The association of DNA methylation with body mass index: distinguishing between predictors and biomarkers

Zoe E. Reed[1*], Matthew J. Suderman[1], Caroline L. Relton[1], Oliver S.P. Davis[1,2], Gibran Hemani[1]

1 Medical Research Council Integrative Epidemiology Unit (MRC IEU), Department of Population Health Sciences, Bristol Medical School, University of Bristol, UK

2 The Alan Turing Institute, British Library, 96 Euston Road, London, UK

*Corresponding author (zoe.reed@bristol.ac.uk)

Abstract

Background:

DNA methylation is associated with body mass index (BMI), but it is not clear if methylation scores are biomarkers for extant BMI, or predictive of future BMI. Here we explore the causal nature and predictive utility of DNA methylation measured in peripheral blood with BMI and cardiometabolic traits.

Methods:

Analyses were conducted across the life course using the ARIES cohort of mothers (n=792) and children (n=906), for whom DNA methylation and genetic profiles and BMI at multiple time points (3 in children at birth, in childhood and in adolescence, 2 in mothers during pregnancy and in middle age) were available. Genetic and DNA methylation scores for BMI were derived using published associations between BMI and DNA methylation and genotype. Causal relationships between methylation and BMI were assessed using Mendelian randomisation and cross-lagged models.

Results:

The DNA methylation scores in adult women explained 10% of extant BMI variance. However, less extant variance was explained by scores generated in the same women during pregnancy (2% BMI variance) and in older children (15-17 years; 3% BMI variance). Similarly, little extant variance was explained in younger children (at birth and at 7 years; 1% and 2%, respectively). These

associations remained following adjustment for smoking exposure and education levels. The DNA methylation score was found to be a poor predictor of future BMI using linear and cross-lagged models, suggesting that DNA methylation variation does not cause later variation in BMI. However, there was some evidence to suggest that BMI is predictive of later DNA methylation. Mendelian randomisation analyses also support this direction of effect, although evidence is weak. Finally, we find that DNA methylation scores for BMI are associated with extant cardiometabolic traits independently of BMI and genetic score.

Conclusion:

The age-specific nature of DNA methylation associations with BMI, lack of causal relationship, and limited predictive ability of future BMI, indicate that DNA methylation is likely influenced by BMI and might more accurately be considered a biomarker of BMI and related outcomes than a predictor. Future epigenome-wide association studies may benefit from further examining associations between early DNA methylation and later health outcomes.

Keywords: BMI - DNA methylation - ALSPAC - ARIES - Longitudinal - Mendelian randomization

Background

Obesity has a considerable burden on healthcare and has been shown to be predictive of mortality [1]. In recent years there has been a gradual increase in body mass index (BMI) in many countries [2] and interventions to decrease BMI have had limited success [3,4]. Predicting BMI early on and performing targeted interventions is an alternative strategy that could be more effective.

The aetiology of BMI comprises both genetic and environmental factors, with heritability likely below 0.5 [5,6]. It has been suggested that natural variation in DNA methylation levels may be a risk factor for certain diseases and play a role in the phenotypic variation of many traits [7–9]. In some cases, DNA methylation may provide the molecular link between environmental factors and associated disease risk, for example, recent studies have suggested that DNA methylation may be the mechanism allowing environmental factors or increased BMI to lead to obesity-related health outcomes [10–12]. Therefore, it could be useful as a predictor of such health outcomes.

Recent studies [13–15] suggest that DNA methylation associates with BMI trait variance independent of genetic variation. Associated genes have been shown to be involved in processes such as metabolism, inflammation, metabolic disease and cardiovascular disease amongst others. This suggests that associated DNA methylation loci may belong to causal pathways linking BMI and metabolic, cardiovascular and other obesity-related health outcomes, but this requires further exploration.

Since genetic variants are fixed at conception, genetic variants associated with BMI can be used early in life to predict later BMI [16]. The relationship of DNA methylation at BMI associated loci, however, is more complex. Methylation levels vary over time and may change in response to environmental or phenotypic changes so earlier methylation variation is not guaranteed to predict later BMI levels. Recent work [14] has suggested that change in BMI is more likely to be causal for changes in DNA methylation than vice versa. This would suggest that current DNA methylation scores are simply *biomarkers* for extant BMI. However, the utility of DNA methylation as a *predictor* for future trajectories of BMI would be of considerably greater utility.

The first aim of our study was to investigate if there is a temporal association between DNA methylation and BMI. We approached this question by using genome-wide DNA methylation profiles from the Accessible Resource for Integrated Epigenomics Studies (ARIES) [17] subset of the Avon Longitudinal Study of Parents and Children (ALSPAC) [18,19]. ALSPAC is a prospective cohort of children born in the former county of Avon, England during 1991 and1992. DNA methylation profiles were generated in children from blood collected at three time points (birth, childhood, adolescence) and from their mothers at two time points (during pregnancy and at middle-age). We used these DNA methylation profiles along with multiple measurements of BMI genetic profiles to determine if DNA methylation predicted BMI later in life, independently of genetic variation and BMI itself, and vice versa. In doing so, our objective was to determine if DNA methylation is a predictor for BMI or simply a biomarker that proxies current or previous BMI values.

105 DNA methylation scores for BMI may be useful for the detection of adverse
106 health outcomes related to BMI. For example, *Wahl et al* found that DNA
107 methylation could be used to identify individuals at high risk of incident type 2
108 diabetes, independently of other explanatory factors, including BMI itself [14].
109 Therefore, the second aim of our study was to see if DNA methylation scores for
110 BMI contributed anything above BMI itself in predicting related adverse health
111 outcomes.

112

113

114

115

116

117

118

119

120

121

122

Methods

Cohort description

We used children and mothers data from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort in this study [18,19]. The ALSPAC cohort is a prospective birth cohort study in which 14,541 pregnant women living in Avon, UK, with an expected delivery date from 1st April 1991 to the 31st December 1992 were initially recruited. Of these, 13,988 children were still alive 1 year later and have been followed-up with regular questionnaires and clinical measures, providing behavioural, lifestyle and biological data. When the children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. As a result, when considering variables collected after age 7, the total sample size for those alive at 1 year of age is 14,901. The study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool http://www.bristol.ac.uk/alspac/researchers/our-data/.

We only included participants that were also in the sub-study Accessible Resource for Integrated Epigenomic Studies (ARIES), where methylation data was available for these individuals [17]. After excluding those without methylation or phenotypic data and those that had withdrawn consent, we had data available for analyses from 823 children at birth, 906 for childhood (age 7), 770 for adolescence (age 15) and 792 for pregnant mothers and 726 for middle-aged mothers. The mean age and BMI are presented in **Table 1**. Sex is also included for children only, as adults were all female.

146 Table 1. Cohort description

	Children			Mothers	
	Birth (n=823)	**Childhood (n=906)**	**Adolescence (n=770)**	**Pregnancy (n=792)**	**Middle-age (n=726)**
Mean age in years (SD)	NA	7.45 (0.15)	17.35 (0.88)	28.87 (4.30)	47.66 (4.26)
Percentage of females	51.03%	50.44%	53.38%	100%	100%
Mean BMI (kg/m^2) or birthweight (grams) (SD)	3490.93 (477.76)	16.20 (2.04)	22.58 (3.83)	22.72 (3.63)	26.41 (5.09)

147 *SD = Standard Deviation, BMI= Body Mass Index*

148 Ethics

149 Ethical approval for the study was obtained from the ALSPAC Ethics and Law
150 Committee and the Local Research Ethics Committees. Informed consent for the
151 use of data collected via questionnaires and clinics was obtained from
152 participants following the recommendations of the ALSPAC Ethics and Law
153 Committee at the time.

