

RESEARCH PAPER



Adiposity associated DNA methylation signatures in adolescents are related to leptin and perinatal factors

R.C. Huang^a, P.E. Melton^{b,c,d}, M.A. Burton^e, L.J. Beilin^f, R. Clarke-Harris^g, E. Cook^e, K.M. Godfrey^h, G. C. Burdge^e, T.A. Morif, D. Anderson^g, S. Rauschert^a, J. M. Craig^{g,h}, M.S. Koborⁱ, J.L. MacIsaac^j, A.M. Morinⁱ, W. H. Oddy^d, C.E. Pennell^j, J.D. Holbrook^b, and K.A. Lillycrop^b

^aTelethon Kids Institute, University of Western Australia, Australia; ^bCurtin/UWA Centre for Genetic Origins of Health and Disease, School of Biomedical Sciences, The University of Western Australia, Perth, Australia; ^cSchool of Pharmacy and Biomedical Sciences, Curtin University, Perth, Australia; ^dMenzies Institute for Medical Research, University of Tasmania, Australia; ^eSchool of Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK; ^fMedical School, The University of Western Australia, Australia; ^gCentre for Molecular and Medical Research, School of Medicine, Deakin University, Geelong, Victoria, Australia; ^hEnvironmental & Genetic Epidemiology Research, Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia; ⁱDepartment of Medical Genetics, University of British Columbia, Vancouver Canada; ^jSchool of Medicine and Public Health, Faculty of Medicine and Health, The University of Newcastle, Australia

ABSTRACT

Epigenetics links perinatal influences with later obesity. We identified differentially methylated CpG (dmCpG) loci measured at 17 years associated with concurrent adiposity measures and examined whether these were associated with hsCRP, adipokines, and early life environmental factors. Genome-wide DNA methylation from 1192 Raine Study participants at 17 years, identified 29 dmCpGs (Bonferroni corrected $p < 1.06E-07$) associated with body mass index (BMI), 10 with waist circumference (WC) and 9 with subcutaneous fat thickness. DmCpGs within Ras Association (RalGDS/AF-6), Pleckstrin Homology Domains 1 (*RAPH1*), Musashi RNA-Binding Protein 2 (*MSI2*), and solute carrier family 25 member 10 (*SLC25A10*) are associated with both BMI and WC. Validation by pyrosequencing confirmed these associations and showed that *MSI2*, *SLC25A10*, and *RAPH1* methylation was positively associated with serum leptin. These were also associated with the early environment; *MSI2* methylation ($\beta = 0.81$, $p = 0.0004$) was associated with pregnancy maternal smoking, *SLC25A10* (CpG2 $\beta = 0.12$, $p = 0.002$) with pre- and early pregnancy BMI, and *RAPH1* ($\beta = -1.49$, $p = 0.036$) with gestational weight gain. Adjusting for perinatal factors, methylation of the dmCpGs within *MSI2*, *RAPH1*, and *SLC25A10* independently predicted BMI, accounting for 24% of variance. *MSI2* methylation was additionally associated with BMI over time (17 years old $\beta = 0.026$, $p = 0.0025$; 20 years old $\beta = 0.027$, $p = 0.0029$) and between generations (mother $\beta = 0.044$, $p = 7.5E-04$). Overall findings suggest that DNA methylation in *MSI2*, *RAPH1*, and *SLC25A10* in blood may be robust markers, mediating through early life factors.

ARTICLE HISTORY

Received 24 July 2020
Revised 4 December 2020
Accepted 4 January 2021

KEYWORDS

DNA methylation; body Composition; inflammation; high-sensitivity C-reactive Protein; adolescence; youth; leptin; epigenetic Variation; maternal; body mass index; adiposity; the Raine Study

Introduction

Studies suggest that susceptibility to obesity in later life may be influenced by the early life environment. In animal models, variations in maternal body composition and diet can induce persistent changes in metabolism and body composition in the offspring [1], while in humans, higher maternal BMI and excessive gestational weight gain (GWG) independently increase obesity risk in the offspring [2].

Mechanisms by which the environment influences adiposity are suggested to involve the epigenetic regulation of genes. In animal models, variations in

maternal diet induced changes in the methylation of key metabolic genes, which often persist through the lifecourse, accompanied by changes in gene expression and the metabolic processes they control [3]. Human studies also suggest a role for epigenetics in mediating the effects of the early life environment on later disease risk. For example, maternal carbohydrate intake and offspring adiposity have been associated with differential methylation of the promoter of the *RXRA* gene in the umbilical cord [4], while perinatal methylation of CpGs within the promoter of the long non-coding RNA ANRIL is associated with antenatal growth faltering and higher adiposity at age 6 years

CONTACT R.C. Huang ✉ rae-chi.huang@telethonkids.org.au 📍 Telethon Kids Institute, Monash Avenue, Nedlands, WA 6009, Australia

📎 Supplemental data for this article can be accessed [here](#).

© 2021 Informa UK Limited, trading as Taylor & Francis Group

[5]. Although initial studies suggest that the epigenome is most susceptible to environmental factors in early life, there is increasing evidence that the epigenome retains a degree of plasticity throughout life [6]. Thus, modulation of the epigenome may represent a key mechanism by which the environment influences disease risk across the lifecourse.

Epigenome-wide DNA methylation association studies (EWAS) have identified differentially methylated CpG (dmCpG) loci in blood that are associated with concurrent adiposity [3,7–18] and related measures of inflammation [19,20]. Most studies have been carried out during adulthood, rather than in adolescence/childhood [3,16–18] when the influences of early life exposures and plasticity of the epigenome may be stronger. Few studies have explored the longitudinal nature of these associations or the associations across generations. Here, we have undertaken EWAS on whole blood from 1192 participants of the Raine Study (www.rainestudy.org.au) at age 17 years (generation 2), to identify dmCpGs associated with concurrent measures of adiposity in adolescence, and their relationship with early life environmental factors. Subsequent pyrosequencing of the adiposity-associated dmCpGs at 17 and 20 years of age, and in matched mother–child pairs, was undertaken in order to determine whether variation in CpG methylation levels was associated with multiple measures of adiposity, serum levels of the adipokine leptin and adiponectin, and C-reactive protein as a marker of inflammation. We also assessed whether the associations between DNA methylation levels and anthropometry remain stable over a 3-year time period in the same individuals and consistent between mother and offspring. Further, we assessed stability with adjustments for genetic and early life factors.

Methods

The Raine study

The Raine Study (www.rainestudy.org.au) recruited pregnant women between 1989 and 1991 ($N = 2,900$) in Perth, Australia. Women were invited if they were between 16 and 20 weeks pregnant, had sufficient proficiency in English, were expected to deliver at the hospital, and intended to remain in the jurisdiction [21]. There were 2868 live births (including 2 742 singletons, 60 sets of twins, and 2 sets of

triplets.) By the 17-year follow-up, there had been attrition (36 deaths, 480 withdrawn, 598 lost or deferred) [22].

Maternal (Gen1) data

Maternal weight was obtained from medical records at mean 16.5 ± 2.2 and mean 34.1 ± 1.5 weeks' gestation. GWG was defined as the weight gained from 16 to 34 weeks. Maternal smoking during pregnancy was assessed from questionnaires (yes/no) (validated against serum cotinine) [23], and a stress score ascertained using the 11-item questionnaire derived from the Life Stress Inventory at 18 and 34 week gestation. Total life event index scores were calculated throughout pregnancy from the sum of events [24]. Breast feeding duration was obtained from questionnaires 12-months post-partum.

Offspring (Gen2) data

Details of anthropometry (weight, height, waist circumference (WC), and skinfold thicknesses) at 17 and 20 years old are previously described [25]. At 17 years old, ultrasound (Siemens Antares, Mountain View, CA, USA) adipose thickness was measured by ultrasonographers using standardized criteria [26]. At 20 years old, dual-energy x-ray absorptiometry (DEXA) (Norland XR-36 densitometer, Norland Medical Systems, Fort Atkinson, WI, USA) measured whole body fat, lean, and soft tissue mass (g). Venous blood was sampled after overnight fasting; serum high-sensitivity C-reactive protein (hsCRP), leptin, and adiponectin were measured at the PathWest Laboratory (Royal Perth Hospital) as previously described [27].

Diet was assessed at 17 years old using a semi-quantitative food frequency questionnaire (FFQ) developed by the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Adelaide, Australia, to assess study adolescents' usual dietary intake over the previous year. The FFQ was able to correctly rank most nutrient intakes when compared with a 3-day food diary [28]. Factor analysis was used to reduce the food group intakes measured by the FFQ into a smaller number of underlying factors or

dietary patterns that could explain variations in dietary intake. Two major dietary patterns were identified: a 'Healthy' pattern, high in fresh fruit, vegetables, whole grains, and grilled or canned fish, and a 'Western' pattern, high in take-away foods, confectionery, soft drinks, and chips [29].

Steps in the study

Shown in Figure 1, the steps of recruitment and follow-ups of the Raine Study are depicted, alongside the subpopulations within which the multiple epigenetic analyses were performed.

Infinium HumanMethylation 450 K BeadChip array

In step 1 (Figure 1), DNA was extracted from 5-ml samples of ethylenediaminetetraacetic acid

(EDTA) blood from 17 years old adolescents from the Raine study using a Puregene DNA isolation kit based on a simple salting out technique [30]. 1 µg of the genomic DNA was treated with sodium bisulphite using Zymo EZ DNA Methylation-Gold kit (ZymoResearch, Irvine, California, USA, D5007), and processing of the HumanMethylation450K (Infinium Methylation 450 K; Illumina, Inc. CA, USA) platform was carried out by the Centre for Molecular Medicine and Therapeutics (CMMT) (<http://www.cmmt.ubc.ca>).

