

Inference about causation between body mass index and DNA methylation in blood from a twin family study

Shuai Li,¹ Ee Ming Wong,^{2,3} Minh Bui,¹ Tuong L Nguyen,¹ Ji-Hoon Eric Joo,^{2,3} Jennifer Stone,⁴ Gillian S Dite,¹ Pierre-Antoine Dugué,^{1,5} Roger L Milne,^{1,5} Graham G Giles,^{1,5} Richard Saffery,^{6,7} Melissa C Southey,^{2,3} John L Hopper^{1*}

1. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Parkville, Victoria, Australia
2. Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne, Parkville, Victoria, Australia
3. Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia
4. Centre for Genetic Origins of Health and Disease, Curtin University and the University of Western Australia, Perth, Western Australia, Australia
5. Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Melbourne, Victoria, Australia
6. Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia
7. Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia

The authors declare no conflicts of interest.

* Corresponding author

Address: Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, 207 Bouverie Street, Parkville, Victoria 3010, Australia

Email: j.hopper@unimelb.edu.au

Abstract

Background

Several studies have reported DNA methylation in blood to be associated with body mass index (BMI), but only a few have investigated causal aspects of the association. We used a twin family design to assess this association at two life points and applied a novel analytical approach to investigate the evidence for causality.

Methods

The methylation profile of DNA from peripheral blood was measured for 479 Australian women (mean age 56 years) from 130 twin families. Linear regression was used to estimate the associations of methylation at ~410 000 cytosine-guanine dinucleotides (CpG), and of the average methylation at ~20 000 genes, with current BMI, BMI at age 18-21 years, and the change between the two (BMI change). A novel regression-based methodology for twins, Inference about Causation through Examination of Familial Confounding (ICE FALCON), was used to assess causation.

Results

At 5% false discovery rate, nine, six and 12 CpGs at 24 loci were associated with current BMI, BMI at age 18-21 years and BMI change, respectively. The average methylation of *BHLHE40* and *SOC33* loci was associated with current BMI, and of *PHGDH* locus was associated with BMI change. From the ICE FALCON analyses with BMI as the predictor and methylation as the outcome, a woman's methylation level was associated with her co-twin's BMI, and the association disappeared conditioning on her own BMI, consistent with BMI causing methylation. To the contrary, using methylation as the predictor and BMI as the outcome, a woman's BMI was not associated with her co-twin's methylation level, consistent with methylation not causing BMI.

Conclusion

For middle-aged women, peripheral blood DNA methylation at several genomic locations is associated with current BMI, BMI at age 18-21 years and BMI change. Our study suggests that BMI has a causal effect on peripheral blood DNA methylation.

Introduction

DNA methylation, that a methyl group is typically added to a cytosine-guanine dinucleotide (CpG), modifies gene expression without changing DNA sequence. Sensitive to exposures and lifestyle factors associated with health, DNA methylation has been proposed to play a critical role in the etiology of complex traits and diseases^{1, 2}.

Several studies have investigated the association between DNA methylation in blood and obesity. Differences in methylation have been observed between obese and lean people³⁻⁷. Epigenome-wide association studies (EWAS) have reported approximately 500 CpGs to be differentially methylated in relation to body mass index (BMI)⁸⁻¹⁶. Methylation at some BMI-related CpGs has also been suggested to be associated with other obesity traits such as adult waist circumference^{9, 10} and BMI change^{10, 16}, and obesity-related traits such as type 2 diabetes^{14, 17} and metabolic syndrome¹².

Most of the reported associations are from cross-sectional designs, thus the causal nature of the association, i.e. whether DNA methylation has a causal effect on BMI or *vice versa*, is unknown. There is also a possibility that the observed association is due to familial confounders¹⁸. Mendelian randomization (MR) has been proposed as a method to assess causation¹⁹, and several studies have applied MR to make causal inferences between BMI and DNA methylation^{8, 13, 14}.

As well as requiring knowledge and measurement of genetic variants, the validity of MR depends on certain assumptions, some of which are difficult to verify and need to be given attention²⁰. Especially in the context of BMI and methylation, BMI-related genetic variants are potentially associated with DNA methylation through other obesity or metabolism pathways, which violates the assumption of no directional pleiotropy. Additionally, MR requires large-scale studies with genetic variants and DNA methylation data available for the same subjects, which might not be achievable. Obviously, other approaches for assessing causation are needed if using MR is not possible.

In this study, we aimed to investigate the association between BMI and blood DNA methylation, to replicate associations reported by previous EWAS, and to investigate the causal nature of the association using a regression-based approach to data of twins.

Subjects and Methods

Study sample

The sample comprised women from the Australian Mammographic Density Twins and Sisters Study (AMDTSS)²¹. A telephone-administered questionnaire was used to collect self-reported demographic

information, height, current weight at interview and recalled weight at age 18-21 years. A total of 479 women, including 66 monozygotic twin (MZ) pairs, 66 dizygotic twin (DZ) pairs and 215 of their sisters from 130 families was selected²². Of these, one was excluded due to her current reported BMI being an outlier. The study was approved by the Human Research Ethics Committee of the University of Melbourne. All subjects provided written informed consent.