154 Phenotypic measures

155 The time points investigated in this study are pregnancy and middle-age data for
156 the mothers and birth, childhood (age 7) and adolescence (age 15-17 for
157 methylation data and age 17 for BMI data) data for the children.

158 Measures of height and weight were collected at research clinic visits. Height
159 was measured to the nearest millimetre using the Harpenden Stadiometer.

Weight was measured using the Tanita Body Fat Analyser to the nearest 50g. BMI was calculated by dividing weight (kilograms) by height (meters) squared (kg/m^2). For measurements at birth, birth weight was collected and is used here instead of BMI.

Data for smoking and highest education level of mothers, age, sex, sample type and cell type proportions (B cell, CD4T, CD8T, Gran, Mono and NK) were used as covariates in analyses (for children, information on matched mothers smoking and maternal education were used). Various answers from questions regarding smoking in mothers (in pregnancy and currently) were used, from questionnaire data. These included the number of cigarettes smoked per day before pregnancy, during the first three months of pregnancy and the number of cigarettes smoked in the last 2 weeks during or just after pregnancy. These measures were combined to create a variable for whether the mother smoked in pregnancy. We also used data on the number of cigarettes smoked per day, whether they responded as being a current smoker and the time passed since they stopped smoking if this was within the last 12 months to create a variable for current or recent (within the last 12 months) smoker. Education level for the mother's was also taken from questionnaire data, where participants were asked "What educational qualifications do you, your partner, your mother, and your father have?". They were asked to select all options that applied to them and we used the highest education qualification for the participant. The options for this were CSE/none, vocational, O level, A level or degree.

Variables for cardiovascular outcomes for middle-aged mothers and adolescents were also used. These variables were from blood samples, obtained during clinic

184 visits and included fasting glucose and insulin, triglycerides and low-density
185 lipoprotein (LDL). Sitting diastolic and systolic blood pressures (SBP) from the
186 right arm were collected during clinic visits. An Omron M6 upper arm blood
187 pressure/pulse monitor was used to take 2 readings of blood pressure and then
188 mean values were used. The data for triglycerides, glucose and insulin were
189 skewed so we log transformed this data for use in analyses.

190 Methylation data

191 Methylation profiling in the ARIES subset was conducted using DNA samples
192 from blood taken at clinic visits or after delivery from the umbilical cord in the
193 case of the birth time point. Blood from 1,018 mother–child pairs (children at
194 three time points and their mothers at two time points) were selected for
195 analysis as part of the Accessible Resource for Integrative Epigenomic Studies
196 (ARIES, http://www.ariesepigenomics.org.uk/) [17]. Following DNA extraction,
197 samples were bisulphite converted using the Zymo EZ DNA Methylation™ kit
198 (Zymo, Irvine, CA, USA). Following conversion, genome-wide methylation was
199 measured using the Illumina Infinium HumanMethylation450 (HM450)
200 BeadChip. The arrays were scanned using an Illumina iScan, with initial quality
201 review using GenomeStudio. ARIES data were pre-processed and normalised
202 using the *meffil* R package [20]. ARIES consists of 5,469 DNA methylation profiles
203 obtained from 1,022 mother-child pairs measured at five time points. Low
204 quality profiles were removed from further processing, and the remaining 4,593
205 profiles were normalised using the Functional Normalization algorithm [21]
206 with the top 10 control probe principal components. From the ARIES dataset,
207 sample type and normalised methylation data was extracted and cell type

proportion data were estimated using the Houseman method [22]. Full details of the pre-processing and normalization of ARIES has been described previously [20].

Genotyping

Genetic data were collected from blood samples obtained in clinic visits. Genotyping was conducted with the Illumina HumanHap550 quad chip for children and the Illumina human660W-quad array for mothers. Quality control measures were carried out and haplotypes estimated using ShapeIT. A phased version of the 1000 genomes reference panel from the Impute2 reference data repository was used and Imputation of the target data was performed with this, using all reference haplotypes. A large proportion of the cohort has genome-wide data from these samples and a subset of this data is used in this study [18].

Genetic and epigenetic scores

To investigate whether reported DNA methylation associations with BMI could be observed in an independent cohort, we calculated DNA methylation scores from published effect sizes for 135 CpG sites from the *Mendelson et al* [15] meta-analysis of DNA methylation and BMI. Scores were obtained for each ARIES methylation profile by multiplying the CpG site methylation levels in that profile with the corresponding published effects estimates and then summing the products.

Genetic scores were similarly derived using effect sizes for 97 SNPs from the GIANT consortium BMI genome-wide association study (GWAS) [23]. Scores

were created using Plink V1.9 (https://www.cog-genomics.org/plink2) with the score and sum commands, however one of the SNPs did not meet imputation score filters (rs12016871), so the score was constructed using only 96 SNPs.

To perform simplified versions of Mendelian randomisation (MR), we used summary statistics from methylation quantitative trait loci (mQTL). mQTLs are genetic variants associated with DNA methylation [24]. To identify mQTLs we looked these up in mQTLdb [25], which contains mQTLs below a conservative p-value threshold of 1e-07. If multiple mQTLs were identified for an individual CpG site, the one with the lowest p-value reported in the GIANT study for BMI was used as the mQTL for the MR analysis. If these mQTLs were unavailable, then proxies of these SNPs were obtained. These were SNPs with the next lowest p-value for that CpG, which were also present in the BMI GWAS data. We used the last p-value available in the BMI GWAS for each mQTL to maximise power. Of the 135 CpG sites we queried, 89 had an instrument at this threshold.

Statistical analysis

Observational associations at the same time point

Linear regression models, with adjustments for covariates, were used to test observational associations. When testing for association between genetic and/or methylation scores of BMI, BMI was the dependent variable and the methylation and/or genetic score the independent variables. For models including a genetic score, age was included as a covariate. For models including a methylation score, the covariates included were age, sample type and estimated blood cell type

proportions. Sex was additionally included as a covariate in all models analysing child data. For predicting BMI, BMI was used as the dependent variable, and when predicting methylation, methylation score was used as the dependent variable.

To compare the relative contributions of genetic score and methylation score to BMI, an analysis of variance (ANOVA) test was carried out comparing the following three models, with the full model (model 3) being compared to each of the reduced models (models 1 and 2):

1. BMI ~ methylation score + covariates
2. BMI ~ genetic score + covariates
3. BMI ~ methylation score + genetic score + covariates

Finally, we investigated how BMI and DNA methylation change over the life course. Firstly, we calculated correlations of BMI and DNA methylation score at different time points for mothers and children separately. We then examined the correlation of BMI and DNA methylation scores between paired children and mothers at the different time points. Thirdly, we also calculated correlations for all individual CpG sites across the different timepoints and between paired child's cord blood DNA and mother's antenatal peripheral blood DNA values.

<u>Observational associations across the life course</u>

To investigate whether DNA methylation might be predictive of later BMI or vice versa, we assessed associations between different time points in mothers and

children, using linear regression models, similar to those used in the observational analyses within the same time point.

Exploration of a causal relationship between BMI and DNA methylation

Cross-lagged model

We analysed the temporal relationship of BMI and DNA methylation, using a cross-lagged model. This approach allows exploration of the relationships between earlier BMI and later methylation score in two separate systems, one in the children (using childhood and adolescence) and one in the mothers (using the antenatal and middle age time points). The R package OpenMx (version 2.13.2) [26] was used to build a cross-lagged model, shown in Figures 3 and 4. Values for each of the free parameters or paths are estimated in the model. The paths were from earlier BMI to methylation at the same time point, later methylation and later BMI; and from earlier methylation to later BMI and methylation; and from later BMI to methylation at the same time point. Each path was sequentially tested in a sub model analysis, where that path was fixed to 0 and this sub model was compared against the full model using a likelihood ratio test. If a sub model had a significantly worse fit then that path was retained, but otherwise dropped because it was not important to the overall system.