Infinium HumanMethylation 450 K BeadChip array data processing

QC used R and Bioconductor packages *shinyMethyl* [31], *MethylAid* [32], and *RnBeads* [33]. Four participants (3 outliers, 1 gender misclassification) were removed. Three outliers were

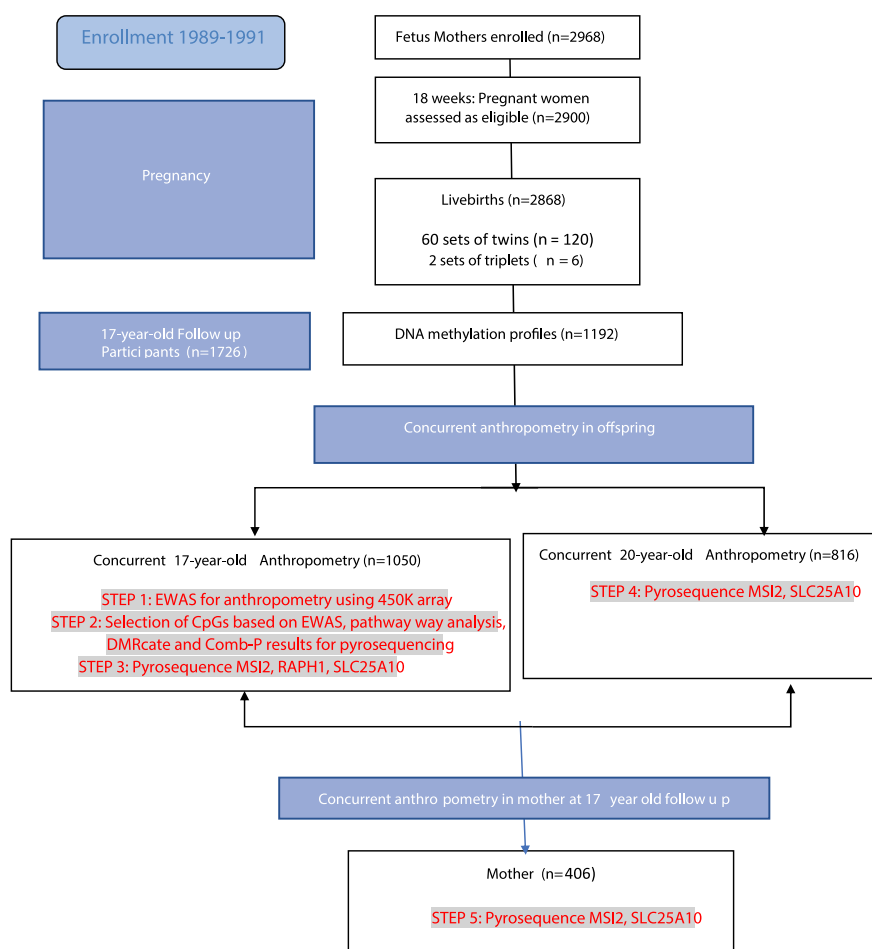


Figure 1. Study steps shown, including recruitment and follow-ups of the Raine Study. The epigenetic analyses, specifically Infinium 450 K array and pyrosequencing at different time points and generations is shown in highlighted red font.

removed due to poor probe quality and were identified as outliers using principal components for both methylAID and shinymethyl using default thresholds. These default thresholds for the 450 K array correspond to [1] median of methylated/unmethylated ratio = 10.5 [2], overall sample-dependent ratio = 11.75 [3], bisulphite conversion = 12.75 [4], hybridization control probe quality = 13.25 [5], detection p -value = 0.95. The fourth sample was removed as there was sex chromosome discrepancy based on the average signal intensity for those participants. Sixty-five intentional SNP CpGs, 11,648 sex chromosome CpGs, and 10,777 CpGs with a detection p -value >0.05 in any sample were removed. One hundred and sixty probes with bead counts <3 in >5% samples were removed. Beta-mixture quantile normalization [34] was applied to each CpG. As the batch effects were still present, technical covariates (plate, slide, and well number) were included in all statistical models performed.

Infinium HumanMethylation 450 K BeadChip array data analysis

As cellular heterogeneity is known to influence DNA methylation profiles and drive some of the observed differences across individual blood samples, we adjusted methylation values for cell counts using the estimated Housman method [35] as implemented in the R statistical package, *minfi* [36] for six cell types (CD8T, CD4T, NK, B cell, monocytes, and granulocytes). Linear mixed effects examined the association of each autosomal CpG with adiposity measures, accounting for age, sex, cell count, well, and row, with plate number as the random effect. Changes in BMI per unit change in methylation (0–1 scale, corresponding to 0–100% change in methylation) were reported.

Differentially Methylated Regions (DMRs)

For the identification of DMRs, we utilized two methods, Comb-p [37] and DMRcate [38]. Comb-p and DMRcate utilize one-step Sidak correction [39] and the Benjamini–Hochberg method for false-discovery rate (FDR) [40], finding significant CpGs <1,000 nucleotides. For the DMR analysis, the Stouffer correction was applied to calculate the

p -value. Within a genomic region, we set the minimum number of CpGs at 2. To maximize robustness, we only considered a DMR significant when it was detected in both Comb-p and DMRcate. Resulting DMRs were annotated with the UCSC Refgene panel from the Illumina annotation file in DMRcate [37].

Pathway analysis

Gene set enrichment analysis used the Bioconductor package *missMethyl* [41]. Pathways enriched by CpGs with $p < 0.001$ were reported as significant, with FDR corrected p -value <0.05. These CpGs with $p < 0.001$ were annotated with the nearest UCSC hg19 human genome reference panel to determine their Entrez Gene ID. To account for potential probe number bias, where a gene with more CpGs is likely to be associated with a gene, we employed a modified hypergeometric test as implemented in *missMethyl*. We also accounted for ‘multi-gene bias’ as approximately 10% of CpGs on the 450 K are associated with more than one gene and is accounted for within the gene set testing function within *missMethyl*.

Pyrosequencing

Bisulphite pyrosequencing was undertaken on DNA from 17 ($n = 1050$) and 20 ($n = 817$) year olds (gene2)(Step 4 in Figure 1) and their mothers ($n = 406$) (gen1) (Step 5 in Figure 1). It measured DNA methylation (%) of each CpG (Pyro Q-CpG 1.0.9 software). Bisulphite conversion, PCR, and primer design details are provided in supplementary Table 1. Pyrosequencing covered two adjacent CpG loci in *MSI2* locus and one additional CpG in the *SLC25A10* locus (Supplementary Table 2).

Statistical analysis of pyrosequencing data

Statistical analysis was carried out using SPSS (v24.0) and R (v3.3). Where appropriate, outcomes were transformed using natural logarithms to satisfy the assumption of normality. Agreement between methylation from pyrosequencing and Infinium arrays was assessed by Bland-Altman plots [42]. The same tests were used to assess agreement between pyrosequenced methylation within corresponding loci in mother and offspring.

Table 1. General characteristics of participants. Number of participants, general characteristics, phenotypic information, and DNA methylation levels of pyrosequenced *MSI2*, *SLC25A10*, and *RAPH1*.

