation or other related epigenetic factors at 2 years of age. Elucidation of causal pathways linking environmental exposure, methylation and other epigenetic factors, gene expression, and postnatal phenotype will require mechanistic investigations in cell or animal models. For example, a mouse model has demonstrated an association between high estrogen exposure in early pregnancy and higher free T4, *PAX8* promoter hypomethylation, and increased *PAX8* expression in offspring (67). However, we note that our region of interest is at least partially absent from several species commonly used as models of development including mouse, zebrafish, and *Xenopus tropicalis*. Furthermore, the *PAX8-AS1* lncRNA present in humans does not appear to exist in other species. Its putative role in regulating *PAX8* expression and thyroid development may therefore only exist in humans.

In summary, we have demonstrated that individual variation in DNA methylation at a region of the *PAX8* gene sensitive to periconceptional nutrition is significantly associated with total thyroid volume and free T4 levels in Gambian children. Our work has potential implications for understanding the fetal origins of health and disease and may contribute to our understanding of the epigenetic drivers of thyroid development and function.

MATERIALS AND METHODS

Experimental design

We used a “recall-by-epigenotype” design to examine links between *PAX8* DNA methylation measured at 2 years of age and thyroid gland function and development in the same Gambian children aged 5 to 8 years. We used an existing longitudinal cohort of Gambian children [children from the ENID (37) study now aged between 5 and 8 years, $n = 493$] to recruit the top (high) and bottom (low) quantiles for DNA methylation at a region of the *PAX8* gene previously identified as sensitive to the periconceptional environment from banked DNA at 2 years of age. Participants were assessed for thyroid volume, function (free T3, free T4, TSH, and Tg), UIC, and body composition and bone measures by whole-body DXA scan. Figure 1 provides an overview of the study, and further details are provided below.

ENID trial and cohort

The ENID trial (ISRCTN49285450) (37) was a combined trial of pregnancy and infancy nutritional supplementation conducted in West Kiang, a rural region of The Gambia, recruiting pregnant mothers between January 2010 and February 2014. Briefly, women were recruited in early pregnancy (10 to 20 weeks) and randomized

to receive either (i) iron-folate (standard care); (ii) MMN; (iii) energy, protein, and lipid with iron-folate; or (iv) energy, protein, and lipid with MMN supplements for the remainder of their pregnancy. There were no differences in maternal (BMI, age, and parity) or infant characteristics (birthweight, birth length, sex, or gestational age) across the study arms (68). From 6 to 18 months of age, infants were further randomized to a lipid-based nutritional supplement, with or without additional MMN. A total of 875 women were randomized in pregnancy to one of the four study arms and 686 participants completed follow-up to 2 years of age (69). Routine blood samples were collected from mothers and infants at various time points including a peripheral blood cell DNA samples stored at age 2 years. At the time of this current study, children from the ENID study, now aged between 5 and 8 years of age ($n = 493$), were being followed up monthly in a separate longitudinal observational study (see Fig. 1).

PAX8 methylation measurement at 2 years of age

Children from the original ENID trial had DNA isolated from whole blood at age 2 years. DNA was enriched for a panel of candidate regions and bisulfite-converted using a custom Agilent SureSelect Methyl-Seq targeted capture system on a subset of these children with sufficient DNA available for processing ($n = 521$) (70). Target-enriched DNA, including the *PAX8* region of interest, was sequenced using the Illumina NovaSeq platform at the Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA. Reads were mapped to the human genome (hg38) using Bismark v0.20.0 (71) with default options, which was also used to extract methylation values after mapping. Methylation calls from opposite strands of the same CpG site were combined. Within each individual, CpG sites were considered “covered” if they had a read depth of at least 20 \times ; uncovered sites were excluded from analyses.

Selection of PAX8 region of interest

We selected a subset of CpGs in the *PAX8* region for which there was independent evidence of systemic interindividual variation and sensitivity to early environment from previous studies (23, 25, 27). From these, four CpGs of interest, with coverage in a large number of samples and strong correlation with nearby CpGs, were chosen (see fig. S10): chr2:113,235,186; chr2:113,235,228; chr2:113,235,251; and chr2:113,235,267. All of these CpGs lie in intron 9 of *PAX8* and in a promoter region of antisense *PAX8-AS1* (see Fig. 2; all genomic coordinates hg38).

Participant selection

From the 493 children being followed up in the ENID cohort at age 5 to 8 years (see Fig. 1), a “recall by epigenotype” design was used whereby participants were selected by methylation level at CpG chr2:113,235,228. This CpG was chosen due to it having the best overall correlations with nearby CpG methylation levels (see fig. S10). Individuals with at least one other informative CpG among the four CpGs of interest and without large differences (≥ 0.2) in quantile among the CpGs of interest ($n = 217$) were then placed into high or low groups based on the chr2:113,235,228 methylation level (see Fig. 3). In total, 125 participants (low *PAX8* methylation group $n = 64$, high *PAX8* methylation group $n = 61$) were identified for potential recruitment, with 118 participants (94%) consenting to participate in the study ($n = 7$ declined to participate).

Thyroid volume assessment

Thyroid ultrasound was conducted by T.C. who was blinded to the participant's *PAX8* methylation group. The length (*l*), width (*w*), and depth (*d*) of each thyroid lobe (in centimeters) were measured on transverse and longitudinal scans using Sonosite MicroMaxx (10-Hz probe). All measurements were made in triplicate, and the mean was used for analysis. The volume (Vol) of each lobe (in milliliters) was estimated by the modified formula for an ellipsoid; $\text{Vol(ml)} = (0.479 \times d \times w \times l) / 72$, and the total thyroid volume was calculated as the sum of the volumes of both lobes.

Mid-childhood blood collection and biochemical assessments

Recruited children had a morning (between 8 a.m. and 10.30 a.m.) venous blood sample taken into a serum sample collection tube in April 2019. Aliquoted serum was frozen at -70°C . The cellular fraction was discarded. During the same study visit as the blood sample and thyroid ultrasound, a fasted morning urine sample was collected in iodine-free tubes and stored at -20°C . The urine and serum samples were shipped to the University Hospital of Wales, UK for thyroid hormone and urinary iodine measurements. Serum TSH, free T3, and free T4 were measured by the automated ALINITY System (Abbott Laboratories, USA), and Tg was measured by the Beckman Access DxI. Urinary iodine and urinary creatinine were measured by inductively coupled plasma mass spectrometry. Iodine sufficiency was defined as a Tg $< 40 \mu\text{g/liter}$ and/or a UIC $> 100 \mu\text{g/liter}$ (63).

Other phenotypic measures

Standing height was calculated as the mean of measures taken in triplicate to the nearest millimeter using a portable stadiometer (Seca 213). Weight was similarly calculated from measures in triplicate to the nearest 0.1 kg using electronic scales (Seca 803), with participants clothed, but with shoes and coat removed. BMI was calculated as weight (kilograms) divided by height² (m²). HAZ, WAZ, and BMI SD score for each participant were calculated using World Health Organization reference ranges (73). A whole-body DXA scan was performed using the GE-Lunar Prodigy scanner (GE Medical, Waltham, MA; software version 13.60.033) on 113 children (5 did not attend for DXA scan). Bone-related outputs included areal BMD (g/cm²), bone mineral content (BMC, g), and bone area (BA, cm²). For children, it is recommended to use TBLH for bone-related measurements (74). The analyzed outcome measure for bone was TBLH BMD and was calculated as $\text{TBLH BMD (g/cm}^3) = \text{TBLH BMC (g)} / \text{TBLH BA (cm}^2)$. The analyzed outcome measure for body fat was FMI, which was calculated as fat mass (kg) divided by height² (m²). The analyzed outcome measure for lean mass was LMI, which was calculated as lean mass (kg) divided by height² (m²).