Mendelian randomisation

To investigate causal inference more directly, an approach based on MR was adopted. To test if changes in BMI cause changes in DNA methylation, we calculated genetic scores for BMI, as previously described, and tested the

association of this score with each of the 135 BMI-associated CpG sites. A Fisher's test was then applied to combine the association p-values for all 135 association tests. To increase power, the adolescent and middle-aged mother's data was subsequently combined, and the association was tested again using a mixed model to account for relatedness.

Two-sample MR was applied to explore the reverse direction, i.e. a causal effect of DNA methylation on BMI. In this approach, summary statistics from the BMI GWAS were obtained for the mQTLs (or proxies of these, if these SNPs were unavailable) for the 135 BMI-associated CpG sites.

Confounder analysis

To investigate whether any associations found between methylation and BMI were due to confounding by smoking or education, we compared linear models of BMI and DNA methylation with and without smoking (prenatal smoking during pregnancy for children and own smoking for adults) and education as covariates.

Cardiovascular trait analyses

Linear regression models were used to test observational associations between the methylation and genetic scores for BMI against cardiovascular outcomes. These were performed with and without adjustment for BMI and using other covariates, as with other models. ANOVA tests were used to compare the relative contributions of BMI and the genetic and methylation scores and results are reported for these comparisons.

Results

Establishing associations between DNA methylation and BMI within time-points

The methylation score for BMI, derived from the *Mendelson et al* epigenome-wide association study (EWAS) [15], was strongly associated with BMI in middle aged mothers, explaining 10% of the variation in BMI (p=1.58E-23). The association was weaker in mothers during pregnancy (2% variance explained, p=2.31E-06) and in children at birth (2% variance explained, p=6.83E-05), childhood (1% variance explained, p=2.23E-04) and adolescence (3% variance explained, p=3.91E-11). Full results are presented in **Table 2**.

Table 2. Associations between methylation and genetic scores and BMI/birth weight at the same time point.

	N	Associations between methylation score and BMI/birth weight			Associations between genetic score and BMI/birth weight		
		Effect size[1], in kg/m^2 per SD in methylation score (CI)	p-value	Adjusted R-squared[3]	Effect size[2] (CI)	p-value	Adjusted R-squared[3]
Birth	823	69.27 (35.35, 103.18)	6.83E-05	0.02	-219.75 (-412.59, -26.90)	0.03	0.004
Childhood	906	0.25 (0.12, 0.39)	2.23E-04	0.01	2.24 (1.45, 3.02)	2.72E-08	0.03
Adolescence	770	0.97 (0.69, 1.26)	3.91E-11	0.03	5.25 (3.65, 6.85)	2.12E-10	0.05
Pregnancy	792	0.62 (0.36, 0.87)	2.31E-06	0.02	2.91 (1.38, 4.43)	1.95E-04	0.02

Middle-age	726	2.06 (1.67, 2.46)	1.58E-23	0.10	4.72 (2.54, 6.90)	2.45E-05	0.02

[1] *From model adjusting for age (except at birth), sex (where applicable), sample type (where applicable), and cell type proportions (B cell, CD4T, CD8T, Gran, Mono and NK).*

[2] *From model adjusting for age and sex (where applicable)*

[3] *Adjusted R-squared obtained from model with only the methylation score or genetic score and no other covariates.*

Genetic scores for BMI, derived from published GWAS summary statistics [23], were also associated with BMI, with the strongest association found for children in adolescence with 5% variance explained in BMI (p=2.12E-10) and weaker associations found at all other time points (see **Table 2** for full results).

The genetic and methylation score associations appear to be mostly independent (**Table 3**), as the combined model with both genetic and methylation scores performed better than both the methylation score (ANOVA test p-values ranged from 1.93E-10 to 0.04) and genetic score models (ANOVA test p-values ranged from 4.55E-21 to 1.28E-03) alone for all time points. This validates previous findings that the genotype and DNA methylation explain independent subsets of BMI variation [13,15].

Table 3 Results from combined model and ANOVA comparing this with models for methylation and genetic scores.

	Combined model adjusted R-squared[1]	**ANOVA (model 1 vs model 3)**	**ANOVA (model 2 vs model 3)**
Birth	0.03	0.04	1.56E-09
Childhood	0.04	9.03E-09	1.28E-03
Adolescence	0.08	1.93E-10	5.28E-11
Pregnancy	0.04	2.22E-04	5.36E-05
Middle-age	0.12	2.77E-05	4.55E-21

[1] *Adjusted R-squared obtained from model with both the methylation and the genetic scores and no other covariates*

Model 1 includes the methylation score, model 2 includes the genetic score and model 3 is the combined model including the methylation and genetic score. All models also included the relevant covariates

Stability of phenotypic values over time and between generations

We evaluated the extent to which individual BMI levels correlated over time and between mothers and children. The strongest BMI correlations were observed in children between age 7 and adolescence and in mothers between pregnancy and middle age (R ~ 0.7). Intermediate correlations were observed between children and mothers at all time points (R ~ 0.3) except birth. Lowest BMI correlations (R ~ 0.15) were observed with birth, likely because birthweight is a different measure than BMI (**Figure 1**).

Figure 1. Correlation matrix of BMI in children (birth, childhood, adolescence) and mothers (pregnancy and middle-age).

Legend: This correlation matrix shows the correlations over time for BMI at all time points in children and mothers. There is a correlation of BMI in children and mothers over time and between paired mother and children's BMI.

The BMI methylation score correlations exhibited similar patterns but were generally lower than for BMI. Strongest correlations were observed in children between age 7 and adolescence and in mothers between pregnancy and middle age ($R \sim 0.5$). All other correlations were between 0.2 and 0.25. Thus, methylation scores at birth were more highly correlated with later time points and with maternal methylation scores than birthweight and BMI (**Figure 2**). Given the weak association of methylation in childhood with BMI, factors other than BMI likely contribute to the correlation of DNA methylation over time and between mothers and children.

Figure 2. Correlation matrix of methylation score in children (birth, childhood, adolescence) and mothers (pregnancy and middle-age).

Legend: This correlation matrix shows the correlations over time for overall methylation score at all time points in children and mothers. There is a correlation of methylation score in children and mothers over time and between paired mother and children's methylation scores.

Finally, to examine whether there are particular CpG sites that correlate more strongly over time and between paired children and mothers, we tested the correlation of each site at different time points (Supplementary Table 1, Additional file 1). We observe that median correlations across all CpG sites follow a similar pattern to correlations of the methylation scores over time,

where the strongest correlations were observed in children between age 7 and adolescence and in mothers between pregnancy and middle age (R ~ 0.2). Only these two sets of timepoints have CpG sites with correlation R > 0.5. There are six such CpG sites for each and two of these are common to both (cg16611584, cg24145109). Both of these CpG sites are highly correlated across all time points.

Predicting future BMI with past DNA methylation scores

We next investigated whether the BMI methylation score was predictive of BMI at later timepoints (**Table 4**) or vice versa (if BMI was predictive of methylation score at later timepoints; **Table 5**). We observed some evidence that methylation score in childhood could be predictive of BMI in adolescence (p=0.004), although the association disappeared when adjusting for childhood BMI (p=0.20) and there was stronger evidence for the converse, that is BMI in childhood predicting adolescent DNA methylation (p=1.52E-06 even when adjusting for childhood methylation score). We observed the same in the mothers between pregnancy and middle age; that is, the association between antenatal (earlier) methylation score and middle-age (later) BMI (p=0.02) essentially disappears when adjusting for antenatal (earlier) BMI (p=0.13). Also, the association between antenatal (earlier) BMI and middle age (later) DNA methylation is much stronger even when adjusting for antenatal (earlier) methylation score (p=5.48E-11).