	All	Males	Females
17 year old			
N	1050	532	518
Age (years)	17.1 (0.3)	17.1 (0.3)	17.1 (0.3)
Family Income (n = 858)			
<\$35,000	125 (15%)	62 (14%)	63 (15%)
\$35,001 to \$70,000	222 (26%)	114 (26%)	108 (26%)
>\$70,000	511 (60%)	269 (60%)	242 (59%)
Maternal Ethnicity (Caucasian)	929 (88.5%)	467 (88%)	462 (89%)
Paternal Ethnicity (Caucasian)	924 (88%)	463 (87%)	461 (89%)
Weight (kg)	68.8 (14.9)	73.0 (15.0)	64.4 (13.6)
Height (cm)	172.3 (9.2)	178.3 (7.2)	166.2 (6.5)
BMI (kg/m ²) (n = 878)	23.1 (4.5)	22.9 (4.3)	23.3 (4.7)
BMI Categories			
<25 kg/m ²	670 (78%)	347 (79%)	323 (76%)
Overweight (25–29.9 kg/m ²)	122 (14%)	63 (14%)	59 (14%)
Obese (≥30 kg/m ²)	72 (8%)	30 (7%)	42 (10%)
Waist circumference (cm) (n = 847)	79.5 (11.5)	80.9 (11.0)	78.0 (11.7)
Fasting Insulin (mIU/L)	9.3 (10.8)	8.5 (6.6)	10.2 (14.0)
Glucose (mmol/L)	4.7 (0.5)	4.8 (0.5)	4.6 (0.4)
HOMA (molar units)	2.1 (2.7)	2.2 (3.2)	2.2 (3.2)
hsCRP (mg/L)	2.0 (5.4)	1.5 (4.0)	2.5 (6.5)
Leptin (μg/L)	19.3 (22.5)	6.7 (10.0)	32.7 (24.4)
Adiponectin (mg/L)	9.5 (5.3)	7.9 (3.8)	11.4 (6.1)
Exercise (n = 643)			
< 1 time/week	130 (20%)	45 (14%)	85 (26%)
1–3 times/week	342 (53%)	170 (53%)	172 (54%)
≥ 4 times/week	170 (26%)	106 (33%)	64 (20%)
Dietary Patterns (n = 626)			
Healthy	0.00 (0.90)	−0.14 (0.86)	0.12 (0.88)
Western	−0.03 (0.83)	0.30 (0.86)	−0.31 (0.69)
Adolescent smoking (in last 4 weeks) (n = 787)	134 (17%)	62 (15%)	72 (19%)
DNA methylation			
<i>MSI2</i> CpG 1 (%)	82.5 (4.5)	83.1 (4.0)	81.8 (4.8)*
<i>MSI2</i> CpG 2 (%)	73.6 (3.4)	74.2 (3.3)	72.9 (3.4)*
<i>MSI2</i> CpG 3 (%)	60.5 (5.0)	60.9 (5.1)	72.9 (3.4)*
<i>SLC25A10</i> CpG 1 (%)	79.2 (6.0)	79.2 (6.2)	79.2 (5.8)
<i>SLC25A10</i> CPG 2 (%)	77.9 (5.4)	77.9 (5.4)	77.8 (5.3)
<i>Raph1</i> CpG 1 (%)	78.9 (4.5)	79.5 (4.9)	78.3 (4.0)
20 year old follow-up			
N	817	414	403
Weight (kg)	72.9 (17.0)	78.5 (16.2)	67.2 (15.8)
Height (m)	172.0 (9.4)	179.0 (7.2)	166 (6.4)
BMI (kg/m ²)(n = 817)	24.5 (5.1)	24.5 (4.5)	24.5 (5.7)
BMI Categories			
(<25 kg/m ²)	580 (71%)	290 (70%)	291 (72%)
Overweight (25–29.9 kg/m ²)	131 (16%)	74 (18%)	56 (14%)
Obese (≥30 kg/m ²)	106 (13%)	50 (12%)	56 (14%)
Waist (cm) (n = 823)	80.2 (12.9)	82.7 (12.1)	77.1 (13.1)
Total fat mass (kg) (n = 749)	22.380 (12.450)	17.885 [11,279]	27,061 [11,886]
Total soft tissue mass (kg)	69.385 (15.442)	74.790 (14.907)	63,758 [13,920]
Total lean mass (kg)	47.003 (12.090)	56.904 (7.883)	36,696 [5,017]
DNA methylation			
<i>MSI2</i> CpG 1 (%)	82.1 (4.2)	83.0 (3.9)	81.2 (4.3)*
<i>MSI2</i> CpG 2 (%)	73.8 (3.0)	74.6 (2.6)	72.9 (3.2)*
<i>MSI2</i> CpG 3 (%)	60.1 (4.4)	60.8 (4.1)	59.2 (4.5)*
<i>SLC25A10</i> CpG 1 (%)	78.5 (5.0)	78.4 (4.8)	78.5 (5.1)
<i>SLC25A10</i> CpG 2 (%)	77.0 (5.2)	76.9 (5.0)	77.1 (5.4)
Mother and Newborn/Infant Early life			
N	1050	532 (51%)	518 (49%)
Birthweight (kg)	3307 (585)	3368 (569)	3244 (595)
Birth Length (cm)	50.0 (2.7)	49.4 (2.6)	48.5 (2.7)
Gestational Age (weeks)	39.2 (2.1)	39.4 (1.9)	39.1 (2.3)
Prepregnancy BMI (n = 1031)	22.4 (4.4)	22.3 (4.4)	22.6 (4.5)
Gestational weight gain rate (g/wk) (n = 992)	502 (195)	503 (193)	501 (197)
Maternal stress (n = 1031)	1.18 (1.23)	1.15 (1.25)	1.21 (1.22)
Smoking during pregnancy	137 (13%)	62 (11.7%)	75 (14.5%)
Breast feeding duration (months) (n = 998)	8.1 (7.3)	8.2 (7.1)	8.1 (7.4)

(Continued)

Table 1. (Continued).

	All	Males	Females
17 year old			
Mother at time of DNA collection			
N			406
Age (yrs)	-	-	45.3 (5.5)
BMI(kg/m ²)	-	-	27.38 (6.2)
DNA methylation			
MSI2 CpG 1 (%)	-	-	86.3 (3.9)
MSI2 CpG 2 (%)	-	-	77.1 (2.5)
MSI2 CpG 3 (%)	-	-	64.4 (3.9)
SLC25A10 CpG 1 (%)	-	-	77.8 (5.0)
SLC25A10 CPG 2 (%)	-	-	76.0 (5.7)

* indicates that males and females are statistically different with *P*-value <0.05 for that variable.

Multivariate linear regression was undertaken with the outcome being adiposity or inflammatory marker (hsCRP, leptin, adiponectin) and independent variable being the pyrosequenced CpG(%). Adiposity was measured as BMI (kg/m²), WC (cm), subcutaneous and visceral fat thickness, fat, and lean mass (g). Coefficients (β) represent the log-transformed unit of change in outcome/SD change in methylation with associated FDR adjusted p-value (q-values). Analyses adjusted for age, sex, cell count where available [43], and SNPs identified as affecting CpGs. SNPs associated with CpGs were evaluated by the Gene Environment Methylation (GEM) package [44], which detects genome-wide, methylation markers associated in *cis* and *trans* genetic polymorphisms according to the model $M = \alpha + \beta G + \text{age} + \text{sex} + \text{cell count} + \epsilon$ (where M is the methylation matrix and G is the genotype matrix). Collinearity was avoided by removing covariates with high variance inflation factors [45]. All models were tested for sex interaction and reported separately if a significant interaction was detected.

In assessing effects of early environment, linear regression was undertaken with CpG methylation as the outcome. Environmental factors from early life (prepregnancy BMI, GWG, smoking during pregnancy, maternal stress, and breast feeding duration) were independent variables. A final model predicting BMI at 17 years old considered the covariates of all pyrosequenced CpGs, cell counts, age, sex, and environmental factors (associated with CpGs at $p < 0.05$).

Backward selection determined maximal parsimony with variables removed if the significance level of its F value is greater than the removal value of 0.10. The model summary R^2 was taken to indicate the variance in BMI at 17 years old that is predicted by the variables in the model.

Results

Participants

Of 1192 individuals with DNA methylation profiles, 1050 (51% males) with anthropometry measurements were analysed. Participants had a mean (SD) age of 17.1 (0.3) years, BMI 23.1 (4.5) kg/m², and WC 79.5 (11.5) cm. At age 20 years, 817 of these had anthropometry, DEXA (mean BMI 24.5 (5.1) kg/m², WC 80.2 (12.9) cm), and DNA methylation measured (Table 1). Additionally, 406 mothers had concurrent BMI and DNA methylation measured.

Identification of dmCpGs and DMRs associated with adiposity measures at 17 years

DNA from peripheral blood of the 17 year old adolescents from The Raine study was interrogated for genome-wide DNA methylation using the Infinium 450 K array. Adjusting for cell count heterogeneity and batch effects, there were associations (Bonferroni corrected $p < 1.06\text{e-}07$) between DNA methylation and concurrent BMI, WC, and subcutaneous fat (SC) (supplementary Table 3),

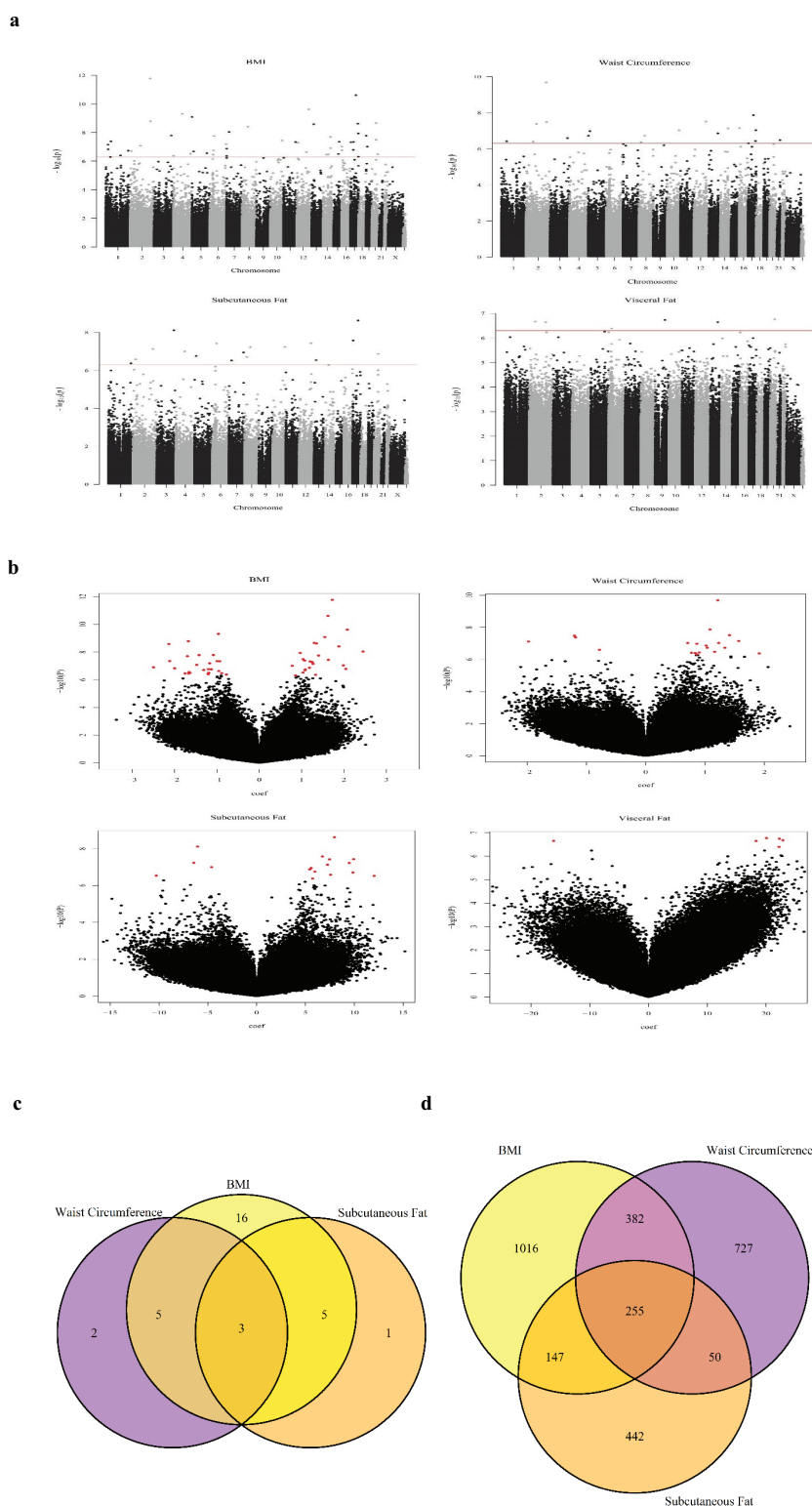


Figure 2. Epigenome Wide Association Study (EWAS) results which show DNA methylation associations with adiposity at 17 years of age.