Maternal biomarker data

Women recruited to the ENID trial provided a 10-ml venous blood sample at the time of enrolment in early pregnancy (< 20 weeks of gestation). Plasma was stored at -70°C . A subset of 350 women were previously selected for biomarker analysis on the stored samples; $n = 303$ women with maternal biomarker data had *PAX8* methylation data available for their offspring. These women were selected to give an even distribution by month of enrolment and to provide data from the earliest gestational age, i.e., sample collection

nearest to time of conception (median 12.1 weeks, IQR 3.5). In our analysis, we examined a core set of 10 nutritional biomarkers involved in one-carbon metabolism [homocysteine, methionine, cysteine, choline, betaine, dimethylglycine (DMG), vitamin B12, folate, PLP, and vitamin B2 (riboflavin)]. We also considered an extended panel of biomarkers to capture other metabolic pathways potentially influencing one-carbon metabolism [α -1-acid glycoprotein (AGP), aspartate, threonine, serine, glutamate, glycine, alanine, valine, isoleucine, leucine, tyrosine, phenylalanine, lysine, histidine, arginine, proline, uridine, and uracil]. Biomarkers were analyzed at the BC Children's Hospital, Canada, using liquid chromatography-tandem mass spectrometry (choline, betaine, DMG, homocysteine, cysteine, methionine, PLP, riboflavin, uracil, and uridine), Abbott AxSYM autoanalyzer (folate and vitamin B12), and Hitachi L-8900 amino acid analyzer (additional amino acids: serine, glycine, alanine, arginine, aspartic acid, glutamic acid, histidine, isoleucine, leucine, lysine, phenylalanine, proline, threonine, tryptophan, tyrosine, and valine). The inflammatory marker AGP was measured using the Cobas Integra 400 plus autoanalyzer at the MRC Unit The Gambia, Keneba field station.

Stability of *PAX8* methylation from mid-childhood to after puberty/adolescence

Stability of the four selected *PAX8* CpGs with age was assessed by analyzing blood samples from 49 Gambians with samples collected in mid-childhood (7 years old) and adolescence (17 years old) (25) by pyrosequencing (see Supplementary Methods).

PAX8 gene expression and DNA methylation in blood and thyroid tissues

We accessed data from TCGA, an open access resource providing genomic data for more than 20,000 cancer and noncancerous matched samples obtained from biopsy from live individuals. We downloaded *PAX8* gene expression data on 50 individuals with noncancerous thyroid samples, reported as both FPKM upper quartile (from GDC v18.0 PANCAN HTSeq hosted by Xena, <https://gdc.xenahubs.net>) and RSEM (data from TCGA Wanderer, <http://maplab.imppc.org/wanderer/>) from Illumina HiSeq RNA-seq data. TCGA methylation data for the same samples measured on the Illumina HumanMethylation450 array were available for two of our CpGs of interest (chr2:113,235,186 and 113,235,267) from TCGA Wanderer.

A comparison of *PAX8* methylation levels in blood and thyroid tissue was carried out through quantitative analysis of the four *PAX8* CpGs of interest by bisulfite pyrosequencing in paired thyroid tissue, and whole blood samples from 86 donors were obtained from the GTEx postmortem tissue database (38) (see Supplementary Methods for sample details).

Genotyping from methyl-seq data

We used the targeted methyl sequencing (methyl-seq) data to call individual genotypes within the *PAX8* gene. A modified version of BS-Snperr (75, 76) was used to call SNP genotypes from the bisulfite-converted reads. Genotypes with a nonreference allele frequency (AF) of ≥ 0.05 and with a call rate (proportion of individuals with called genotypes) of 20% or higher were considered for the analysis ($n = 14$ SNPs). Of these, SNP rs10193733 (chr2:113,235,047 T > C; MAF, 0.25) was the closest to the *PAX8* variably methylated region analyzed here (see Fig. 2) and had a high call rate in the methyl-seq dataset ($> 99\%$ of available samples with methylation data; $n = 523$), enabling

analysis of interactions between methylation, genotype, and thyroid phenotypes. Furthermore, we verified that this SNP is in LD with several other SNPs on either side of the variably methylated region using Gambian reference data from the Gambian Genome Variation Project (GGVP) (77) and 1000 Genomes Project, Phase3 [Gambia West Division (GWD)] (78) and confirmed this in our study population (see fig. S11). This SNP and the nearby SNPs in LD all had similar AFs to the GWD reference population from GGVP, demonstrating the reliability of the methyl-seq-derived genotypes (with the exception of rs7576384 chr2:113235808 C > G, which had a much lower AF than expected, likely due to the loss of informative reads after bisulfite conversion). A 2–base pair (bp) indel polymorphism not detectable by the genotype caller, rs35724515 (chr2:113235224 CCC > C), is located within our region of interest and is in high LD with rs10193733 (1000 Genomes GWD LD $r^2 = 0.976$).

Statistical analysis

Baseline characteristics and crude between-group comparisons

Differences between high and low *PAX8* methylation groups for the categorical variables of sex, pregnancy and infant supplementation, and iodine sufficiency (yes/no) were assessed using Pearson chi-square tests. Birth weight, WAZ, BMI *z* score, HAZ, total thyroid volume, free T4, and free T3 were normally distributed, and group differences were assessed using Student's *t* test with mean and SD reported. Age, UIC, urinary iodine:creatinine ratio, TSH, and Tg were not normally distributed, and group differences were assessed by Mann-Whitney *U* tests with median and IQR reported.

Multiple linear regression models

Unless otherwise indicated, all outcome variables were normally distributed.

Thyroid volume. Thyroid volume as the dependent variable was regressed against *PAX8* methylation group and adjusted for sex, age, BMI, and UIC. BMI was included to ensure an adjustment for body size as (i) there was a significant difference between the two groups at recruitment and (ii) BMI also provided the best model fit (indicated by lowest Akaike information criterion) compared to other measures of body size (e.g., WAZ, HAZ, and body surface area). UIC was included as there was a significant difference between the groups at recruitment and iodine deficiency can be associated with goiter/thyroid size.

Measures of thyroid function. Measures of thyroid function (free T4, free T3, TSH, and Tg) were regressed against *PAX8* methylation group and adjusted for sex, age, and UIC. TSH and Tg were normally distributed after log transformation before regression analysis. Iodine deficiency can be associated with hypothyroidism and an increase in Tg, and therefore, UIC was included in all models. Correction for multiple tests was not made as the markers of thyroid function are not independent.

DXA-derived measures. FMI was normally distributed after log transformation. FMI was regressed against free T4 and adjusted for age, sex, and weight (to adjust for lean mass). LMI was regressed against free T4, with LMI adjusted for age, sex, and weight (to adjust for fat mass). TBLH BMD was normally distributed after log transformation and regressed against free T4, adjusted for age, sex, weight, and height.