Table 4. Associations between methylation score and BMI at later time points.

	N	**Unadjusted for current BMI/birthweight[1]**	**Adjusted for current BMI/birthweight[1]**

		Effect size[1]	CI	p-value	Adjusted R-squared[2]	Effect size[1]	CI	p-value
Methylation score at birth, BMI in childhood	830	-0.03	-0.18, 0.13	0.72	-0.0004	-0.07	-0.22, 0.07	0.36
Methylation score at birth, BMI in adolescence	712	0.06	-0.25, 0.36	0.70	-0.001	-0.05	-0.35, 0.25	0.73
Methylation score in childhood, BMI in adolescence	762	0.41	0.13, 0.69	0.004	0.005	0.13	-0.07, 0.34	0.20
Antenatal methylation score, BMI in middle-age	765	0.80	0.44, 1.17	1.92E-05	0.02	0.19	-0.06, 0.44	0.13

[1] *From model adjusting for age (except at birth), sex (where applicable), sample type (where applicable), and cell type proportions (B cell, CD4T, CD8T, Gran, Mono and NK).*

[2] *Adjusted R-squared obtained from model with only the methylation score and no other covariates*

Table 5. Associations between BMI/birthweight and methylation score at later time points.

	N	Unadjusted for current methylation score[1]				Adjusted for current methylation score[1]		
		Effect size[1]	CI	p-value	Adjusted R-squared[2]	Effect size[1]	CI	p-value
Birthweight, methylation score in childhood	814	-0.0001	-0.0002, 0.00004	0.17	0.003	-0.0002	-0.0003, -0.00006	0.006
Birthweight, methylation	812	0.000003	-0.0001, 0.0001	0.96	0.002	-0.00008	-0.0002, 0.00005	0.22

score in adolescence								
BMI in childhood, methylation score in adolescence	891	0.09	0.06, 0.12	3.05E-09	0.02	0.06	0.04, 0.09	1.52E-06
Antenatal BMI, methylation score in middle-age	851	0.07	0.05, 0.08	1.42E-16	0.06	0.05	0.03, 0.05	5.48E-11

[1] *From model adjusting for age (except at birth), sex (where applicable), sample type (where applicable), and cell type proportions (B cell, CD4T, CD8T, Gran, Mono and NK).*

[2] *Adjusted R-squared obtained from model with only the methylation score and no other covariates*

Exploration of temporal relationships between BMI and DNA methylation

To further evaluate the temporal associations between DNA methylation and BMI, we used cross-lagged models to test which paths from earlier trait measures and scores were important for later trait measures and scores. Agreeing with the results from adjusted linear models, these did reveal a pathway between BMI in childhood and methylation score in adolescence. **Figure 3** shows the estimates and variances/covariances obtained from the main model. The only path that could be dropped from the model without affecting model fit was between childhood methylation and adolescent BMI ($p=0.35$ for this path and $p < 1.79E-4$ for all other paths; Supplementary Figure 1, Additional file 1). Cross-lagged model fits in mothers (**Figure 4**) also revealed a pathway from (earlier) BMI in pregnancy to (later) DNA methylation in middle-age. The only path that could be dropped from the model without affecting

model fit was from DNA methylation in pregnancy (earlier) and BMI in middle-age (later) ($p=0.20$ for this path and $p<3.65E-09$ for all other paths; Supplementary Figure 2, Additional file 1).

Figure 3. Pathway diagram for the cross-lagged model for childhood and adolescence.

Legend: This diagram shows the observed variables in boxes. Single headed arrows indicate linear regressions and double headed, curved arrows indicate variances/covariances. Estimates for the linear relationships are shown on the arrows, as are the values for variances and covariances.

Figure 4. Pathway diagram for the cross-lagged model for pregnancy and middle-aged mothers.

Legend: This diagram shows the observed variables in boxes. Single headed arrows indicate linear regressions and double headed, curved arrows indicate variances/covariances. Estimates for the linear relationships are shown on the arrows, as are the values for variances and covariances.

Mendelian randomisation does not support a causal relationship of DNAm on BMI

We used two-sample MR to explore causal relationships between DNA methylation and BMI (Supplementary Table 2, Additional file 1). Using the BMI genetic score as an instrumental variable for BMI, we found little evidence for a causal link of BMI on each of the 135 CpG sites used to construct the methylation score (p-value range: 1.63E-03 – 9.99E-01, for all timepoints, with a Bonferroni-

adjusted p-value threshold of 3.7E-04). A combined p-value for all 135 CpG sites obtained using Fisher's method similarly indicated no strong association between the genetic score and methylation (p-value range: 0.82 – 1.00, for all timepoints). Furthermore, even when this analysis was repeated with a mixed model including data from both adolescents and middle-aged mothers to increase the power (Supplementary Table 3, Additional file 1), there was still little evidence of association (p-value range: 2.13E-02 – 9.98E-01).

The reverse causal direction, methylation variation causing BMI variation, was investigated using mQTLs for the 135 methylation score CpG sites as instrumental variables. These individual tests did not indicate a causal link (Bonferroni-adjusted p-value threshold of 3.7E-04; Supplementary Table 4, Additional file 1) although combining the test p-values using Fisher's method did provide weak evidence for a causal association (Fisher's p-value = 0.03).

Confounder analyses

Sensitivity analyses showed that associations between BMI and methylation score remained unaffected by the inclusion of potential confounders in the majority of models (Supplementary Table 5, Additional file 1). Smoking and education appeared to be associated with methylation score in some models, however, most of the associations between BMI and methylation score survived these adjustments ($p < 0.007 = 0.05/7$).

Associations with cardiovascular traits

Finally, given that BMI is a risk factor for cardiovascular disease, we tested within-timepoint associations between the DNA methylation and genetic scores for BMI and cardiovascular traits to examine if the specificity of these scores. Firstly, we observed that methylation associations were partially independent of BMI for LDL (p=0.02) and glucose (p=0.03) in adolescence and triglycerides (p=3.00E-03), LDL (p=0.01) and SBP (p=0.05) in mothers at middle age. Similarly, we observed that methylation associations were partially independent of genetic effects on SBP in adolescence (p=0.05) and triglycerides in mothers (p=3.27E-03) (Supplementary Table 6, Additional file 1). We also observed that genetic effects were partially independent of BMI for LDL (p=0.03), glucose (0.05) and SBP (0.03) in adolescence and triglycerides (2.01E-04), LDL (0.02) and SBP (0.02) in mothers. Finally, we observed that the genetic effects were partially independent of methylation and BMI for LDL (p=0.02) and glucose (p=0.03) in adolescence and triglycerides (p=3.80E-03), LDL (p=0.01) and SBP (p=0.02) in mothers (Supplementary Table 7, Additional file 1).

Discussion

In this study we have demonstrated strong associations between DNA methylation and genetic scores for BMI [15] in both adults and children. Importantly, the use of temporal data indicates that the DNA methylation scores are not predictive. While the association between DNA methylation scores and BMI within time point are strong, the associations between earlier methylation scores later BMI are weak, and these signals do not improve the model of simply using earlier BMI as a predictor for later BMI. Hence, it may be more appropriate to term the DNA methylation score as a biomarker rather than a cause or predictor of BMI.

We observed within-timepoint associations between DNA methylation score for BMI and health outcomes for which BMI is a risk factor. These associations were independent of BMI and genotype, suggesting that the DNA methylation – BMI associations might arise due to unmeasured confounders that also influence those outcomes. DNA methylation could be used as a biomarker for these outcomes, above and beyond BMI and genetic variation.