Figure 2 Manhattan (a) and volcano plots (b) showing the $-\log_{10}$ p-values for individual autosomal CpG probe associations for BMI, WC, subcutaneous and visceral fat thickness. The value of the dotted line in Panel A is the Bonferroni cut-off of P-value = 1.06×10^{-7} . (C/D) Venn Diagrams showing overlap between CpGs associated with BMI, WC and subcutaneous fat; (C) includes CpGs that are significant after Bonferroni. (D) includes CpGs that are significant by the threshold $p < 0.001$. 8 CpGs were shared between BMI and subcutaneous fat thickness, 8 between WC and subcutaneous fat thickness, and 3 (cg21139312 [*MSI2*], cg07390598 [*RAPH1*] and cg10773745 [intergenic region]) between all three adiposity measurements

but no associations between CpG methylation and visceral fat thickness. Manhattan and volcano plots (Figure 2(a,b)) show the $-\log_{10}$ p-values for individual autosomal CpG probe associations for BMI, WC, SC, and visceral fat thickness. Q-Q plots for evaluation of model fit are shown in Supplementary Figure 1. The lambda values for BMI, WC, subcutaneous fat, and visceral fat were 1.5, 1.7, 1.1 and 3.2, respectively. Twenty-nine dmCpGs were associated with BMI, the top 2 dmCpGs were located within pleckstrin homology domains containing protein 1 (*RAPH1*) (cg07390598) and Musashi RNA-binding protein 2 Ras-associated (*MSI2*) (cg21139312) genes. Ten dmCpGs were associated with WC, the top 2 being cg07390598 and cg21139312, the same dmCpG sites within *RAPH1* and *MSI2* that were most strongly associated with BMI. Nine dmCpGs associated with SC fat thickness, the top dmCpG being cg21139312 located within *MSI2*. Overlap was observed between dmCpGs associated with BMI, WC, and SC; 8 dmCpGs were shared between BMI and SC fat thickness, 8 between WC and SC fat thickness, and 3 (cg21139312 [*MSI2*], cg07390598 [*RAPH1*], and cg10773745 [intergenic region]) between all three adiposity measurements (Figure 2c).

Regional methylation associations were explored using *DMRcate* and *Comb-p*, identifying 9 and 22 (6 overlapping) DMRs, respectively, associated with BMI; 8 and 2 (2 overlapping) DMRs associated with WC, and 8 and 6 (2 overlapping) with SC (supplementary Table 4). The DMR in a 60 bp region within *MSI2* was associated with BMI by *DMRcate* (p-value = $7.01\text{E-}12$) and *Comb-p* (p-value = $1.23\text{E-}09$), and with WC by *DMRcate* (p-value = $4.42\text{E-}06$). Both DMR models identified the *MSI2* region in association with SC fat (DMRcate: p-value = $4.42\text{E-}06$, comb-p: p-value = $7.0\text{E-}12$).

Pathways associated with inflammatory response and histone modification were enriched among dmCpGs associated with adiposity

To explore the functional significance of associations between methylation and adiposity measures, pathway analysis was undertaken. For BMI, the top pathway in the Gene Ontology category was

inflammatory response ($p_{\text{FDR}} = 5.59\text{E-}04$), for WC, there was enrichment for histone modifications and methyltransferase activity ($p_{\text{FDR}} = 0.033$). SC fat showed no significant pathways (supplementary table 5). For BMI and WC or SC, no KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathways were significant.

Validation by pyrosequencing of adiposity-related dmCpGs

The criteria for choosing CpGs to proceed to pyrosequencing included prioritizing those that consistently passed FDR on three measures of adiposity (BMI, WC, Subcutaneous fat), and then two measures of adiposity (BMI and WC). Within the FDR significant CpGs, those whose associations were less likely to be confounded by technical factors and more likely to be interpretable and clinically meaningful were favoured. For instance, we chose those that showed significance within pathway analysis, *DMRcate* and *Comb-P*. Those within CpG islands were also given priority as they were more likely to influence transcription. We gave lower priority to those CpGs with extreme mean methylation (<10% and >90%), as technical accuracy of the pyrosequencing method is decreased at the limits of the range.

MSI2 cg21139312 and *RAPH1* cg07390598 were selected for pyrosequencing as the most highly ranked dmCpGs for association with BMI and WC; *MSI2* cg21139312 was also the top ranked dmCpG with SC fat. The region of chromosome 17:55,663,225 to 55,663,284 encompassing the *MSI2* cg21139312 (chr17:55,663,225) was the second ranked enriched region for BMI by *DMRcate* ($p = 7.01\text{E-}12$) and *Comb-p* (p-value = $9.09\text{E-}09$), and 7th ranked for WC (p-value = $4.4\text{E-}06$). The third dmCpG chosen for validation was *SLC25A10* cg09996840, as it was associated with both BMI and WC and belonged to an enriched 990 base region for WC (p-value = $5.58\text{E-}07$) (supplementary Table 4). For *MSI2*, the pyrosequencing assay covered two additional CpG sites, one upstream (CpG1) and one downstream (CpG3) of cg21139312 (CpG2). For *SLC25A10*, the pyrosequencing assay covered one additional CpG site (CpG2). Bland-Altman plots comparing DNA methylation levels

measured by 450 K array and pyrosequencing at these CpGs are shown (supplementary Figure 2a).

Consistent with the methylation array, pyrosequencing analysis showed *MSI2* cg21139312 (CpG2) associated with BMI ($\beta = 0.026$, $p\text{-value} = 1.3\text{e-}04$), WC ($\beta = 0.019$, $p\text{-value} = 2.1\text{e-}04$), and SC fat thickness ($\beta = 0.068$, $p\text{-value} = 0.003$). The two flanking *MSI2* CpG sites were also associated with BMI, WC, and SC fat thickness. All three *MSI2* CpGs were associated positively with suprailiac, subscapular, and triceps skinfold thicknesses (Table 2). Methylation of *SLC25A10* CpGs measured by pyrosequencing were associated with BMI (CpG1 $\beta = 0.020$, $p\text{-value} = 0.011$; CpG2 $\beta = 0.030$, $p\text{-value} = 1.4\text{e-}04$), WC (CpG2 $\beta = 0.018$, $p\text{-value} = 0.004$), and SC fat thickness (CpG1 $\beta = 0.080$, $p\text{-value} = 0.004$; CpG2 $\beta = 0.087$, $p\text{-value} = 0.003$). *RAPH1* methylation was associated with BMI, WC, or SC fat thickness before cell-type adjustment, but there was no association with these measures of adiposity after adjustment (Table 2, Figure 3). *SLC25A10* and *RAPH1* methylations were not associated with skinfold thicknesses. No CPSs are associated with visceral fat (Table 2).

DNA methylation at 17 years of all three *MSI2* and *Raph1* CpGs was positively associated with BMI measured at 20 years. Changes in DNA methylation between 17 and 20 years were not associated with BMI at 20 years of age. (Supplemental Table 6)

Association of DNA methylation with serum inflammatory markers and adipokines at 17 years of age

At 17 years, *MSI2* methylation was associated with hsCRP (CpG1 $\beta = 0.053$, $p\text{-value} = 0.02$; CpG3 $\beta = 0.08$, $p\text{-value} = 0.001$) and serum leptin (CpG1 $\beta = 0.034$, $p\text{-value} = 0.039$; CpG2 $\beta = 0.056$, $p\text{-value} = 0.0002$; CpG3 $\beta = 0.069$, $p\text{-value} = 0.0018$) (supplementary table 7, Figure 4). Methylation of *SLC25A10* CpG1 ($\beta = 0.036$, $p\text{-value} = 0.018$; CpG2 $\beta = 0.06$, $p\text{-value} = 0.00017$) and *Raph1* CpG1 ($\beta = 0.138$, $p\text{-value} = 0.00016$) were associated with leptin levels (Figure 4). No associations were detected between adiponectin and any of the

pyrosequenced CpGs. All sex interactions were non-significant.

Influence of genotype on dmCpGs associated with adiposity

To assess how genotype influences pyrosequenced DNA methylation loci, the GEM R package [44] was used. *SLC25A10* (cg09996840) was associated with genotype at 12 *cis* acting SNPs (supplementary table 8a) all located on the north shore of a CpG island at chr17:79,682,525–79,682,767. The most significant SNP was rs6565624 ($q\text{-value} = 2.4 \times 10^{-152}$ (FDR)). Neither *MSI2* (cg21139312) nor *RAPH1* (cg07390598) were significantly associated with genotype at any of the SNP sites measured. The identified SNPs were tested alongside the pyrosequenced *SLC25A10* CpGs for predicting BMI or WC. Genotype at tSNPs (rs6565624_G, rs3830068_G, rs12452184_A) was consistently associated with BMI and WC, while abolishing the effects of *SLC25A10* CpG1 and CpG2. (Supplementary table 8b and model 3 in Figure 2).