Predictors of *PAX8* methylation analysis

In analyses where *PAX8* methylation was considered as an outcome, 2-year methylation measurements were taken from the wider ENID cohort (i.e., not just the high and low *PAX8* methylation

groups; $n = 303$ to 521 depending on the predictor considered) and treated as a continuous variable. Methylation across the four CpGs of interest was highly correlated (Pearson *R* between 0.871 and 0.952; fig. S10). A univariate composite measure was therefore used in all regression models, calculated as the mean *z* score (over all CpG sites) of the logit-transformed methylation level at each CpG site (referred to as *PAX8* mean logit methylation *z* score). Methylation levels of 1.0 were reduced to 0.99 to prevent infinite values after logit transformation.

All maternal biomarkers were preadjusted for gestational age, maternal BMI, maternal age, and inflammation (AGP) and then back-extrapolated to date of conception using previously described methods (29). Those biomarkers not normally distributed were log-transformed, and all biomarkers were scaled and centered to enable comparison of standardized coefficients. Multiple linear regression models with *PAX8* mean logit methylation *z* score as the dependent variable were fitted individually with each biomarker as a predictor and adjusted for sex.

For other predictors, *PAX8* mean logit methylation *z* score was regressed against maternal BMI, infant BMI *z* score, infant WAZ, infant sex, and SoC in separate models. SoC was defined as “rainy” (January to June) and “dry” (July to December) as previously described (79) with the conception date calculated from a gestational age estimation obtained from antenatal ultrasound at ENID trial recruitment.

All model covariates were assessed for multicollinearity, and standard tests were performed to ensure that linear modeling assumptions were met. Where reported, coefficients (β) associated with log-transformed dependent variables were back transformed using $(\exp(\beta) - 1) \times 100$ to represent percentage change in dependent variable per unit increase in the corresponding predictor.

Causal mediation analysis

We performed a causal mediation analysis to test the hypothesis that the observed effect of *PAX8* methylation group on free T4 is mediated by its effect on thyroid volume. We used the Mediation package (v4.5.0) in R, with confidence intervals for direct and indirect effects calculated using a nonparametric bootstrap with 10,000 simulations. All statistical analyses were performed using R version 3.6.2 (80).

Analyses by genotype

Genetic analyses focused on a single SNP (rs10193733) that tagged an LD block proximal to our *PAX8* region of interest (see above for justification for choosing this SNP). We performed two sets of analyses stratified by rs10193733 genotype. First, Student's *t* tests were used to compare the *PAX8* mean logit methylation *z* score between SoC (rainy versus dry) separately for individuals with each genotype. Second, we fitted multiple linear regression models with total thyroid volume and free T4 as dependent variables adjusted for relevant covariables, again stratified by rs10193733 genotype. We also assessed the influence of the rs10193733 “C” allele on free T4 and thyroid volume (outcome variables) in multiple regression models with allelic dosage (predictor) coded as C/C = 2, C/T = 1, T/T = 0.

Ethics

Ethical approvals for the ENID trial and for this study were given by The Gambia Government/MRC Joint Ethics Committee (SCC1126v2, SCC1640, and L2015.51v2). Consent was gained by signature or

thumb print from mothers for their own participation and that of their child. All data were anonymized before analysis.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abj1561>