Our study builds upon previous research in this field showing that both genetic and environmental factors contribute to variance in BMI [5,23]. Previous work investigating the relationship between DNA methylation and BMI has found associations with specific DNA methylation sites [10–12]. However results from these studies are fairly inconsistent and practical implications of these associations have not yet been identified [27]. In extension of this previous research we have used 135 CpG sites, identified in an EWAS meta-analysis [15],

to create methylation scores that are also associated with BMI in an independent cohort. The main novel component of our study is that we look across the life course in children and adults. Another study investigating DNA methylation profile and obesity in children aged 6-14 years also found an association between childhood obesity and a separate set of differentially methylated CpG sites, supporting our finding that DNA methylation is associated with BMI in adolescents [28]. Our findings in younger children were however weaker and could be due to basing our analysis on methylation at CpG sites associated with BMI in adults.

We have also investigated the nature of these associations further and found that there may be some predictive capability of earlier BMI to later DNA methylation at multiple time points, however we found no evidence to suggest this association is causal, which could be due to several reasons. Firstly, it may be that there is a lack of power in these analyses for MR to detect a causal relationship. The *Wahl et al* paper [14] suggests that with larger sample sizes the direction of effect is likely to be from BMI to DNA methylation. Therefore, it is possible that with a larger sample size we could confirm this direction of causality. It could also be that there are other unknown confounders, which are mediating this effect and future research should focus on investigating this further. Our MR analyses suggest a weak aggregate causal association from DNA methylation to BMI, however due to the DNAm instruments being enriched for genic regions, this association is unlikely to be stronger than expected against a more appropriate null that reflects that BMI is more strongly influenced by genetic variants in genic regions [23]. We also found associations of the BMI

539 DNAm scores with cardiovascular outcomes. If BMI causes changes in DNAm, as
540 our analyses seem to suggest, then DNAm changes may fall on the causal
541 pathway between BMI and these cardiovascular outcomes. Therefore, DNAm
542 scores may have potential use as predictors of cardiovascular outcomes,
543 although generalisability to other populations would need to be confirmed. This
544 has been suggested in previous studies for diabetes [14,29] and insulin-
545 resistance [30].

546 Limitations

547 This study is subject to a number of limitations. Firstly, whilst the size of the
548 ARIES cohort is larger than or equivalent to samples used in other DNA
549 methylation studies, it may still be too small to detect some associations, and this
550 may be why we do not find any causal association for BMI to DNA methylation,
551 as discussed above. The ARIES mQTL database used in the two-sample MR had a
552 total sample size of 3,948, although this was split across three time points in
553 approximately the same children and two time points in the mothers. We used
554 the GIANT cohort for BMI in the two-sample MR and included 235,522
555 participants in these analyses. The GIANT sample size is much larger and
556 therefore it is difficult to directly compare the strength of causal evidence in each
557 direction between these samples. Secondly, whilst it is a strength of our study
558 that we use multiple time points, in the mothers one of these is during pregnancy
559 so these findings may not be generalizable outside of pregnancy, or generalizable
560 to males. As with any epidemiological study there may also be measurement
561 error present in our phenotypic data, for example smoking data was collected via
562 self-report and there may be biases present in this data because of this. In

addition, measurement error may be present in the genotyping data and methylation data. Finally, the DNA methylation data was collected from blood samples which may not be indicative of other tissue methylation levels such as adipose tissue which may be more relevant in a study looking at potential causal relationships between BMI and DNAm, however we assume the two may be correlated to an extent.

Conclusions

In conclusion, our study finds that DNA methylation scores, as have so far been generated, are unlikely to be predictive of future BMI using earlier DNA methylation levels, and therefore is more appropriately considered a biomarker. This indicates that conducting EWAS using DNAm values and trait values measured at the same time point is not an effective strategy when attempting to create predictors. Therefore, future studies should perform EWAS that test for early DNAm values against later health outcomes to evaluate whether this may enable creating an effective predictor, which could then be tested in other populations.

583 **Abbreviations:**

584 BMI: Body mass index; ARIES: Accessible Resource for Integrated Epigenomic
585 Studies; ALSPAC: Avon Longitudinal Study of Parents and Children; LDL: Low-
586 density lipoprotein; SBP: Systolic blood pressure; MR: Mendelian randomisation;
587 mQTL: methylation quantitative trait loci; ANOVA: Analysis of variance; EWAS:
588 Epigenome-wide association study; GWAS: Genome-wide association study.

589 **Declarations:**

590 Ethics approval and consent to participate:

591 Ethical approval for the study was obtained from the ALSPAC Ethics and Law
592 Committee and the Local Research Ethics Committees. Informed consent for the
593 use of data collected via questionnaires and clinics was obtained from
594 participants following the recommendations of the ALSPAC Ethics and Law
595 Committee at the time.

596 Consent for publication:

597 Not applicable

598 Availability of data and material:

599 The ALSPAC data management plan (
600 http://www.bristol.ac.uk/alspac/researchers/data-access/documents/alspac-
601 data-management-plan.pdf) describes in detail the policy regarding data sharing,
602 which is through a system of managed open access. The steps below highlight

603 how to apply for access to the data included in this paper and all other ALSPAC
604 data. The datasets used in this analysis are linked to ALSPAC project number
605 B2154; please quote this project number during your application.

606 • 1. Please read the ALSPAC access policy (PDF, 627 kB) which describes
607 the process of accessing the data and samples in detail, and outlines the
608 costs associated with doing so.

609 • 2. You may also find it useful to browse the fully searchable ALSPAC
610 research proposals database, which lists all research projects that have
611 been approved since April 2011.

612 • 3. Please submit your research proposal for consideration by the ALSPAC
613 Executive Committee. You will receive a response within 10 working days
614 to advise you whether your proposal has been approved.

615 If you have any questions about accessing data, please email alspac-
616 data@bristol.ac.uk.

617 Competing Interests:

618 The authors declare that they have no competing interests.

619 Funding:

620 This work was supported in part by the UK Medical Research Council Integrative
621 Epidemiology Unit at the University of Bristol (Grant ref: MC_UU_00011/5). The
622 UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the

623 University of Bristol provide core support for ALSPAC. This publication is the
624 work of the authors and ZER, MJS, CLR, OSPD and GH will serve as guarantors for
625 the contents of this paper. A comprehensive list of grants funding is available on
626 the ALSPAC website
627 (http://www.bristol.ac.uk/alspac/external/documents/grant-
628 acknowledgements.pdf); This research was specifically funded by Wellcome
629 Trust (086676/Z/08/Z), Wellcome Trust and Medical Research Council
630 (076467/Z/05/Z), Wellcome Trust (WT092830/Z/10/Z), the British Heart
631 Foundation (SP/07/008/24066). ZR was supported by a Wellcome Trust PhD
632 studentship (Grant ref: 109104/Z/15/Z). MJS was supported by the ESRC
633 (ES/N000498/1). Methylation data in the ALSPAC cohort were generated as part
634 of the UK BBSRC funded (BB/I025751/1 and BB/I025263/1) Accessible
635 Resource for Integrated Epigenomic Studies
636 (ARIES, http://www.ariesepigenomics.org.uk). Child GWAS data was generated
637 by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and
638 LabCorp (Laboratory Corporation of America) using support from 23andMe.
639 Preparation of mothers GWAS data was specifically funded by the Wellcome
640 Trust (WT088806). GH was funded by the Wellcome Trust and the Royal Society
641 [208806/Z/17/Z].

642 Authors' contributions:

643 ZER, MJS, GH and CLR contributed to the planning and design of the analysis and
644 drafting of this paper. OSPD advised on the use of the cross-lagged model design
645 and contributed to final drafts of this paper.