Early environmental factors are associated with altered *MSI2* and *SLC25A10* methylation in adolescence

Methylation of *SLC25A10* CpGs 1 and 2 was positively associated with maternal pre-pregnancy BMI (CpG1 $\beta = 0.07$, $p\text{-value} = 0.022$; CpG2 $\beta = 0.12$, $p\text{-value} = 0.002$) (Supplementary table 9), while *MSI2* CpG 2 and 3 methylation was positively associated with smoking at 18 weeks (CpG2 $\beta = 0.81$, $p\text{-value} = 4.8\text{e-}04$; CpG3 $\beta = 0.76$, $p\text{-value} = 0.006$) but not with smoking at 28 week gestation. *RAPH1* methylation was associated with GWG rate ($\beta = -1.49$, $p\text{-value} = 0.036$) (g/wk).

A model was formed by backward selection with pre-pregnancy BMI, maternal smoking at 18-week gestation and GWG, all pyrosequenced CpGs, cell counts and potential me-QTLs (rs6565624_G, rs3830068_G, rs12452184_A) for the outcome of BMI at 17 years. Variables were removed if the significance level of its F value was greater than 0.10. The final model ($r^2 = 0.243$ (SE = 0.15), (Table 3) included independent effects of rs6565624_G, *MSI2* CpG1, *RAPH1*, maternal

Table 2. Associations of *MSI2*, *RAPH1* and *SLC25A10* DNA methylation with anthropometry Associations of *MSI2*, *RAPH1* and *SLC25A10* DNA methylation with anthropometrical measures of adiposity (BMI, waist, skin folds) at ages 17 and 20 years and associations with measures of body composition (subcutaneous and visceral fat thickness at age 17 years, and DEXA measurements at age 20 years). All models were adjusted for age, sex and where available for cell count (CD4, CD8, NK, B Cell, Monocytes). Covariates with high variance inflation factors were removed to avoid collinearity as well as Granulocytes. All β -coefficients are expressed per SD change of CpG. q-values which pass FDR accounting for multiple testing are shown in darker font. # Log transformed.

			17 Year old			20 Year old		
Outcome	Gene	CpG	β	P-Value	FDR (q-value)	β	P-Value	FDR (q-value)
Anthropometry								
BMI#	MSI2	CpG1	0.027	2.5×10^{-4}	0.003	0.026	4.0×10^{-4}	0.003
		CpG2	0.026	1.3×10^{-4}	0.002	0.027	3×10^{-4}	0.003
		CpG3	0.032	0.001	0.004	0.027	3×10^{-4}	0.003
	SLC25A10	CpG1	0.020	0.011	0.028	0.015	0.041	0.068
		CpG2	0.030	1.4×10^{-4}	0.002	0.016	0.028	0.051
Waist#	RAPH1		0.013	0.102	0.157			
	MSI2	CpG1	0.020	4.3×10^{-4}	0.003	0.018	0.002	0.006
		CpG2	0.019	2.1×10^{-4}	0.003	0.024	0.000	0.000
		CpG3	0.028	1.6×10^{-4}	0.002	0.014	0.017	0.037
	SLC25A10	CpG1	0.009	0.126	0.177	0.014	0.017	0.037
CpG2		0.018	0.004	0.012	0.012	0.035	0.061	
RAPH1		−0.001	0.893	0.922				
Skinfolds								
Suprailiac#	MSI2	CpG1	0.070	0.004	0.012	0.050	0.018	0.038
		CpG2	0.075	4.7×10^{-4}	0.003	0.067	0.001	0.006
		CpG3	0.080	0.007	0.018	0.058	0.007	0.018
	SLC25A10	CpG1	0.020	0.406	0.488	0.058	0.007	0.018
		CpG2	0.024	0.444	0.514	0.044	0.037	0.063
RAPH1		−0.023	0.345	0.426				
Subscapular#	MSI2	CpG1	0.060	0.001	0.004	0.052	0.001	0.006
		CpG2	0.056	0.001	0.004	0.052	0.001	0.006
		CpG3	0.071	0.002	0.007	0.039	0.018	0.038
	SLC25A10	CpG1	0.023	0.219	0.285	0.029	0.083	0.133
		CpG2	0.029	0.158	0.218	0.027	0.102	0.157
RAPH1		−0.012	0.522	0.583				
Triceps#	MSI2	CpG1	0.055	0.003	0.001	0.023	0.189	0.253
		CpG2	0.056	0.001	0.004	0.037	0.033	0.058
		CpG3	0.050	0.026	0.049	0.014	0.403	0.488
	SLC25A10	CpG1	0.028	0.127	0.177	0.040	0.021	0.043
		CpG2	0.056	0.017	0.038	0.039	0.022	0.044
RAPH1		−0.002	0.923	0.937				
Measures of body composition								
Subcutaneous fat#	MSI2	CpG1	0.082	0.001	0.004			
		CpG2	0.068	0.003	0.001			
		CpG3	0.028	1.6×10^{-4}	0.002			
	SLC25A10	CpG1	0.080	0.004	0.012			
		CpG2	0.087	0.003	0.001			
RAPH1		0.015	0.577	0.630				
Visceral fat#	MSI2	CpG1	0.013	0.443	0.514			
		CpG2	−0.016	0.322	0.402			
		CpG3	−0.039	0.451	0.516			
	SLC25A10	CpG1	0.011	0.548	0.605			
		CpG2	0.000	0.985	0.985			
RAPH1		−0.039	0.122	0.176				
Lean Mass#	MSI2	CpG1				0.013	0.016	0.037
		CpG2				0.013	0.014	0.034
		CpG3				0.019	4.0×10^{-4}	0.003
	SLC25A10	CpG1				−0.009	0.110	0.160
		CpG2				−0.002	0.690	0.728
RAPH1					0.014	0.026	0.046	
Fat mass#	MSI2	CpG1				0.026	0.176	0.238
		CpG2				0.043	0.026	0.049
		CpG3				0.022	0.267	0.343
	SLC25A10	CpG1				0.043	0.029	0.053
		CpG2				0.031	0.108	0.160
RAPH1					0.013	0.533	0.601	

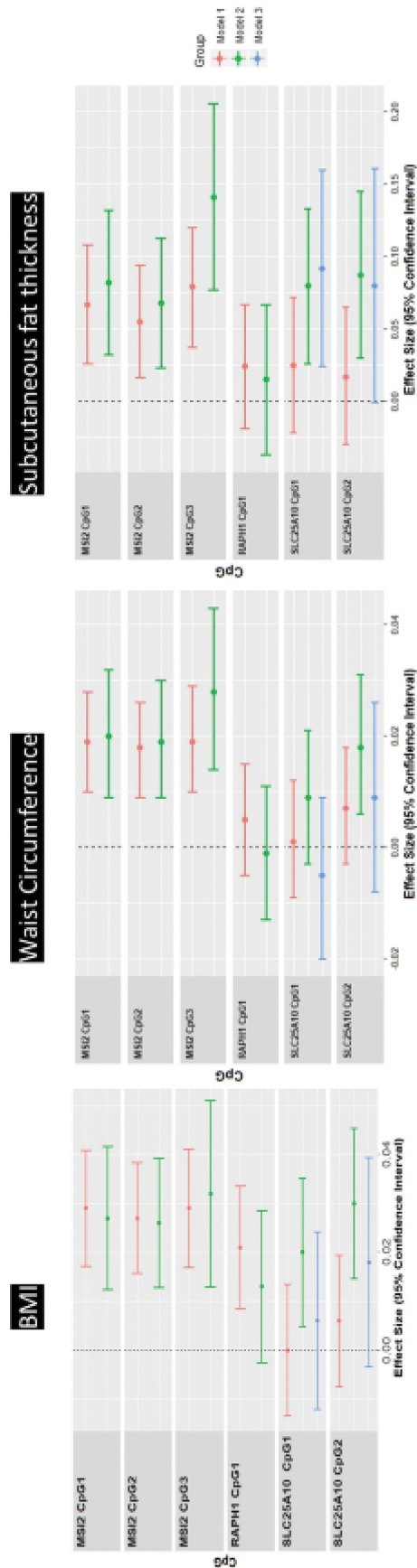


Figure 3. Linear regression associations with adiposity at age 17 years of age.

Linear regression models between DNA methylation levels at pyrosequenced CpGs (*MS12*, *SLC25A10*, *RAPH1*) on the outcomes of BMI, WC and subcutaneous fat thickness. Beta coefficient and 95% CI are shown, adjusted for age and sex (model 1), adjusted additionally for cell count (model 2) and for me-QTL SNP, where identified (model 3). All β -coefficients are expressed per SD change of CpG. For *SLC25A10*, the relevant me-QTL in model 3 is rs6565624.