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. K. Boelaert, J. A. Franklyn, Thyroid hormone in health and disease. *J. Endocrinol.* **187**, 1–15 (2005).
2. P. Hanley, K. Lord, A. J. Bauer, Thyroid disorders in children and adolescents: A review. *JAMA Pediatr.* **170**, 1008–1019 (2016).
3. P. N. Taylor, S. Razvi, S. H. Pearce, C. M. Dayan, Clinical review: A review of the clinical consequences of variation in thyroid function within the reference range. *J. Clin. Endocrinol. Metab.* **98**, 3562–3571 (2013).
4. L. Mosso, A. Martinez, M. P. Rojas, G. Latorre, P. Margozzini, T. Lyng, J. Carvajal, C. Campusano, E. Arteaga, L. Boucai, Early pregnancy thyroid hormone reference ranges in Chilean women: The influence of body mass index. *Clin. Endocrinol.* **85**, 942–948 (2016).
5. P. N. Taylor, R. Richmond, N. Davies, A. Sayers, K. Stevenson, W. Woltersdorf, A. Taylor, A. Groom, K. Northstone, S. Ring, O. Okosieme, A. Rees, D. Nitsch, G. R. Williams, G. D. Smith, J. W. Gregory, N. J. Timpson, J. H. Tobias, C. M. Dayan, Paradoxical relationship between body mass index and thyroid hormone levels: A study using mendelian randomization. *J. Clin. Endocrinol. Metab.* **101**, 730–738 (2015).
6. L. H. Duntas, Thyroid disease and lipids. *Thyroid* **12**, 287–293 (2002).
7. V. Panicker, Genetics of thyroid function and disease. *Clin. Biochem. Rev.* **32**, 165–175 (2011).
8. W. Zhou, B. Brumpton, O. Kabil, J. Gudmundsson, G. Thorleifsson, J. Weinstock, M. Zawistowski, J. B. Nielsen, L. Chaker, M. Medici, A. Teumer, S. Naitza, S. Sanna, U. T. Schultheiss, A. Cappola, J. Karjalainen, M. Kurki, M. Oneka, P. Taylor, L. G. Fritsche, S. E. Graham, B. N. Wolford, W. Overton, H. Rasheed, E. B. Haug, M. E. Gabrielsen, A. H. Skogholt, I. Surakka, G. D. Smith, A. Pandit, T. Roychowdhury, W. E. Hornsby, J. G. Jonasson, L. Senter, S. Liyanarachchi, M. D. Ringel, L. Xu, L. A. Kiemeny, H. He, R. T. Netea-Maier, J. I. Mayordomo, T. S. Plantinga, J. Hrafnkelsson, H. Hjartarson, E. M. Sturgis, A. Palotie, M. Daly, C. E. Citterio, P. Arvan, C. M. Brummett, M. Boehnke, A. de la Chapelle, K. Stefansson, K. Hveem, C. J. Willer, B. O. Åsvold, GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. *Nat. Commun.* **11**, 3981 (2020).
9. C. Peters, A. S. P. van Trotsenburg, N. Schoenmakers, Diagnosis of endocrine disease: Congenital hypothyroidism: Update and perspectives. *Eur. J. Endocrinol.* **179**, R297–R317 (2018).
10. M. Castanet, S. Lyonnet, C. Bonaiti-Pellié, M. Polak, P. Czernichow, J. Léger, Familial forms of thyroid dysgenesis among infants with congenital hypothyroidism. *N. Engl. J. Med.* **343**, 441–442 (2000).
11. R. Perry, C. Heinrichs, P. Bourdoux, K. Khoury, F. Szots, J. H. Dussault, G. Vassart, G. Van Vliet, Discordance of monozygotic twins for thyroid dysgenesis: Implications for screening and for molecular pathophysiology. *J. Clin. Endocrinol. Metab.* **87**, 4072–4077 (2002).
12. S. K. Hall, A. C. J. Hutchesson, J. M. Kirk, Congenital hypothyroidism, seasonality and consanguinity in the West Midlands, England. *Acta Paediatr.* **88**, 212–215 (1999).
13. Y.-H. Gu, T. Kato, S. Harada, H. Inomata, T. Saito, K. Aoki, Seasonality in the incidence of congenital hypothyroidism in Japan: Gender-specific patterns and correlation with temperature. *Thyroid* **17**, 869–874 (2007).
14. J. Guo, D. Teng, X. Shi, Y. Li, J. Ba, B. Chen, J. Du, L. He, X. Lai, Y. Li, H. Chi, E. Liao, C. Liu, L. Liu, G. Qin, Y. Qin, H. Quan, B. Shi, H. Sun, X. Tang, N. Tong, G. Wang, J.-A. Zhang, Y. Wang, Y. Xue, L. Yan, J. Yang, L. Yang, Y. Yao, Z. Ye, Q. Zhang, L. Zhang, J. Zhu, M. Zhu, Z. Shan, W. Teng, Exposure to the Chinese great famine in early life and thyroid function and disorders in adulthood: A cross-sectional study. *Thyroid* **31**, 563–571 (2021).
15. L. P. Fernández, A. López-Márquez, P. Santisteban, Thyroid transcription factors in development, differentiation and disease. *Nat. Rev. Endocrinol.* **11**, 29–42 (2015).
16. S. Ruiz-Llorente, E. C. de Pau, A. Sastre-Perona, C. Montero-Conde, G. Gómez-López, J. A. Fagin, A. Valencia, D. G. Pisano, P. Santisteban, Genome-wide analysis of Pax8 binding provides new insights into thyroid functions. *BMC Genomics* **13**, 147 (2012).
17. M. De Felice, R. Di Lauro, Thyroid development and its disorders: Genetics and molecular mechanisms. *Endocr. Rev.* **25**, 722–746 (2004).
18. A. Mansouri, K. Chowdhury, P. Gruss, Follicular cells of the thyroid gland require Pax8 gene function. *Nat. Genet.* **19**, 87–90 (1998).
19. M. Komatsu, T. Takahashi, I. Takahashi, M. Nakamura, I. Takahashi, G. Takada, Thyroid dysgenesis caused by PAX8 mutation: The hypermutability with CpG dinucleotides at codon 31. *J. Pediatr.* **139**, 597–599 (2001).
20. E. R. Gibney, C. M. Nolan, Epigenetics and gene expression. *Heredity* **105**, 4–13 (2010).
21. A. M. Deaton, A. Bird, CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
22. P. A. Jones, Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).
23. N. J. Kessler, R. A. Waterland, A. M. Prentice, M. J. Silver, Establishment of environmentally sensitive DNA methylation states in the very early human embryo. *Sci. Adv.* **4**, eaat2624 (2018).
24. V. K. Rakyan, M. E. Blewitt, R. Druker, J. I. Preis, E. Whitelaw, Metastable epialleles in mammals. *Trends Genet.* **18**, 348–351 (2002).
25. M. J. Silver, N. J. Kessler, B. J. Hennig, P. Dominguez-Salas, E. Laritsky, M. S. Baker, C. Coarfa, H. Hernandez-Vargas, J. M. Castellino, M. N. Routledge, Y. Y. Gong, Z. Herczeg, Y. S. Lee, K. Lee, S. E. Moore, A. J. Fulford, A. M. Prentice, R. A. Waterland, Independent genome-wide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. *Genome Biol.* **16**, 118 (2015).
26. P. Kühnen, D. Handke, R. A. Waterland, B. J. Hennig, M. Silver, A. J. Fulford, P. Dominguez-Salas, S. E. Moore, A. M. Prentice, J. Spranger, A. Hinney, J. Hebebrand, F. L. Heppner, L. Walzer, C. Gröttinger, J. Gromoll, S. Wiegand, A. Grüters, H. Krude, Interindividual variation in DNA methylation at a putative POMC metastable epiallele is associated with obesity. *Cell Metab.* **24**, 502–509 (2016).
27. S. Finer, M. S. Iqbal, R. Lowe, B. W. Ogunkolade, S. Pervin, C. Mathews, M. Smart, D. S. Alam, G. A. Hitman, Is famine exposure during developmental life in rural Bangladesh associated with a metabolic and epigenetic signature in young adulthood? A historical cohort study. *BMJ Open* **6**, e011768 (2016).
28. R. A. Waterland, R. L. Jirtle, Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* **23**, 5293–5300 (2003).
29. P. Dominguez-Salas, S. E. Moore, M. S. Baker, A. W. Bergen, S. E. Cox, R. A. Dyer, A. J. Fulford, Y. Guan, E. Laritsky, M. J. Silver, G. E. Swan, S. H. Zeisel, S. M. Innis, R. A. Waterland, A. M. Prentice, B. J. Hennig, Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat. Commun.* **5**, 3746 (2014).
30. R. A. Waterland, K. B. Michels, Epigenetic epidemiology of the developmental origins hypothesis. *Annu. Rev. Nutr.* **27**, 363–388 (2007).
31. R. A. Waterland, R. Kellermayer, E. Laritsky, P. Rayco-Solon, R. A. Harris, M. Travisano, W. Zhang, M. S. Torskaya, J. Zhang, L. Shen, M. J. Manary, A. M. Prentice, Season of conception in rural gambia affects DNA methylation at putative human metastable epialleles. *PLoS Genet.* **6**, e1001252 (2010).
32. R. Abu-Khudir, F. Magne, J.-P. Chanoine, C. Deal, G. Van Vliet, J. Deladoëy, Role for tissue-dependent methylation differences in the expression of FOXE1 in nontumoral thyroid glands. *J. Clin. Endocrinol. Metab.* **99**, E1120–E1129 (2014).
33. P. James, S. Sajjadi, A. S. Tomar, A. Saffari, C. H. D. Fall, A. M. Prentice, S. Shrestha, P. Issarapu, D. K. Yadav, L. Kaur, K. Lillycrop, M. Silver, G. R. Chandak; EMPHASIS Study group, Candidate genes linking maternal nutrient exposure to offspring health via DNA methylation: A review of existing evidence in humans with specific focus on one-carbon metabolism. *Int. J. Epidemiol.* **47**, 1910–1937 (2018).
34. P. Dominguez-Salas, S. E. Moore, D. Cole, K. A. Da Costa, S. E. Cox, R. A. Dyer, A. J. C. Fulford, S. M. Innis, R. A. Waterland, S. H. Zeisel, A. M. Prentice, B. J. Hennig, DNA methylation potential: Dietary intake and blood concentrations of one-carbon metabolites and cofactors in rural African women. *Am. J. Clin. Nutr.* **97**, 1217–1227 (2013).
35. R. C. Richmond, G. C. Sharp, G. Herbert, C. Atkinson, C. Taylor, S. Bhattacharya, D. Campbell, M. Hall, N. Kazmi, T. Gaunt, W. McArdle, S. Ring, G. D. Smith, A. Ness, C. L. Relton, The long-term impact of folic acid in pregnancy on offspring DNA methylation: Follow-up of the Aberdeen folic acid supplementation trial (AFASST). *Int. J. Epidemiol.* **47**, 928–937 (2018).
36. A. Saffari, S. Shrestha, P. Issarapu, S. Sajjadi, M. Betts, S. A. Sahariah, A. S. Tomar, P. James, A. Dedaniya, D. K. Yadav, K. Kumaran, A. M. Prentice, K. A. Lillycrop, C. H. D. Fall, G. R. Chandak, M. J. Silver; EMPHASIS Study Group, Effect of maternal preconceptional and pregnancy micronutrient interventions on children's DNA methylation: Findings from the EMPHASIS study. *Am. J. Clin. Nutr.* **112**, 1099–1113 (2020).
37. S. E. Moore, A. J. C. Fulford, M. K. Darboe, M. L. Jobarteh, L. M. Jarjou, A. M. Prentice, A randomized trial to investigate the effects of pre-natal and infant nutritional supplementation on infant immune development in rural Gambia: The ENID trial: Early Nutrition and Immune Development. *BMC Pregnancy Childbirth* **12**, 107 (2012).
38. GTex Consortium, The genotype-tissue expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
39. P. N. Taylor, E. Porcu, S. Chew, P. J. Campbell, M. Traglia, S. J. Brown, B. H. Mullin, H. A. Shihab, J. Min, K. Walter, Y. Memari, J. Huang, M. R. Barnes, J. P. Beilby, P. Charoen, P. Danecsek, F. Dudbridge, V. Forgetta, C. Greenwood, E. Grundberg, A. D. Johnson, J. Hui, E. M. Lim, S. McCarthy, D. Muddyman, V. Panicker, J. R. B. Perry, J. T. Bell, W. Yuan, C. Relton, T. Gaunt, D. Schlessinger, G. Abecasis, J. C. A. G. L. Surdulescu, W. Woltersdorf, E. Zeggini, H.-F. Zheng, D. Toniolo, C. M. Dayan, S. Naitza, J. P. Walsh, T. Spector, G. D. Smith, R. Durbin, J. B. Richards, S. Sanna, N. Soranzo, N. J. Timpson,