646 Acknowledgements:

647 We are extremely grateful to all the families who took part in this study, the
648 midwives for their help in recruiting them, and the whole ALSPAC team, which
649 includes interviewers, computer and laboratory technicians, clerical workers,
650 research scientists, volunteers, managers, receptionists and nurses.

651 Additional files:

652 Additional file 1: This contains supplementary tables S1-S7 and supplementary
653 figures S1-S2. File format is .docx.

654

655

656

657

658

659

660

661

662

663 **References**

664 1. Whitlock G, Lewington S, Sherliker P, Clarke R, Emberson J, Halsey J, et al.
665 Body-mass index and cause-specific mortality in 900 000 adults: collaborative
666 analyses of 57 prospective studies. Lancet. 2009;373:1083–96.

667 2. Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ, et al.
668 National, regional, and global trends in body-mass index since 1980: systematic
669 analysis of health examination surveys and epidemiological studies with 960
670 country-years and 9·1 million participants. Lancet. 2011;377:557–67.

671 3. Tam G, Yeung MPS. A systematic review of the long-term effectiveness of
672 work-based lifestyle interventions to tackle overweight and obesity. Prev. Med.
673 (Baltim). Academic Press; 2018. p. 54–60.

674 4. Sim LA, Lebow J, Wang Z, Koball A, Murad MH. Brief Primary Care Obesity
675 Interventions: A Meta-analysis. Pediatrics. American Academy of Pediatrics;
676 2016;138:e20160149–e20160149.

677 5. Hemani G, Yang J, Vinkhuyzen A, Powell JE, Willemsen G, Hottenga JJ, et al.
678 Inference of the genetic architecture underlying bmi and height with the use of
679 20,240 sibling pairs. Am. J. Hum. Genet. Elsevier; 2013;93:865–75.

680 6. Elks CE, den Hoed M, Zhao JH, Sharp SJ, Wareham NJ, Loos RJF, et al. Variability
681 in the heritability of body mass index: a systematic review and meta-regression.
682 Front. Endocrinol. (Lausanne). 2012;3:29.

683 7. Fazzari MJ, Greally JM. Introduction to epigenomics and epigenome-wide
684 analysis. Methods Mol. Biol. 2010;620:243–65.

685 8. Dogan M V, Grumbach IM, Michaelson JJ, Philibert RA. Integrated genetic and
686 epigenetic prediction of coronary heart disease in the Framingham Heart Study.
687 PLoS One. Public Library of Science; 2018;13:e0190549.

688 9. Zhang Y, Wilson R, Heiss J, Breitling LP, Saum KU, Schöttker B, et al. DNA
689 methylation signatures in peripheral blood strongly predict all-cause mortality.
690 Nat. Commun. Nature Publishing Group; 2017;8:14617.

691 10. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aïssi D, Wahl S, et al. DNA
692 methylation and body-mass index: a genome-wide analysis. Lancet. Elsevier;
693 2014;383:1990–8.

694 11. Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, et al.
695 Epigenome-wide study identifies novel methylation loci associated with body
696 mass index and waist circumference. Obesity. 2015;23:1493–501.

697 12. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou Y-H, et
698 al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist
699 circumference in African American adults identifies multiple replicated loci.
700 Hum. Mol. Genet. 2015;24:4464–79.

701 13. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, et al.
702 Improving Phenotypic Prediction by Combining Genetic and Epigenetic
703 Associations. Am. J. Hum. Genet. Elsevier; 2015;97:75–85.

14. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2016;

15. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman ÅK, Aslibekyan S, et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. Lewis C, editor. PLoS Med. Public Library of Science; 2017;14:e1002215.

16. Khera A V., Chaffin M, Wade KH, Zahid S, Brancale J, Xia R, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. Cell. Cell Press; 2019;177:587-596.e9.

17. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, et al. Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). Int. J. Epidemiol. 2015;

18. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int. J. Epidemiol. 2013;42:97–110.

19. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int. J. Epidemiol. 2013;42:111–27.

20. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M. Meffil: efficient

normalization and analysis of very large DNA methylation datasets. Bioinformatics. Oxford University Press; 2018;34:3983–9.

21. Fortin J-P, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome Biol. BioMed Central; 2014;15:503.

22. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.

23. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2015;518:197–206.

24. Smith AK, Kilaru V, Kocak M, Almli LM, Mercer KB, Ressler KJ, et al. Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. BMC Genomics. BioMed Central; 2014;15:145.

25. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic identification of genetic influences on methylation across the human life course. Genome Biol. BioMed Central; 2016;17:61.

26. Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, et al. OpenMx: An Open Source Extended Structural Equation Modeling Framework.

746 Psychometrika. 2011;76:306–17.

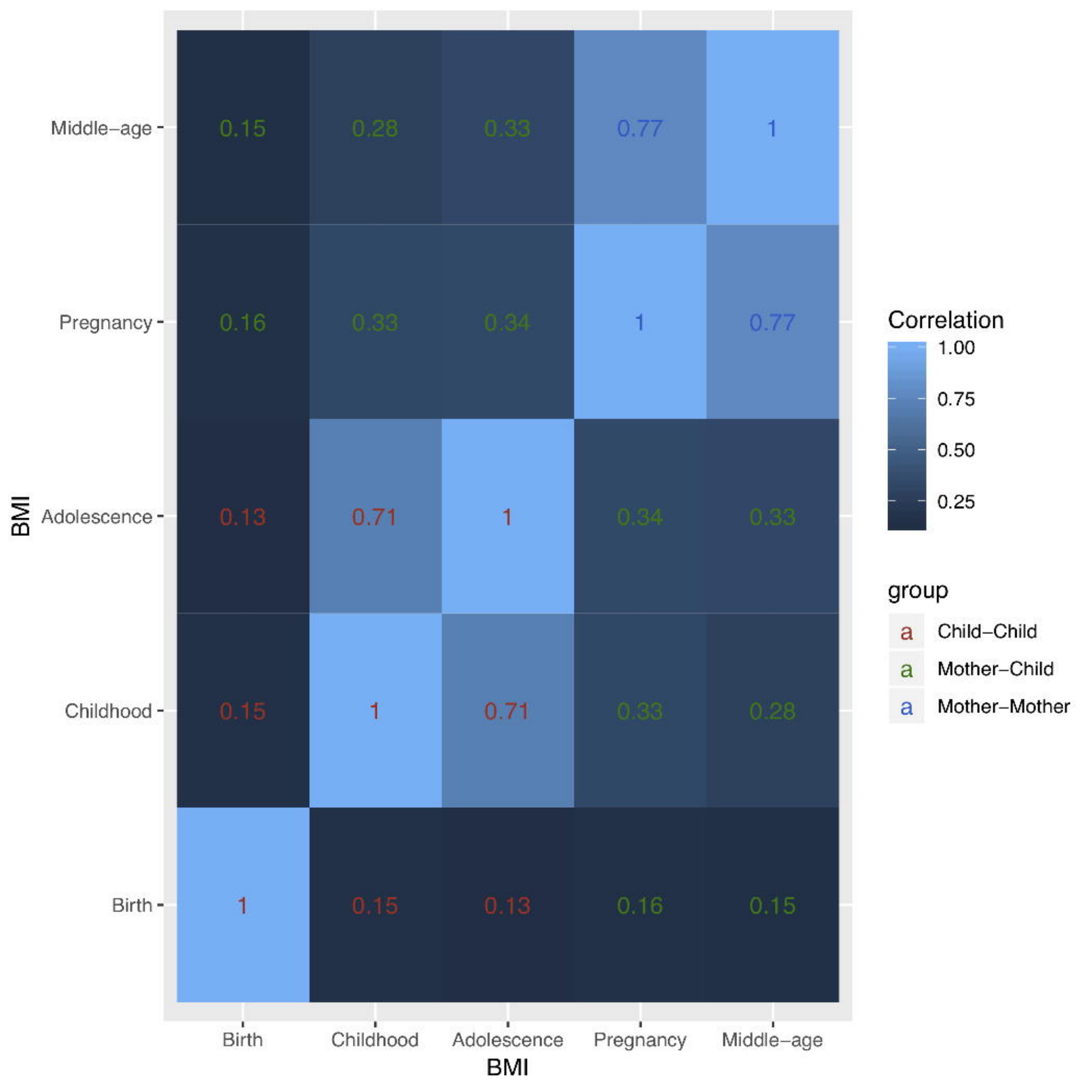
747 27. van Dijk SJ, Molloy PL, Varinli H, Morrison JL, Muhlhausler BS, Buckley M, et
748 al. Epigenetics and human obesity. Int. J. Obes. Nature Publishing Group;
749 2015;39:85–97.