BMI traits

We studied three BMI traits at using measures at two ages: current BMI, BMI at age 18-21 years and BMI change between the two time-points. Current BMI and BMI at age 18-21 years were calculated using current height and the reported weight at the corresponding age.

DNA methylation data

DNA was extracted from dried blood spots stored on Guthrie cards using a method previously described²³. DNA was sodium bisulfite converted using the EZ DNA Methylation-Gold protocol as per manufacturers' instructions (Zymo Research, Irvine, CA) and eluted in 20 µl elution buffer. DNA Methylation was measured using the Illumina Infinium HumanMethylation450K BeadChip (HM450) array. All laboratory work was performed at the Genetic Epidemiology Laboratory, University of Melbourne.

Raw intensity data were processed by the Bioconductor *minfi* package²⁴, which included normalization of data using Illumina's reference factor-based normalization methods (*preprocessIllumina*) and subset-quantile within array normalization (SWAN)²⁵ for type I and II probe bias correction. An empirical Bayes batch-effects removal method *ComBat*²⁶ was applied to minimise the technical variation across batches. All samples passed quality control. Probes with missing value (detection *P*-value >0.01) for one or more samples, with documented SNPs at the target CpG, binding to multiple locations²⁷ or binding to the X chromosome, and 65 control probes were excluded, leaving 411 394 probes remaining for analysis. See Li *et al.*²² for more details.

Association analyses for CpG-specific methylation

We used a linear regression model to investigate the association between CpG-specific methylation values and each BMI trait, in which the methylation M-value, the logit transformation of the percentage of methylation, was the outcome and BMI trait was the predictor. The model was adjusted for age and cell-type proportions (monocytes, B cells, natural killers, CD4+ T cells, CD+8 T cells, granulocytes) estimated using the Houseman method²⁸. Parameters were estimated using the Generalised Estimating Equations (GEE) method, with family as cluster, fitted using the *geeglm()*

function from the R package *geepack*. To account for multiple testing, associations with false discovery rate (FDR)²⁹ < 0.05 were considered statistically significant.

Association analyses for gene-average methylation

We investigated associations between gene-average methylation value and each BMI trait. We annotated CpGs to genes using the column ‘UCSC_RefGene_Name’ from the Illumina’s annotation file, resulting in a total of 19 823 genes. For each gene, the average methylation was calculated as the average Beta-value, the percentage of methylation, across CpGs annotated to that gene. The gene-average methylation was then logit transformed and the association was investigated using the same model as that for CpG analysis.

Replication of previously reported CpGs and genes

462 CpGs and 332 genes at which there are CpGs previously reported to be associated with BMI⁸⁻¹⁶ were included in our study after quality control. For these CpGs and genes, we checked their associations with current BMI. For genes, the gene-average methylation was used. CpGs or genes with nominal $P < 0.05$ were considered replicated.

Causal inference analyses

We performed causal inference using Inference about Causation through Examination of FAMILiaL CONfounding (ICE FALCON), a regression-based methodology for analysing twin data³⁰⁻³⁴. By causal, it meant that if it were possible to vary a predictor measure experimentally then the expected value of the outcome measure would change.

As shown in Figure 1, suppose there are two variables, X and Y, measured for pairs of twins. Assume that X and Y are positively associated within an individual. Let S denote the unmeasured genetic and non-genetic factors that affect both twins; S_X represents those factors that influence X values only, S_Y those that influence Y values only, and S_{XY} those that influence both X and Y values. For the purpose of explanation, let ‘self’ refer to an individual and ‘co-twin’ refer to the individual’s twin, but recognise that these labels can be swapped and both twins within a pair are used in the analysis.

If there is a correlation between Y_{self} and $X_{\text{co-twin}}$, it might be due to a familial confounder, S_{XY} (Figure 1a). It could also be due to X having a causal effect on Y within an individual, provided X_{self} and $X_{\text{co-twin}}$ are correlated (Figure 1b), or to Y having a casual effect on X, provided Y_{self} and $Y_{\text{co-twin}}$ are correlated (Figure 1c). Note that the confounders specific to an individual, C_{self} and $C_{\text{co-twin}}$, do not of themselves result in a correlation between Y_{self} and $X_{\text{co-twin}}$.

Using GEE to take into account any correlation in Y between twins within the same pair, we fit three models:

$$\text{Model 1: } E(Y_{\text{self}}) = \alpha + \beta_{\text{self}} X_{\text{self}}$$

$$\text{Model 2: } E(Y_{\text{self}}) = \alpha + \beta_{\text{co-twin}} X_{\text{co-twin}}$$

$$\text{Model 3: } E(Y_{\text{self}}) = \alpha + \beta'_{\text{self}} X_{\text{self}} + \beta'_{\text{co-twin}} X_{\text{co-twin}}$$

If the correlation between Y_{self} and $X_{\text{co-twin}}$ is solely due to familial confounders (Figure 1a), the marginal association between Y_{self} and X_{self} (β_{self} in Model 1) and the marginal association between Y_{self} and $X_{\text{co-twin}}$ ($\beta_{\text{co-twin}}$ in Model 2) must both be non-zero. Adjusting for X_{self} , however, the conditional association between Y_{self} and $X_{\text{co-twin}}$ ($\beta'_{\text{co-twin}}$ in Model 3) is expected to attenuate from $\beta_{\text{co-twin}}$ in Model 2 towards the null. Similarly, adjusting for $X_{\text{co-twin}}$ (Model 3), the conditional association between Y_{self} and X_{self} (β'_{self} in Model 3) is expected to attenuate from β_{self} towards the null.