- S. G. Wilson, S. Al Turki, C. Anderson, R. Anney, D. Antony, M. S. Artigas, M. Ayub, S. Balasubramaniam, J. C. Barrett, I. Barroso, P. Beales, J. Benthams, S. Bhattacharya, E. Birney, D. Blackwood, M. Bobrow, E. Bochukova, P. Bolton, R. Bounds, C. Boustred, G. Breen, M. Calissano, K. Carss, K. Chatterjee, L. Chen, A. Ciampi, S. Cirak, P. Clapham, G. Clement, G. Coates, D. Collier, C. Cosgrove, T. Cox, N. Craddock, L. Crooks, S. Curran, D. Curtis, A. Daly, A. Day-Williams, I. N. M. Day, T. Down, Y. Du, I. Dunham, S. Edkins, P. Ellis, D. Evans, S. Faraogi, G. Fatemifar, D. R. Fitzpatrick, P. Flicek, J. Floyd, A. R. Foley, C. S. Franklin, M. Futema, L. Gallagher, M. Geihs, D. Geschwind, H. Griffin, D. Grozeva, X. Guo, X. Guo, H. Gurling, D. Hart, A. Hendricks, P. Holmans, B. Howie, L. Huang, T. Hubbard, S. E. Humphries, M. E. Hurles, P. Hysi, D. K. Jackson, Y. Jamshidi, T. Jing, C. Joyce, J. Kaye, T. Keane, J. Kemp, J. Kemp, K. Kennedy, A. Kolb-Kococinski, G. Lachance, C. Langford, D. Lawson, I. Lee, M. Lek, J. Liang, H. Lin, R. Li, Y. Li, R. Liu, J. Lönngqvist, M. Lopes, V. Iotchkova, D. MacArthur, J. Marchini, J. Maslen, M. Massimo, I. Mathieson, G. Marenne, P. McGuffin, A. McIntosh, A. G. McKechnie, A. McQuillin, S. Metrustry, H. Mitchell, A. Moayyeri, J. Morris, F. Muntioni, K. Northstone, M. O'Donovan, A. Onoufriadis, S. O'Rahilly, K. Ouakacha, M. J. Owen, A. Palotie, K. Panoutsopoulou, V. Parker, J. R. Parr, L. Paternoster, T. Paunio, F. Payne, O. Pietiläinen, V. Plagnol, L. Quaye, M. A. Quail, L. Raymond, K. Rehnström, S. Ring, G. R. S. Ritchie, N. Roberts, D. B. Savage, P. Scambler, S. Schiffls, M. Schmidt, N. Schoenmakers, R. K. Semple, E. Serra, S. I. Sharp, S.-Y. Shin, D. Skuse, K. Small, L. Southam, O. Spasic-Boskovic, D. S. Clair, J. Stalker, E. Stevens, B. S. Pourcain, J. Sun, J. Suvisaari, I. Tachmazidou, M. D. Tobin, A. Valdes, M. Van Kogelenberg, P. Vijayarangakannan, P. M. Visscher, L. V. Wain, J. T. R. Walters, G. Wang, J. Wang, Y. Wang, K. Ward, E. Wheeler, T. Whyte, H. Williams, K. A. Williamson, C. Wilson, K. Wong, C. Xu, J. Yang, F. Zhang, P. Zhang; UK10K Consortium, Whole-genome sequence-based analysis of thyroid function. *Nat. Commun.* **6**, 5681 (2015).
40. A. Teumer, R. Rawal, G. Homuth, F. Ernst, M. Heier, M. Evert, F. Dombrowski, U. Völker, M. Nauck, D. Radke, T. Ittermann, R. Biffar, A. Döring, C. Gieger, N. Klopp, H.-E. Wichmann, H. Wallaschofski, C. Meisinger, H. Völzke, Genome-wide association study identifies four genetic loci associated with thyroid volume and goiter risk. *Am. J. Hum. Genet.* **88**, 664–673 (2011).
41. N. Lafontaine, P. J. Campbell, J. E. Castillo-Fernandez, S. Mullin, E. M. Lim, P. Kendrew, M. Lewer, S. J. Brown, R.-C. Huang, P. E. Melton, T. A. Mori, L. J. Beilin, F. Dudbridge, T. D. Spector, M. J. Wright, N. G. Martin, A. F. McRae, V. Panicker, G. Zhu, J. P. Walsh, J. T. Bell, S. G. Wilson, Epigenome-wide association study of thyroid function traits identifies novel associations of FT3 With KLF9 and DOT1L. *J. Clin. Endocrinol. Metab.* **106**, e2191–e2202 (2021).
42. P. T. James, P. Dominguez-Salas, B. J. Hennig, S. E. Moore, A. M. Prentice, M. J. Silver, Maternal one-carbon metabolism and infant DNA methylation between contrasting seasonal environments: A case study from the gambia. *Curr. Dev. Nutr.* **3**, nzy082 (2019).
43. M. E. Brosnan, L. MacMillan, J. R. Stevens, J. T. Brosnan, Division of labour: How does folate metabolism partition between one-carbon metabolism and amino acid oxidation? *Biochem. J.* **472**, 135–146 (2015).
44. L. Rios-Avila, B. Coats, Y.-Y. Chi, Ø. Midttun, P. M. Ueland, P. W. Stacpoole, J. F. Gregory III, Metabolite profile analysis reveals association of vitamin B-6 with metabolites related to one-carbon metabolism and tryptophan catabolism but not with biomarkers of inflammation in oral contraceptive users and reveals the effects of oral contraceptives. *J. Nutr.* **145**, 87–95 (2015).
45. C. G. Howe, B. Cox, R. Fore, J. Jungius, T. Kvist, S. Lent, H. E. Miles, L. A. Salas, S. Rifas-Shiman, A. P. Starling, P. Yousefi, C. Ladd-Acosta, A. Baccarelli, E. B. Binder, V. L. Chatzi, D. Czamara, D. Dabelea, D. L. DeMeo, A. Ghanous, Z. Herczeg, E. Kajantie, J. M. T. Lahti, D. A. Lawlor, A. Litonjua, T. S. Nawrot, E. A. Nohr, E. Oken, C. Pizzi, M. Plusquin, K. Rääkkönen, C. L. Relton, G. C. Sharp, T. I. A. Sorensen, J. Sunyer, M. Vrijheid, W. Zhang, M.-F. Hivert, C. V. Breton, Maternal gestational diabetes mellitus and newborn DNA methylation: Findings from the pregnancy and childhood epigenetics consortium. *Diabetes Care* **16**, dc190524 (2019).
46. A. L. Teh, H. Pan, L. Chen, M.-L. Ong, S. Dogra, J. Wong, J. L. MacIsaac, S. M. Mah, L. M. McEwen, S.-M. Saw, K. M. Godfrey, Y.-S. Chong, K. Kwek, C.-K. Khoo, S.-E. Soh, M. F. F. Chong, S. Barton, N. Karnani, C. Y. Cheong, J. P. Buschdorf, W. Stünkel, M. S. Kobar, M. J. Meaney, P. D. Gluckman, J. D. Holbrook, The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res.* **24**, 1064–1074 (2014).
47. C. J. Gunasekara, C. A. Scott, E. Laritsky, M. S. Baker, H. MacKay, J. D. Duryea, N. J. Kessler, G. Hellenthal, A. C. Wood, K. R. Hodges, M. Gandhi, A. B. Hair, M. J. Silver, S. E. Moore, A. M. Prentice, Y. Li, R. Chen, C. Coarfa, R. A. Waterland, A genomic atlas of systemic interindividual epigenetic variation in humans. *Genome Biol.* **20**, 105 (2019).