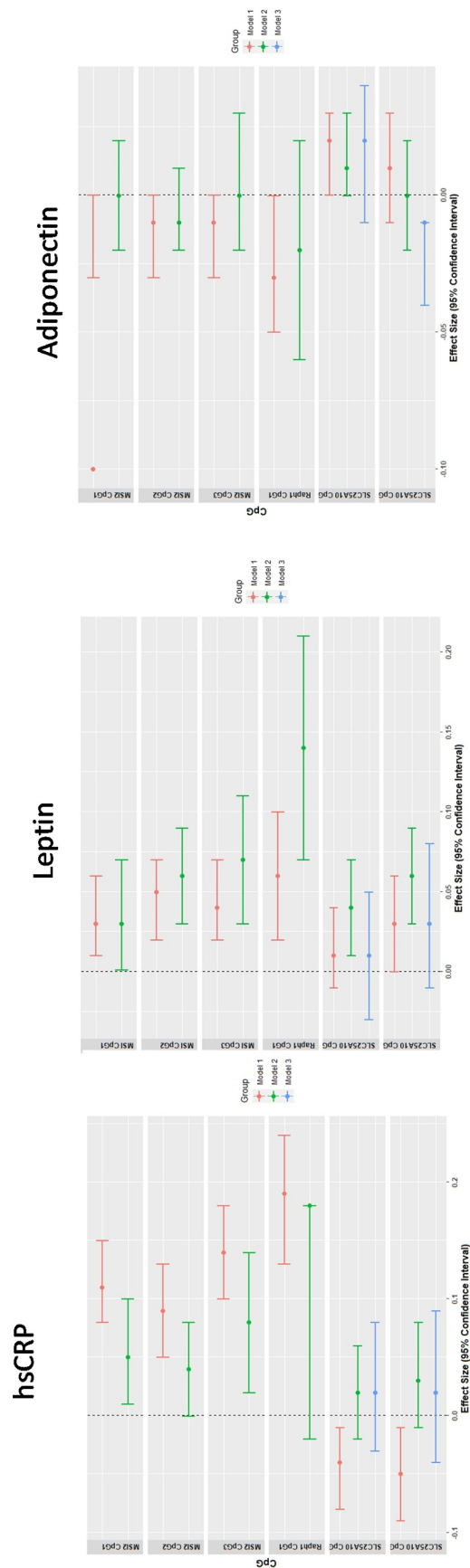


Figure 4. Association between *MSI2*, *RAPH1* and *SLC25A10* DNA methylation with inflammation and adipokines at age 17-years. Linear regression models between DNA methylation levels at pyrosequenced CpGs (*MSI2*, *SLC25A10*, *RAPH1*) on the outcomes of hsCRP, leptin and adiponectin. Beta coefficient and 95% CI are shown, adjusted for age, sex, cell count and for me-QTL SNP, where identified by GEM package [44]. All β -coefficients are expressed per SD change of CpG. ^For *SLC25A10*, the relevant me-QTL is rs6565624 using the additive model.

smoking at 18-weeks, pre-pregnancy BMI, CD8+, CD4 + T cells, and granulocytes (Table 3).

Methylation of *MSI2* and *SLC25A10* are associated with concurrent measures of adiposity at 20 years

In the 20 years old participants, DNA was extracted from peripheral blood and the three *MSI2* and two *SLC25A10* CpGs were pyrosequenced. All *MSI2* CpGs were positively associated with concurrent BMI, WC, suprailiac, and subscapular skinfold fat thicknesses and lean mass; *MSI2* CpG2 was associated with fat mass. *SLC25A10* CpG1 was associated with WC, suprailiac, and triceps skinfold thicknesses, and CpG2 with tricep thickness (all p-values < 0.05) (Table 2).

Methylation of *MSI2* in mothers is associated with maternal BMI and offspring methylation

DNA methylation of *MSI2* and *SLC25A10* CpGs from maternal blood samples was measured by pyrosequencing. *MSI2* methylation was associated with maternal BMI adjusted for maternal age (CpG1 $\beta = 0.031$, p-value = 3.6×10^{-4} ; CpG2 $\beta = 0.044$, p-value = 1.2×10^{-4} ; CpG3 $\beta = 0.043$, p-value = 1.8×10^{-5}); *SLC25A10* CpGs were not significantly associated with maternal BMI (Table 4).

Methylation levels of the CpGs from child and mother were compared by Bland–Altman plots and showed fixed and proportional bias for all CpGs (supplementary Figure 2b). The mean value of the difference differed from 0 on one-sample t-tests (all $p < 0.001$). Adolescents had overall lower levels of DNA methylation for

MSI2 CpGs, compared to their mothers (mean differences: CpG1 -4.2 ± 5.1 ; CpG2 -3.8 ± 3.7 , CpG3 -4.6 ± 5.4). The association of the mean *MSI2* methylation upon the dependent variable for difference in methylation was significant (CpG1 $\beta = 0.28$, p-value = 7.3×10^{-7} ; CpG2 $\beta = 0.38$, p-value = 2.9×10^{-12} ; CpG3 $\beta = 0.44$, p-value = 2.5×10^{-16}). All coefficients were positive, indicating that with increasing overall *MSI2* methylation, the child's *MSI2* methylation was increasingly inflated compared to the mother's.

Discussion

Our EWAS identified novel dmCpGs in peripheral blood associated with measures of adiposity in adolescents, with a considerable overlap in the dmCpGs associated with BMI, WC, and SC fat thickness. DNA methylation at three novel dmCpG loci related to the genes *MSI2*, *SLC25A10*, and *RAPH1* was associated with a wide range of adiposity/body composition measures in adolescents, and with a number of antenatal factors including GWG, maternal BMI, and maternal smoking. Furthermore, for one locus, *MSI2*, the association with BMI was observed over time and between generations, suggesting DNA methylation changes in peripheral blood may be robust markers of adiposity across the lifecourse.

In this study, we identified a consistent positive association between *MSI2* methylation and multiple measures of adiposity. *MSI2* is an RNA-binding protein which plays a key role in post-transcriptional processing [46]. The same CpG site within the *MSI2* gene has previously been reported

Table 3. Final linear regression model predicting BMI at 17 years old. Linear regression model for the outcome of BMI at 17-years was formed by backward selection with pre-pregnancy BMI, maternal smoking at 18-weeks gestation and GWG, all pyrosequenced CpGs, cell counts and potential me-QTLs (rs6565624_G, rs3830068_G, rs12452184_A). The model ($r^2 = 0.243$ (SE = 0.15), includes independent effects of rs6565624_G, *MSI2* CpG1, *RAPH1*, maternal smoking at 18-weeks, pre-pregnancy BMI, CD8+, CD4 + T cells, and granulocytes.

Parameter	Beta	SE	Sig.
rs6565624_G	0.043	0.012	3.8×10^{-4}
RAPH1 (Z score)	0.088	0.020	8.4×10^{-6}
<i>MSI2</i> CpG1 (Z score)	0.021	0.009	0.022
Maternal smoking at 18 weeks pregnancy	0.052	0.018	0.004
Prepregnancy BMI (kg/m ²)	0.015	0.002	1.1×10^{-17}
CD8 + T Cells	0.417	0.220	0.058
CD4 + T Cells	1.069	0.226	2.8×10^{-6}
Granulocytes	0.325	0.143	0.024

Model summary $R^2 = 0.24$

Table 4. Associations of peripheral blood DNA methylation of *MIS2*, *RAPH1* and *SLC25A10* with concurrent BMI in mothers of RAINE adolescents. Associations are adjusted for maternal age. BMI and maternal age were ascertained at the time of blood collection. Total sample size is 406. All β -coefficients are expressed per SD change of CpG. q-values which pass FDR accounting for multiple testing are shown in darker font.

Outcome	Gene	CpG	β	P value	FDR q-value
BMI#	<i>MSI2</i>	CpG1	0.031	3.60E-04	0.0017
		CpG2	0.044	1.20E-04	0.0008
		CpG3	0.043	1.80E-05	0.0003
	<i>SLC25A10</i>	CpG1	0.018	0.1	0.1370
		CpG2	0.018	0.11	0.1429

by Dhana *et al.* to be the most significantly associated dmCpG with BMI in older adults [47], while Shah [48] showed a positive association between cg21139312 and BMI in Dutch adults ($n = 750$) at 45.5 ± 13.3 years. Our data have now demonstrated the same association in adolescence and young adults, suggesting that differential methylation of *MSI2* may be a robust marker of adiposity throughout the lifecourse. *MSI2* methylation was also positively associated with leptin, a hormone predominantly made by adipose cells which regulates energy balance [49] and hsCRP, a marker of systemic inflammation [50]. Whether leptin or hsCRP are drivers of the associations between *MSI2* methylation and adiposity, or simply bystanders, reflecting the pro-inflammatory state and/or leptin-resistant state associated with high BMI remains to be determined.

Associations between *MSI2* methylation and BMI in the mothers of the proband were also observed. Moreover, the associations were in the same direction with a similar or larger estimated effect size compared to their children. This may be attributed to the mothers being approximately 17 years older with a higher BMI than their offspring at the time of the assessment. Compared to our adolescent data, in middle-aged participants, a larger *MSI2* methylation effect size upon BMI was reported in the Atherosclerosis Risk in Communities (ARIC), and a similar effect size was reported in the Rotterdam Study [47], the direction of the effect being the same in all these studies. These observations suggest that the association between *MSI2* methylation and BMI is generally accentuated with age, possibly due to the natural progression of increasing BMI with age. The consistent association of *MSI2* with

adiposity in two generations (mother and offspring) shows further internal replication. Bland–Altman analysis showed that with increasing mean (mother and child) *MSI2* methylation, the child's DNA methylation was relatively higher, compared with a dyad with lower mean methylation. This suggests that higher familial levels of obesity and *MSI2* methylation may compound obesity and *MSI2* methylation in the next generation, potentially transmitting across generations. This increase in BMI and *MSI2* methylation within more obese families may be due to shared environment and/or genetics.