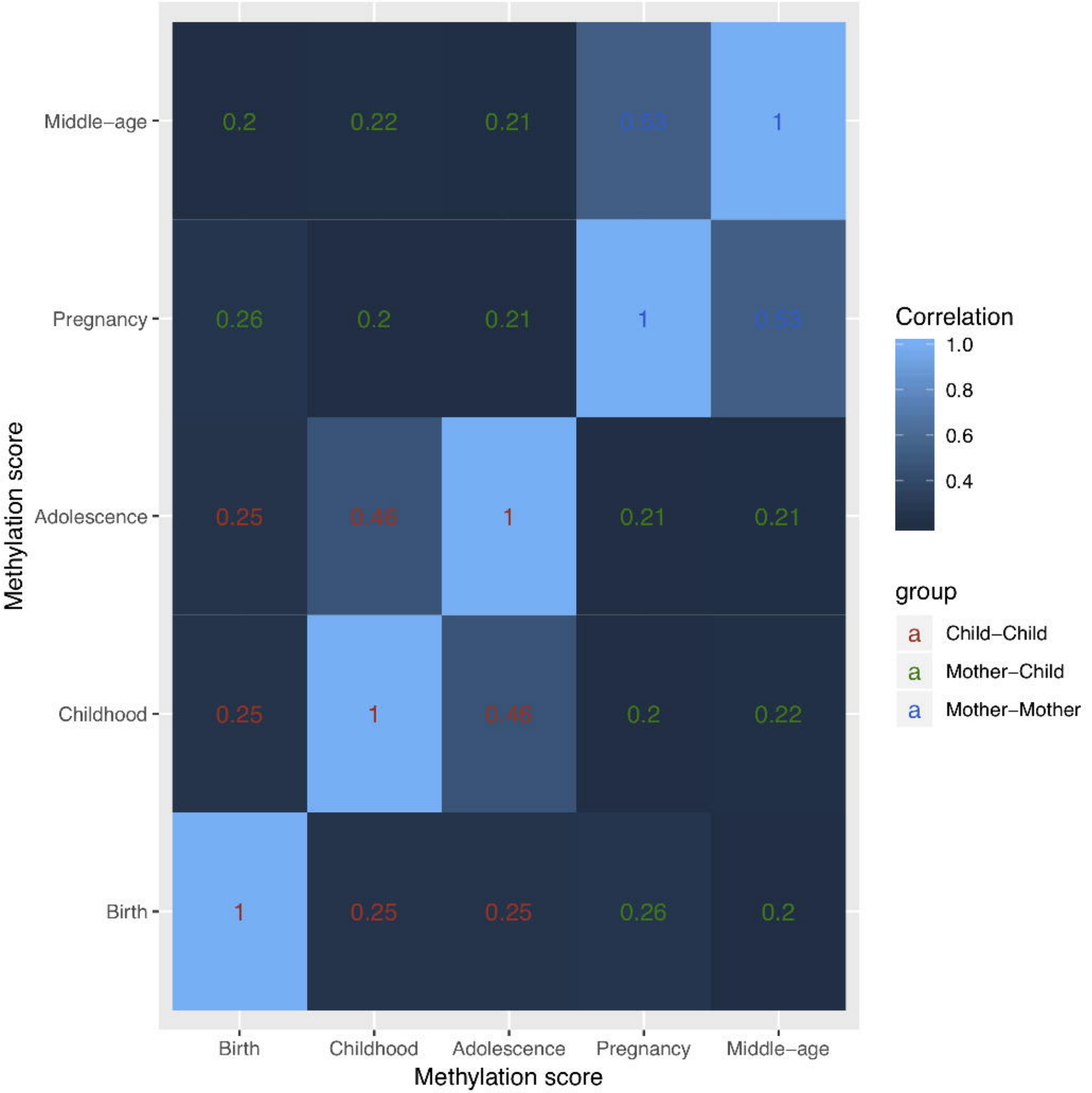
750 28. Huang RC, Garratt ES, Pan H, Wu Y, Davis EA, Barton SJ, et al. Genome-wide
751 methylation analysis identifies differentially methylated CpG loci associated with
752 severe obesity in childhood. Epigenetics. Taylor & Francis; 2015;10:995–1005.

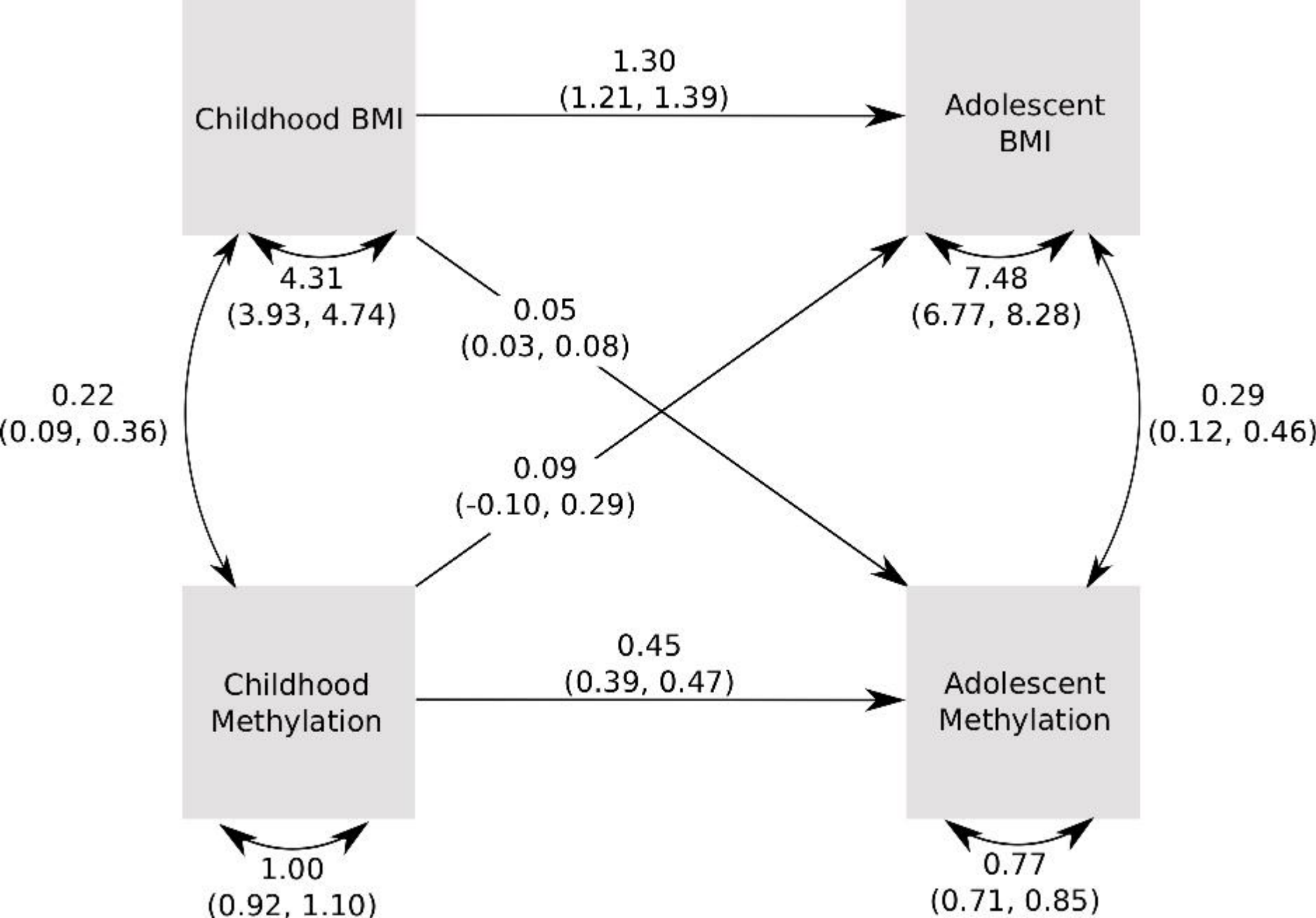
753 29. Dayeh T, Tuomi T, Almgren P, Perfilyev A, Jansson PA, de Mello VD, et al. DNA
754 methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated
755 with future type 2 diabetes risk. Epigenetics. Taylor & Francis; 2016;11:482–8.