48. R. A. Waterland, D. C. Dolinoy, J. R. Lin, C. A. Smith, X. Shi, K. G. Tahiliani, Maternal methyl supplements increase offspring DNA methylation at Axin fused. *Genesis* **44**, 401–406 (2006).
49. M. V. Rastogi, S. H. LaFranchi, Congenital hypothyroidism. *Orphanet J. Rare Dis.* **5**, 17 (2010).
50. E. Szczepanek-Parulska, A. Zybek-Kocik, L. Wartofsky, M. Ruchala, Thyroid hemiagenesis: Incidence, clinical significance, and genetic background. *J. Clin. Endocrinol. Metab.* **102**, 3124–3137 (2017).
51. C. G. Ma, Y. S. Shim, Association of thyroid-stimulating hormone and thyroid hormones with cardiometabolic risk factors in euthyroid children and adolescents aged 10–18 years: A population-based study. *Sci. Rep.* **9**, 15476 (2019).
52. M. R. Licenziati, G. Valerio, I. Vetrani, G. De Maria, F. Liotta, G. Radetti, Altered thyroid function and structure in children and adolescents who are overweight and obese: Reversal after weight loss. *J. Clin. Endocrinol. Metab.* **104**, 2757–2765 (2019).
53. M. A. Emokpae, P. A. Obazelu, The association of triiodothyronine-to-thyroxine ratio with body mass index in obese nigerian children and adolescents. *Med. Sci.* **5**, 36 (2017).
54. M. Barjaktarovic, T. I. M. Korevaar, R. Gaillard, Y. B. de Rijke, T. J. Visser, V. W. V. Jaddoe, R. P. Peeters, Childhood thyroid function, body composition and cardiovascular function. *Eur. J. Endocrinol.* **177**, 319–327.
55. C. D. Wirth, M. R. Blum, B. R. da Costa, C. Baumgartner, T.-H. Collet, M. Medici, R. P. Peeters, D. Aujesky, D. C. Bauer, N. Rodondi, Subclinical thyroid dysfunction and the risk for fractures: A systematic review and meta-analysis. *Ann. Intern. Med.* **161**, 189–199 (2014).
56. J. H. D. Bassett, G. R. Williams, Role of thyroid hormones in skeletal development and bone maintenance. *Endocr. Rev.* **37**, 135–187 (2016).
57. O. Alvarez-Garcia, K. M. Fisch, N. E. Wineinger, R. Akagi, M. Saito, T. Sasho, A. I. Su, M. K. Lotz, Increased DNA methylation and reduced expression of transcription factors in human osteoarthritis cartilage. *Arthritis Rheum.* **68**, 1876–1886 (2016).
58. Y. Guo, Q. Sheng, J. Li, F. Ye, D. C. Samuels, Y. Shyr, Large scale comparison of gene expression levels by microarrays and RNAseq Using TCGA data. *PLOS ONE* **8**, e71462 (2013).
59. M. Kulis, A. C. Queirós, R. Beekman, J. I. Martin-Subero, Intragenic DNA methylation in transcriptional regulation, normal differentiation and cancer. *Biochim. Biophys. Acta* **1829**, 1161–1174 (2013).
60. J. R. Wagner, S. Busche, B. Ge, T. Kwan, T. Pastinen, M. Blanchette, The relationship between DNA methylation, genetic and expression inter-individual variation in untransformed human fibroblasts. *Genome Biol.* **15**, R37 (2014).
61. A. P. Boyle, E. L. Hong, M. Hariharan, Y. Cheng, M. A. Schaub, M. Kasowski, K. J. Karczewski, J. Park, B. C. Hitz, S. Weng, J. M. Cherry, M. Snyder, Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
62. G. Bahari, M. Hashemi, M. Naderi, S. Sadeghi-Bojd, M. Taheri, Long non-coding RNA PAX8-AS1 polymorphisms increase the risk of childhood acute lymphoblastic leukemia. *Biomed. Rep.* **8**, 184–190 (2018).
63. M. B. Zimmermann, I. Aeberli, M. Andersson, V. Assay, J. A. J. Yorg, P. Jooste, T. Jukić, D. Kartono, Z. Kusić, E. Pretell, T. O. L. S. Luis Jr., J. Untoro, A. Timmer, Thyroglobulin is a sensitive measure of both deficient and excess iodine intakes in children and indicates no adverse effects on thyroid function in the IUC range of 100–299 µg/L: A UNICEF/ICCIDD study group report. *J. Clin. Endocrinol. Metab.* **98**, 1271–1280 (2013).
64. K. G. Eriksen, M. Andersson, S. Hunziker, M. B. Zimmermann, S. E. Moore, Effects of an iodine-containing prenatal multiple micronutrient on maternal and infant iodine status and thyroid function: A randomized trial in the gambia. *Thyroid* **30**, 1355–1365 (2020).
65. O. P. Soldin, Controversies in urinary iodine determinations. *Clin. Biochem.* **35**, 575–579 (2002).
66. C. J. Gunasekara, R. A. Waterland, A new era for epigenetic epidemiology. *Epigenomics* **11**, 1647–1649 (2019).
67. P.-P. Lv, S. Tian, C. Feng, J.-Y. Li, D.-Q. Yu, L. Jin, Y. Shen, T.-T. Yu, Y. Meng, G.-L. Ding, M. Jin, X.-J. Chen, J.-Z. Sheng, D. Zhang, H.-F. Huang, Maternal high estradiol exposure is associated with elevated thyroxine and Pax8 in mouse offspring. *Sci. Rep.* **6**, 36805 (2016).
68. S. G. Okala, M. K. Darboe, F. Sosseh, B. Sonko, T. Faye-Joof, A. M. Prentice, S. E. Moore, Impact of nutritional supplementation during pregnancy on antibody responses to diphtheria-tetanus-pertussis vaccination in infants: A randomised trial in The Gambia. *PLoS Med.* **16**, e1002854 (2019).
69. K. G. Eriksen, W. Johnson, B. Sonko, A. M. Prentice, M. K. Darboe, S. E. Moore, Following the World Health Organization's recommendation of exclusive breastfeeding to 6 months of age does not impact the growth of rural gambian infants. *J. Nutr.* **147**, 248–255 (2017).
70. M. P. F. Alicia Alonso, T. J. Kacmarczyk, "Agilent SureSelectXT Methyl-Seq Applications with Low-Input DNA and Smaller Capture Libraries" (2017).
71. F. Krueger, S. R. Andrews, Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
72. J. Brunn, U. Block, G. Ruf, I. Bos, W. P. Kunze, P. C. Scriba, Volumetric analysis of thyroid lobes by real-time ultrasound (author's transl). *Dtsch. Med. Wochenschr.* **106**, 1338–1340 (1981).
73. WHO (World Health Organization), WHO Anthro for personal computers. Software for assessing growth and development of the world's children (2011).