A positive association between *SLC25A10* methylation and adiposity was also observed. *SLC25A10* catalyses transport dicarboxylates across the inner mitochondrial membrane providing substrates for the Krebs cycle and fatty acid synthesis [51]. Increased expression of *SLC25A10* has been reported in white adipose tissue in obese mouse models and decreased expression in a fasted mouse model [52]. Moreover, a global transcriptome profiling study identified *SLC25A10* as a modifier of insulin sensitivity in adipose tissue in obese women and an inhibitor of insulin-stimulated lipogenesis [53]. Me-QTLs were also identified with cg09996840, which when adjusted for, abolished the association between *SLC25A10* CpGs and BMI, suggesting that genetics, rather than environment determines the relationship between methylation and adiposity at this locus. DNA methylation at *SLC25A10* may lay on the regulatory pathway by which the identified SNPs influence BMI, in much the same way Kato *et al.* [54] suggested for SNPs associated with blood pressure.

The third dmCpG associated with BMI was located within *RAPH1*, which encodes a protein

regulating actin and cellular migration [55]. No direct association between *RAPH1* and energy homeostasis has previously been reported. However, the dmCpG site located within *RAPH1* overlaps the cholinesterase 2 locus [56] that codes a protein that binds to butyrylcholinesterase, which in turn has been linked to adiposity and insulin resistance [57]. In contrast with the multiple associations observed between DNA methylation and BMI, WC, and SC, there were no associations observed between DNA methylation and visceral fat, possibly, due to the youth of our participants, in which excess adipose tissue is preferentially laid down subcutaneously, increasing relative sensitivity for detecting associations with subcutaneous fat compared to deeper fat deposits [58].

Methylation of *MSI2*, *RAPH1*, and *SLC25A10* dmCpGs was also associated with antenatal factors including GWG, maternal BMI, and maternal smoking, consistent with the paradigm that the early environment can influence the epigenome with long-term consequences. Interestingly, there were differences in the associations between the dmCpGs, with *RAPH1* methylation being associated with GWG, *MSI2* with maternal smoking, and *SLC25A10* with maternal pre-pregnancy BMI, suggesting that different environmental exposures may drive associations at distinct loci. Interestingly, even after adjusting for these perinatal factors, we found that methylation of the dmCpGs within *MSI2*, *RAPH1*, and *SLC25A10* independently predicted adolescent BMI, accounting for 24% of variance. This suggests that differential methylation of the three dmCpG loci identified in this study captures a significant proportion of the variance associated with BMI and that such marks may be useful indices of adiposity.

The strength of this study is the ability to integrate prospectively collected early life environment with multiple time points across multiple generations, from which we have identified robust associations between specific dmCpG sites and measures of adiposity across a wide age range. Moreover, our study shows a strong link between inflammation and dmCpGs related to adiposity in adolescence. The most significant gene ontology on pathway analysis was ‘inflammatory responses,’

and selected pyrosequenced dmCpGs were highly associated with hsCRP and leptin, consistent with obesity being a chronic inflammatory disease [59].

Limitations of this study include the use of BMI which is an inexact approximation of adiposity, and which is unable to differentiate between muscle and fat tissue. However, we found considerable overlap between dmCpGs and DMRs associated with BMI, WC, SC, and skinfold thickness which increases the validity of the associations. A second limitation is that we measured DNA methylation in whole blood, and whether similar epigenetic changes are present in adipose tissue is not currently known; however, DNA methylation changes in nucleated blood cells may reflect changes in immune function or be related to inflammation. Consistent with this hypothesis, there was enrichment among the dmCpGs in pathways related to inflammation. Thirdly, we report associations within a single cohort. However, pyrosequencing validation was undertaken to increase the accuracy of the DNA methylation measurements with internal validation at two separate follow-up time points, and within the mothers of the children. Finally, although we found that the adiposity-associated dmCpGs were also associated with a variety of maternal factors, it is not known whether these maternal factors drive increased adiposity through changes in the DNA methylation at these dmCpG loci, although as the dmCpGs within *MSI2*, *SLC25A10*, and *RAPH1* all predicted adiposity independently of maternal factors, it would suggest that the likely effect is increased adiposity resulting in altered methylation. Methylation levels within *SLC25A10* are associated with *in cis* genotype and may mediate a genetic influence on adiposity levels.

In conclusion, *MSI2*, *SLC25A10*, and *RAPH1* methylation was identified through EWAS and validated in their associations with multiple measures of adiposity. All three independently associated with adolescent BMI even after adjusting for multiple early life factors.

MSI2, in particular, has been shown in this study to be particularly robust over time and multiple anthropometry measures. Furthermore, the association replicated in the mother and with serum leptin as a marker of adiposity. Such findings suggest that certain DNA methylation changes in peripheral

blood can be robust markers of adiposity. It is not possible to determine whether the DNA methylation changes are a consequence or cause of adiposity and related inflammation, with the study suggesting both direct and indirect mediation through early life factors.

Acknowledgments

We acknowledge the Raine Study participants and their families, The Raine Study Team for cohort co-ordination and data collection, The NH&MRC for their long-term contribution to funding the study over the last 29 years, The Telethon Kids Institute for long-term support of the Study. We also acknowledge The University of Western Australia (UWA), Curtin University, The Telethon Kids Institute, Women and Infants Research Foundation, Edith Cowan University, Murdoch University, The University of Notre Dame Australia, and Raine Medical Research Foundation for providing funding for Core Management of the Raine Study. The DNA methylation work was supported by NHMRC grant 1059711. The 17-year follow-up was supported by the National Health and Medical Research Council Program grant (ID353514) and Project grant (ID403981).

Availability of data and material

The Raine Study holds a rich and detailed collection of data gathered over 30 years for the purpose of health and well-being research. The informed consent provided by each participant does not permit individual-level data to be made available in the public domain (i.e., a public data repository). However, de-identified analytic data sets are available to all researchers for original research or auditing of published findings. All data access is managed through established Raine Study procedures which require data handlers to agree to a code of conduct, outlined in the Raine Study Researcher Engagement Policy, that includes safeguards to protect the identity of participants. Details of the data access processes and code of conduct are available on the Raine Study website.

Authors' contributions

R-CH, KAL, and GCB conceived the project. R-CH, KAL, GCB, KMG, LJB, JDH, JMC, TAM, PEM, WHO developed the methods. MAB, RC-H, EC, and KAL performed the laboratory pyrosequencing. MSK, JLM, AMM performed the laboratory arrays. PEM, SR and DA performed bioinformatics. R-CH, KAL and KMG wrote the manuscript. All co-authors reviewed and edited the manuscript.

Declarations

Ethics approval and consent to participate/Consent for publication

Institutional ethics approval for the original Raine Study and DNA methylation analysis have been obtained from the University of Western Australia Ethics Committee (RA/4/1/6613, 1214-EP, RA-4-1-2646). Written informed consent was obtained from legal guardians or parents of those under the age of 17 years with assent of the child. Written informed consent was obtained for those 17 years and older.

Funding

RCH and TAM are supported by National Health and Medical Research Council Fellowships (grant number 1053384 and 1136046, respectively). S.R is supported by the National Health and Medical Research Council EU grant (1142858) and the Department of Health, Western Australia FutureHealth fund in connection with the European Union's Horizon2020 grant 733206. This work was supported by National Health and Medical Research Council Grant Number 1059711 resources provided by The Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. KMG is supported by the UK Medical Research Council (MC_UU_12011/4), the National Institute for Health Research (NIHR Senior Investigator (NF-SI-0515-10042) and the NIHR Southampton Biomedical Research Centre) and the European Union (Erasmus+ Programme Early Nutrition eAcademy Southeast Asia-573651-EPP-1-2016-1-DE-EPPKA2-CBHE-JP). KAL is supported by the British Heart Foundation (RG/15/17/3174), Diabetes UK (16/0005454); British Heart Foundation [RG/15/17/3174]; Horizon 2020 [733206]; National Health and Medical Research Council [1059711]; National Health and Medical Research Council [353514]; National Health and Medical Research Council [403981]; National Institute for Health Research [Southampton Biomedical Research Centre]; National Institute for Health Research [NF-SI-0515-10042]; Medical Research Council (UK) [MC_UU_12011/4];

Disclosure statement

KMG has received reimbursement for speaking at conferences sponsored by companies selling nutritional products. KMG and KAL are part of an academic consortium that has received research funding from Abbott Nutrition, Nestec and Danone. Other coauthors have nothing to disclose

ORCID

R.C. Huang  <http://orcid.org/0000-0002-8464-6639>

P.E. Melton  <http://orcid.org/0000-0003-4026-2964>

R Clarke-Harris  <http://orcid.org/0000-0002-6888-9518>

K.M. Godfrey  <http://orcid.org/0000-0002-4643-0618>
 D Anderson  <http://orcid.org/0000-0003-0643-4136>
 J. M. Craig  <http://orcid.org/0000-0003-3979-7849>
 W.H. Oddy  <http://orcid.org/0000-0002-6119-7017>
 C.E. Pennell  <http://orcid.org/0000-0002-0937-6165>
 J.D. Holbrook  <http://orcid.org/0000-0003-1791-6894>
 K.A. Lillycrop  <http://orcid.org/0000-0001-7350-5489>