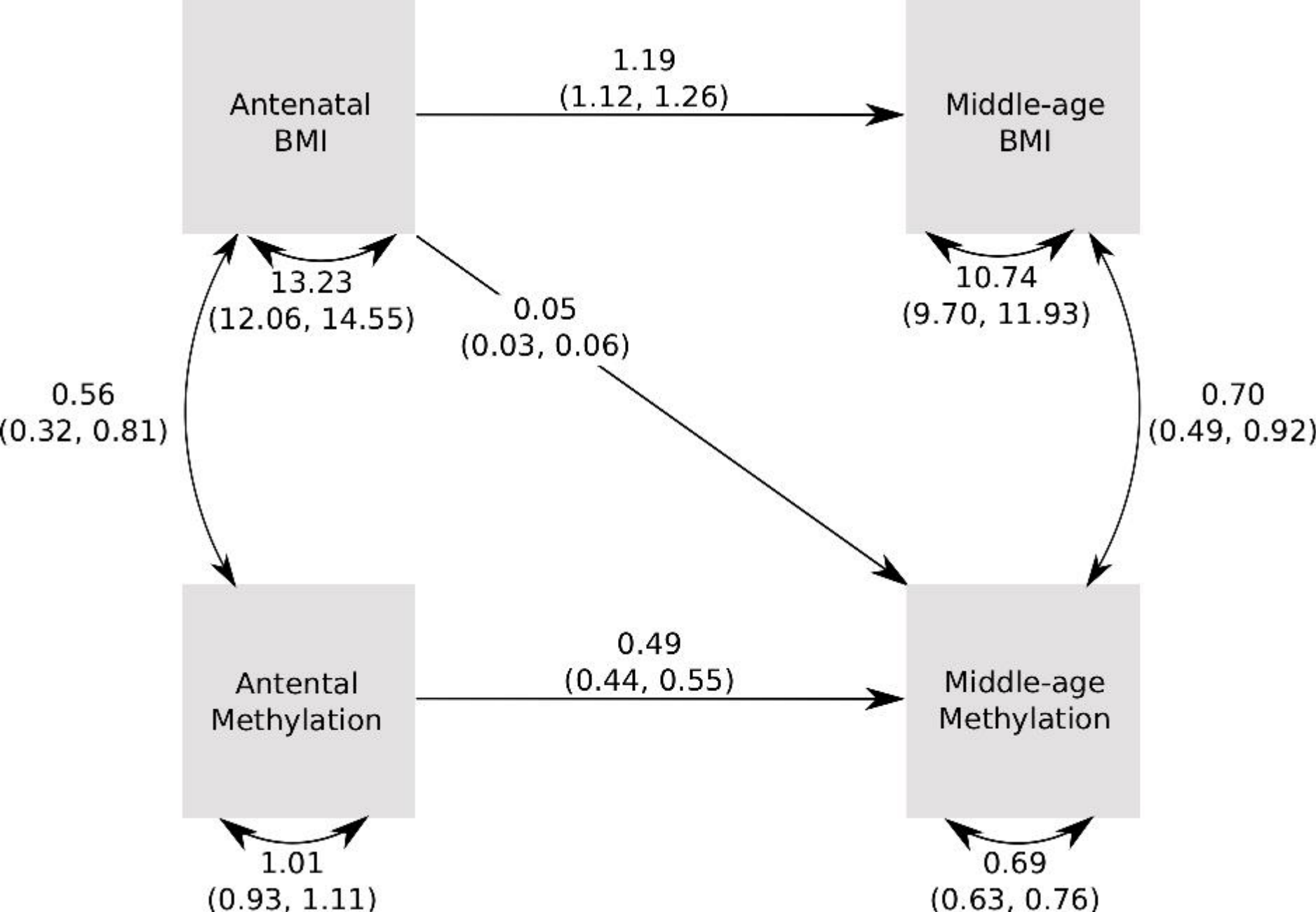
756 30. Krause C, Sievert H, Geißler C, Grohs M, El Gammal AT, Wolter S, et al. Critical
757 evaluation of the DNA-methylation markers ABCG1 and SREBF1 for Type 2
758 diabetes stratification. Epigenomics. Future Medicine Ltd London, UK;
759 2019;11:885–97.

760


Middle–age
Pregnancy
Adolescence
Childhood
Birth
BMI
0.15
0.28
0.33
0.77
1
0.16
0.33
0.34
1
0.77
0.13
0.71
1
0.34
0.33
0.15
1
0.71
0.33
0.28
1
0.15
0.13
0.16
0.15
Birth
Childhood
Adolescence
Pregnancy
Middle–age
BMI
Correlation
1.00
0.75
0.50
0.25
group
a
Child–Child
a
Mother–Child
a
Mother–Mother


Middle−age
Pregnancy
Adolescence
Childhood
Birth
Methylation score
0.2
0.22
0.21
0.53
1
0.26
0.2
0.21
1
0.53
0.25
0.46
1
0.21
0.21
0.25
1
0.46
0.2
0.22
1
0.25
0.25
0.26
0.2
Birth
Childhood
Adolescence
Pregnancy
Middle−age
Methylation score
Correlation
1.0
0.8
0.6
0.4
group
a
Child−Child
a
Mother−Child
a
Mother−Mother


Childhood BMI
1.30
(1.21, 1.39)
Adolescent
BMI
4.31
(3.93, 4.74)
0.05
(0.03, 0.08)
7.48
(6.77, 8.28)
0.22
(0.09, 0.36)
0.29
(0.12, 0.46)
0.09
(-0.10, 0.29)
Childhood
Methylation
0.45
(0.39, 0.47)
Adolescent
Methylation
1.00
(0.92, 1.10)
0.77
(0.71, 0.85)


Antenatal BMI
1.19 (1.12, 1.26)
Middle-age BMI
13.23 (12.06, 14.55)
0.05 (0.03, 0.06)
10.74 (9.70, 11.93)
0.56 (0.32, 0.81)
0.70 (0.49, 0.92)
Antental Methylation
0.49 (0.44, 0.55)
Middle-age Methylation
1.01 (0.93, 1.11)
0.69 (0.63, 0.76)