74. N. J. Crabtree, A. Arabi, L. K. Bachrach, M. Fweltrell, G. El-Hajj Fuleihan, H. H. Keckskemethy, M. Jaworski, C. M. Gordon, Dual-energy X-ray absorptiometry interpretation and reporting in children and adolescents: The revised 2013 ISCD Pediatric Official Positions. *J. Clin. Densitom.* **17**, 225–242 (2014).
75. N. Kessler, knowah/bssnper2: Initial release (2021), doi:10.5281/ZENODO.4647013.
76. S. Gao, D. Zou, L. Mao, H. Liu, P. Song, Y. Chen, S. Zhao, C. Gao, X. Li, Z. Gao, X. Fang, H. Yang, T. F. Ørntoft, K. D. Sørensen, L. Bolund, BS-SNPPer: SNP calling in bisulfite-seq data. *Bioinformatics* **31**, 4006–4008 (2015).
77. G. Band, Q. S. Le, G. M. Clarke, K. Kivinen, C. Hubbard, A. E. Jeffreys, K. Rowlands, E. M. Leffler, M. Jallow, D. J. Conway, F. Sisay-Joof, G. Sirugo, U. d'Alessandro, O. B. Toure, M. A. Thera, S. Konate, S. Sissoko, V. D. Mangano, E. C. Bougouma, S. B. Sirima, L. N. Amenga-Etego, A. K. Ghansah, A. V. O. Hodgson, M. D. Wilson, A. Enimil, D. Ansong, J. Evans, S. A. Ademola, T. O. Apinjoh, C. M. Ndila, A. Manjurano, C. Drakeley, H. Reyburn, N. H. Phu, N. T. N. Quyen, C. Q. Thai, T. T. Hien, Y. Y. Teo, L. Manning, M. Laman, P. Michon, H. Karunajeewa, P. Siba, S. Allen, A. Allen, M. Bahlo, T. M. E. Davis, V. Simpson, J. Shelton, C. C. A. Spencer, G. B. J. Busby, A. Kerasidou, E. Drury, J. Stalker, A. Dilthey, A. J. Mentzer, G. McVean, K. A. Bojang, O. Doumbo, D. Modiano, K. A. Koram, T. Agbenyega, O. K. Amodu, E. Achidi, T. N. Williams, K. Marsh, E. M. Riley, M. Molyneux, T. Taylor, S. J. Dunstan, J. Farrar, I. Mueller, K. A. Rockett, D. P. Kwiatkowski, Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. *Nat. Commun.* **10**, 5732 (2019).
78. A. Auton, G. R. Abecasis, D. M. Altshuler, R. M. Durbin, G. R. Abecasis, D. R. Bentley, A. Chakravarti, A. G. Clark, P. Donnelly, E. E. Eichler, P. Flicek, S. B. Gabriel, R. A. Gibbs, D. Green, M. E. Hurles, B. M. Knoppers, J. O. Korbel, E. S. Lander, C. Lee, H. Leirach, E. R. Mardis, G. T. Marth, G. A. McVean, D. A. Nickerson, J. P. Schmidt, S. T. Sherry, J. Wang, R. K. Wilson, R. A. Gibbs, E. Boerwinkle, H. Doddapaneni, Y. Han, V. Korchina, C. Kovar, S. Lee, D. Muzny, J. G. Reid, Y. Zhu, J. Wang, Y. Chang, Q. Feng, X. Fang, X. Guo, M. Jian, H. Jiang, X. Jin, T. Lan, G. Li, J. Li, Y. Li, S. Liu, X. Liu, Y. Lu, X. Ma, M. Tang, B. Wang, G. Wang, H. Wu, R. Wu, X. Xu, Y. Yin, D. Zhang, W. Zhang, J. Zhao, M. Zhao, X. Zheng, E. S. Lander, D. M. Altshuler, S. B. Gabriel, N. Gupta, N. Gharani, L. H. Toji, N. P. Gerry, A. M. Resch, P. Flicek, J. Barker, L. Clarke, L. Gil, S. E. Hunt, G. Kelman, E. Kulesha, R. Leinonen, W. M. McLaren, R. Radhakrishnan, A. Roa, D. Smirnov, R. E. Smith, I. Streeter, A. Thormann, I. Toneva, B. Vaughan, X. Zheng-Bradley, D. R. Bentley, R. Grocock, S. Humphray, T. James, Z. Kingsbury, H. Leirach, R. Sudbrak, M. W. Albrecht, V. S. Amstislavskiy, T. A. Borodina, M. Lienhard, F. Mertes, M. Sultan, B. Timmermann, M.-L. Yaspo, E. R. Mardis, R. K. Wilson, L. Fulton, R. Fulton, S. T. Sherry, V. Ananiev, D. Beloslyudtsev, N. Bouk, C. Chen, D. Church, R. Cohen, C. Cook, J. Garner, T. Hefferon, M. Kimelman, C. Liu, J. Lopez, P. Meric, C. O'Sullivan, Y. Ostapchuk, L. Phan, S. Ponomarov, V. Schneider, E. Shekhtman, K. Sirotkin, D. Slotta, H. Zhang, G. A. McVean, R. M. Durbin, S. Balasubramaniam, J. Burton, P. Danecek, T. M. Keane, A. Kolb-Kocokinski, S. McCarthy, J. Stalker, M. Quail, J. P. Schmidt, C. J. Davies, J. Gollub, T. Webster, B. Wong, Y. Zhan, A. Auton, C. L. Campbell, Y. Kong, A. Marcketta, R. A. Gibbs, F. Yu, L. Antunes, M. Bainbridge, D. Muzny, A. Sabo, Z. Huang, J. Wang, L. J. M. Coin, L. Fang, X. Guo, X. Jin, G. Li, Q. Li, Y. Li, Z. Li, H. Lin, B. Liu, R. Luo, H. Shao, Y. Xie, C. Ye, C. Yu, F. Zhang, H. Zheng, H. Zhu, C. Alkan, E. Dal, F. Kahveci, G. T. Marth, E. P. Garrison, D. Kural, W.-P. Lee, W. F. Leong, M. Stromberg, A. N. Ward, J. Wu, M. Zhang, M. J. Daly, M. A. DePristo, R. E. Handsaker, D. M. Altshuler, E. Banks, G. Bhatia, G. del Angel, S. B. Gabriel, G. Genovese, N. Gupta, H. Li, S. Kashin, E. S. Lander, S. A. McCarroll, J. C. Nemes, R. E. Poplin, S. C. Yoon, J. Lihm, V. Makarov, A. G. Clark, S. Gottipati, A. Keinan, J. L. Rodriguez-Flores, J. O. Korbel, T. Rausch, M. H. Fritz, A. M. Stütz, P. Flicek, K. Beal, L. Clarke, A. Datta, J. Herrero, W. M. McLaren, G. R. S. Ritchie, R. E. Smith, D. Zerbino, X. Zheng-Bradley, P. C. Sabeti, I. Shlyakhter, S. F. Schaffner, J. Vitti, D. N. Cooper, E. V. Ball, P. D. Stenson, D. R. Bentley, B. Barnes, M. Bauer, R. A. C. Cheetham, A. Cox, M. Eberle, S. Humphray, S. Kahn, L. Murray, J. Peden, R. Shaw, E. E. Kenny, M. A. Batzer, M. K. Konkel, J. A. Walker, D. G. MacArthur, M. Lek, R. Sudbrak, V. S. Amstislavskiy, R. Herwig, E. R. Mardis, L. Ding, D. C. Koboldt, D. Larson, K. Ye, S. Gravel, T. 1000 G. P. Consortium; C. authors; S. committee; P. group; B. C. of Medicine; BGI-Shenzhen; B. I. of M. I. T. and Harvard; C. I. for M. Research; E. B. I. European Molecular Biology Laboratory; Illumina, M. P. I. for M. Genetics; M. G. I. at W. University; U. S. N. I. of Health, U. of Oxford; W. T. S. Institute; A. group, Affymetrix; A. E. C. of Medicine, B. University, B. College, C. S. H. Laboratory; C. University, E. M. B. Laboratory, H. University, H. G. M. Database, I. S. of M. at M. Sinai, L. S. University; M. G. Hospital, M. University, N. I. H. National Eye Institute, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
79. S. E. Moore, T. J. Cole, A. C. Collinson, E. M. Poskitt, I. A. McGregor, A. M. Prentice, Prenatal or early postnatal events predict infectious deaths in young adulthood in rural Africa. *Int. J. Epidemiol.* **28**, 1088–1095 (1999).
80. R Core Team, R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*. **0** (2013), p. (ISBN) 3–900051–07–0; doi:www.R-project.org/.