References

- [1] Bertram CE, Hanson MA. Animal models and programming of the metabolic syndrome. *Br Med Bull.* 2001;60(1):103–121.
- [2] Catalano PM, Kirwan JP. Maternal factors that determine neonatal size and body fat. *Curr Diab Rep.* 2001;1(1):71–77.
- [3] Rzehak P, Covic M, Saffery R, et al. DNA-methylation and body composition in preschool children: epigenome-wide-analysis in the European Childhood Obesity Project (CHOP)-study. *Sci Rep.* 2017;7(1):14349.
- [4] Godfrey KM, Sheppard A, Gluckman PD, et al. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. *Diabetes.* 2011;60(5):1528–1534.
- [5] Lillycrop K, Murray R, Cheong C, et al. ANRIL promoter DNA methylation: a perinatal marker for later adiposity. *Ebiomedicine.* 2017;19:60–72.
- [6] Hoile SP, C-H R, Huang RC, et al. Supplementation with N-3 long-chain polyunsaturated fatty acids or olive oil in men and women with renal disease induces differential changes in the DNA methylation of FADS2 and ELOVL5 in peripheral blood mononuclear cells. *Plos One.* 2014;9(10):e109896.
- [7] Dick KJ, Nelson CP, Tsaprouni L, et al. DNA methylation and body-mass index: a genome-wide analysis. *Lancet.* 2014;383(9933):1990–1998.
- [8] Aslibekyan S, Demerath EW, Mendelson M, et al. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity.* 2015;23(7):1493–1501.
- [9] Al Muftah WA, Al-Shafai M, Zaghlool SB, et al. Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clin Epigenetics.* 2016;8(1):13.
- [10] Ronn T, Volkov P, Gillberg L, et al. Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. *Hum Mol Genet.* 2015;24(13):3792–3813.
- [11] Meeks KAC, Henneman P, Venema A, et al. An epigenome-wide association study in whole blood of measures of adiposity among Ghanaians: the RODAM study. *Clin epigenetics.* 2017;9:9.
- [12] Almen MS, Nilsson EK, Jacobsson JA, et al. Genome-wide analysis reveals DNA methylation markers that vary with both age and obesity. *Gene.* 2014;548(1):61–67.
- [13] Wang X, Zhu H, Snieder H, et al. Obesity related methylation changes in DNA of peripheral blood leukocytes. *BMC Med.* 2010;8:8.
- [14] Xu X, Su S, Barnes VA, et al. A genome-wide methylation study on obesity: differential variability and differential methylation. *Epigenetics.* 2013;8(5):522–533.
- [15] Almen MS, Jacobsson JA, Moschonis G, et al. Genome wide analysis reveals association of a FTO gene variant with epigenetic changes. *Genomics.* 2012;99(3):132–137.
- [16] Huang RC, Garratt ES, Pan H, et al. Genome-wide methylation analysis identifies differentially methylated CpG loci associated with severe obesity in childhood. *Epigenetics.* 2015;10(11):995–1005.
- [17] Richmond RC, Sharp GC, Ward ME, et al. DNA methylation and BMI: investigating identified methylation sites at HIF3A in a causal framework. *Diabetes.* 2016;65(5):1231–1244.
- [18] Rounge TB, Page CM, Lepisto M, et al. Genome-wide DNA methylation in saliva and body size of adolescent girls. *Epigenomics.* 2016;8(11):1495–1505.
- [19] Ligthart S, Marzi C, Aslibekyan S, et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* 2016;17(1):255.
- [20] Su S, Zhu H, Xu X, et al. DNA methylation of the LY86 gene is associated with obesity, insulin resistance, and inflammation. *Twin Res Hum Genet.* 2014;17(3):183–191.
- [21] Newnham JP, Evans SF, Michael CA, et al. Effects of frequent ultrasound during pregnancy - a randomized controlled trial. *Lancet.* 1993;342(8876):887–891.
- [22] Straker L, Mountain J, Jacques A, et al. Cohort profile: the Western Australian pregnancy cohort (Raine) study-generation 2. *Int J Epidemiol.* 2017;46(5):1384–+.
- [23] Stick SM, Burton PR, Gurrin L, et al. Effects of maternal smoking during pregnancy and a family history of asthma on respiratory function in newborn infants. *Lancet.* 1996;348(9034):1060–1064.
- [24] Robinson M, Mattes E, Oddy WH, et al. Prenatal stress and risk of behavioral morbidity from age 2 to 14 years: the influence of the number, type, and timing of stressful life events. *Dev Psychopathol.* 2011;23(2):507–520.
- [25] Huang RC, Burrows S, Mori TA, et al. Lifecourse adiposity and blood pressure between birth and 17 years old. *Am J Hypertens.* 2015;28(8):1056–1063.
- [26] Kim SK, Kim HJ, Hur KY, et al. Visceral fat thickness measured by-ultrasonography can estimate not only visceral obesity but also risks of cardiovascular and metabolic diseases. *Am J Clin Nutr.* 2004;79(4):593–599.
- [27] Huang R-C, Mori TA, Burrows S, et al. Sex dimorphism in the relation between early adiposity and cardiometabolic risk in adolescents. *J Clin Endocrinol Metab.* 2012;97(6):E1014–E22.
- [28] Ambrosini GL, de Klerk NH, O'Sullivan TA, et al. The reliability of a food frequency questionnaire for use among adolescents. *Eur J Clin Nutr.* 2009;63(10):1251–1259.

- [29] Ambrosini GL, Oddy WH, Robinson M, et al. Adolescent dietary patterns are associated with lifestyle and family psycho-social factors. *Public Health Nutr.* **2009**;12(10):1807–1815.
- [30] Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **1988**;16(3):1215.
- [31] Fortin JP, Fertig E, Hansen K. shinyMethyl: interactive quality control of illumina 450k DNA methylation arrays in R. *F1000Research.* **2014**;3(3):175.
- [32] van Iterson M, Tobi EW, Sliker RC, et al. MethylAid: visual and interactive quality control of large illumina 450k datasets. *Bioinformatics.* **2014**;30(23):3435–3437.
- [33] Assenov Y, Muller F, Lutsik P, et al. Comprehensive analysis of DNA methylation data with RnBeads. *Nat Methods.* **2014**;11(11):1138–1140.
- [34] Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in illumina infinium 450 k DNA methylation data. *Bioinformatics.* **2013**;29(2):189–196.
- [35] Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinf.* **2012**;13(1):86.
- [36] Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* **2014**;30(10):1363–1369.
- [37] Pedersen BS, Schwartz DA, Yang IV, et al. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics.* **2012**;28(22):2986–2988.
- [38] Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin.* **2015**;8(1):6.
- [39] Sidak Z. Rectangular confidence regions for the means of multivariate normal distributions. *J Am Stat Assoc.* **1967**;62(318):626–633.
- [40] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser B (Methodol).* **1995**;57(1):289–300.
- [41] Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from illumina's HumanMethylation450 platform. *Bioinformatics.* **2016**;32(2):286–288.
- [42] Understanding Bland GD. Altman analysis. *Biochem Med (Zagreb).* **2015**;25(2):141–151.
- [43] Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinf.* **2012**;13:13.
- [44] Pan H, Holbrook JD, Karnani N, et al. Gene, environment and methylation (GEM): a tool suite to efficiently navigate large scale epigenome wide association studies and integrate genotype and interaction between genotype and environment. *BMC Bioinf.* **2016**;17:17.
- [45] Barton SJ, Melton PE, Titcombe P, et al. In epigenomic studies, including cell-type adjustments in regression models can introduce multicollinearity, resulting in apparent reversal of direction of association. *Front Genet.* **2019**;10:816.
- [46] Sakakibara S, Nakamura Y, Satoh H, et al. Rna-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. *J Neurosci.* **2001**;21(20):8091–8107.
- [47] Dhana K, Braun KVE, Nano J, et al. An epigenome-wide association study of obesity-related traits. *Am J Epidemiol.* **2018**;187(8):1662–1669.
- [48] Shah S, Bonder MJ, Marioni RE, et al. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am J Hum Genet.* **2015**;97(1):75–85.
- [49] Pan WW, Myers MG. Leptin and the maintenance of elevated body weight. *Nat Rev Neurosci.* **2018**;19(2):95–105.
- [50] Cabral M, Severo M, Barros H, et al. Longitudinal association of adiposity and high-sensitivity C-reactive protein from adolescence into early adulthood. *Nutr Metab Cardiovasc Dis.* **2019**;29(6):590–597.
- [51] Das K, Lewis RY, Combatsiaris TP, et al. Predominant expression of the mitochondrial dicarboxylate carrier in white adipose tissue. *Biochem J.* **1999**;344(2):313–320.
- [52] Mizuarai S, Miki S, Araki H, et al. Identification of dicarboxylate carrier Slc25a10 as malate transporter in de novo fatty acid synthesis. *J Biol Chem.* **2005**;280(37):32434–32441.
- [53] Kulyte A, Ehrlund A, Arner P, et al. Global transcriptome profiling identifies KLF15 and SLC25A10 as modifiers of adipocytes insulin sensitivity in obese women. *PLoS One.* **2017**;12(6):6.
- [54] Kato N, Loh M, Takeuchi F, et al. Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation. *Nat Genet.* **2015**;47(11):1282–1293.
- [55] Lafuente EM, van Puijenbroek AA, Krause M, et al. RIAM, an Ena/VASP and profilin ligand, interacts with Rap1-GTP and mediates Rap1-induced adhesion. *Dev Cell.* **2004**;7(4):585–595.
- [56] De Andrade FA, Batistela MS, Amaral SD, et al. Association between RAPH1 gene haplotypes and CHE2 locus phenotypes. *Ann Hum Genet.* **2016**;80(4):203–209.
- [57] Iwasaki T, Yoneda M, Nakajima A, et al. Serum butyrylcholinesterase is strongly associated with adiposity, the serum lipid profile and insulin resistance. *Int Med.* **2007**;46(19):1633–1639.
- [58] Huang TTK, Johnson MS, Figueroa-Colon R, et al. Growth of visceral fat, subcutaneous abdominal fat, and total body fat in children. *Obes Res.* **2000**;8:16S–S.
- [59] Xu HY, Barnes GT, Yang Q, et al. Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. *J Clin Invest.* **2003**;112(12):1821–1830.