Acknowledgments: We thank the women and children of West Kiang who patiently participated in the ENID trial and the participants in this study. We acknowledge the enthusiastic work of the study teams, especially the fieldworkers, village assistants, midwives, clinical staff, data office staff, and laboratory technicians who collected the data and samples. We would like to thank A. Quansah (Consultant Radiologist, Komfo Anokye Teaching Hospital, Ghana), P. Lewis (Royal College of Radiologists, UK), J. Kasznia-Brown (Royal College of Radiologists, UK), and E. Joekees (World Wide Radiology) for support in arranging thyroid ultrasound training for TC in The Gambia. We would like to thank C. Evans and team at the University Hospital of Wales for work processing the thyroid function tests; Y. Li and R. Chen (Human Genome Sequencing Center, Baylor College of Medicine) for performing and troubleshooting the target capture and Illumina sequencing; A. Prentice and her team (Head of Calcium, Vitamin D, Bone Health group at University of Cambridge and MRC Unit in the Gambia at LSHTM) for facilitating the use of the DXA facilities at Keneba Field Station; M. Ceasay, M. Mendy, M. Jammeh, and K. Sawo at MRC The Gambia, Keneba Field station for performing the DXA scans; E. Sise (MRC The Gambia at LSHTM) who performed the AGP analysis in Keneba; A. Ferguson-Smith for discussions regarding the functional implications of indel and SNP variants; and P. Kuhnen for discussions regarding *PAX8* methylation and CHT. This publication uses data from the GGVP, a collaboration between the MRC Unit in the Gambia (www.mrc.gm), the Wellcome Sanger Institute (www.sanger.ac.uk), the MRC Centre for Genomics and Global Health (www.cggh.org/collaborations/mrc-unit-the-gambia) at the University of Oxford, and the MalariaGEN Resource Centre (www.malariagen.net). The *PAX8* and *PAX8-AS1* methylation and expression results here are in part based upon data generated by the TCGA Research Network: www.cancer.gov/tcga. **Funding:** T.C. is funded by an MRC Clinical Research Training Fellowship (MR/S006516/1). The British Society of Paediatric Endocrinology and Diabetes provided funding for this work to T.C. through their Research and Innovation Award 2018. The Gambian studies are supported by core funding from the U.K. Medical Research Council (MRC) MC-A760-5QX00 and by MRC grant MR/M01424X/1. Programme funding for A.P. supported data acquisition and analysis of the whole-body DXA scans (U105960371 and U123261351). **Author contributions:** Conceptualization: T.C., M.L., A.P., M.S. Methodology: T.C., N.K., K.W., D.J., and M.S. Investigation: T.C., C.G., P.J., E.L., M.B., R.D., R.E., R.W., and S.M. Formal analysis: T.C. and N.K. Visualization: T.C., N.K., C.G., R.W., A.P., and M.S. Supervision: M.L., A.P., and M.S. Writing—original draft: T.C. and M.S. Writing—review and editing: All authors reviewed and contributed to the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Data from TCGA (www.cancer.gov/tcga) and GTEx (www.gtexportal.org/home/) used in this analysis are publicly available. N.K.'s code for the modified version of BS-SnpPer is available on Zenodo: <https://doi.org/10.5281/zenodo.4647020>. Requests to access and analyze the Gambian data should be submitted to the corresponding author in the first instance. An application would then need to be made to MRC Unit The Gambia's Scientific Coordinating Committee and the Joint MRC/Gambia Government Ethics Committee. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 23 April 2021
 Accepted 16 September 2021
 Published 5 November 2021
 10.1126/sciadv.abj1561

DNA methylation at a nutritionally sensitive region of the *PAX8* gene is associated with thyroid volume and function in Gambian children

Toby CandlerNoah KesslerChathura GunasekaraKate WardPhilip JamesEleonora LaritskyMaria BakerRoger DyerRajavel ElangoDavid JeffriesRobert WaterlandSophie MooreMarian LudgateAndrew PrenticeMatt Silver

Sci. Adv., 7 (45), eabj1561. • DOI: 10.1126/sciadv.abj1561

View the article online

<https://www.science.org/doi/10.1126/sciadv.abj1561>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of think article is subject to the [Terms of service](#)