If the correlation between Y_{self} and $X_{\text{co-twin}}$ is solely due to a causal effect from X to Y (Figure 1b), by adjusting for X_{self} the pathway between these two variables is blocked and the conditional association ($\beta'_{\text{co-twin}}$ in Model 3) is expected to be null. Adjusting for $X_{\text{co-twin}}$ will have no influence on the pathway between Y_{self} and X_{self} , so the conditional association β'_{self} in Model 3 is expected to be β_{self} in Model 2.

If the correlation between Y_{self} and $X_{\text{co-twin}}$ is solely due to a causal effect from Y to X (Figure 1c), the pathway through X_{self} is blocked due to X_{self} as a collider, and the pathway through S_Y is blocked due to that GEE analysis in effect conditions on S_Y , so there is no marginal association between Y_{self} and $X_{\text{co-twin}}$, and $\beta_{\text{co-twin}}$ of Model 2 is expected to be zero.

We first applied ICE FALCON to methylation at CpGs reported by Wahl *et al.*, the majority of which has been suggested to be consequential to BMI by a bidirectional MR analysis¹⁴. We analysed a methylation score based on the 77 CpGs with $P < 0.05$ in our study. The methylation score was calculated as the sum of the products of the methylation Beta-value and the reported effect size of each CpG. The data for MZ pairs were used. The models were adjusted for age and cell-type proportions as what was done in association analyses. We first used the methylation score to be Y and current BMI to be X and regressed the methylation score on current BMI. We then swapped X and Y to regress current BMI on the methylation score and undertook the same analyses. We made statistical inference about the changes in regression coefficients by bootstrapping. That is, twin pairs were randomly sampled with replacement to generate 1,000 new datasets with the same sample size as the

original dataset. ICE FALCON was then applied to each dataset to calculate the changes in regression coefficients for that dataset.

We then applied ICE FALCON to methylation at CpGs and genes identified in our study. For each BMI trait, we similarly investigated the methylation at CpGs as a methylation score. For a locus containing multiple CpGs, only the CpG with the smallest *P*-value was included in the methylation score. BMI at age 18-21 years and its related methylation were not investigated, given that the methylation score was not correlated within MZ twin pairs ($r = -0.07$, 95% CI: -0.31, 0.17).

Results

Characteristics of the sample

For the women included in the analytic sample, the mean (standard deviation [SD]) age was 56.4 (7.9) years. The mean (SD) current BMI, BMI at age 18-21 years and BMI change was 26.8 (5.7), 21.1 (3.3) and 5.6 (5.0) kg/m², respectively. Current BMI was positively correlated with BMI at age 18-21 years ($r = 0.49$, 95% CI: 0.42, 0.55), and more strongly with BMI change ($r = 0.81$, 95% CI: 0.78, 0.84), and BMI at age 18-21 years was weakly negatively associated with BMI change ($r = -0.11$, 95% CI: -0.20, -0.02).

Associations with CpG-specific methylation

Methylation at nine, six and 12 CpGs was found to be associated with current BMI, BMI at age 18-21 years and BMI change, respectively (Table 1). The genomic inflation factor λ was 1.10, 1.03 and 1.04 for current BMI, BMI at age 18-21 years and BMI change, respectively (Q-Q plots in Supplementary Figure 1; Manhattan plots in Supplementary Figures 2-4). Methylation at two CpGs (cg12992827 and cg00636368) was negatively associated with both current BMI and BMI change. CpGs cg12992827^{10, 14}, cg18181703^{12-14, 16}, cg09349128^{10, 13, 14}, and the *NOD2* locus^{13, 14, 16} have been reported previously. The other CpGs or loci have not been reported before.

Supplementary Table 1 shows the associations between methylation at the identified CpGs for one BMI trait and the other BMI traits. At nominal $P < 0.05$, methylation at all CpGs for current BMI was associated BMI change in the same direction, and *vice versa*. Methylation at approximately 50% CpGs for current BMI was associated with BMI at age 18-21 years in the same direction, and *vice versa*. None of CpGs for BMI change at which methylation was associated with BMI at age 18-21 years, nor *vice versa*. Supplementary Table 2 shows the associations between methylation at the identified CpGs and corresponding BMI trait adjusting for the other BMI traits. At nominal $P < 0.05$,

except for a few CpGs, methylation at all the identified CpGs was associated with corresponding BMI trait independent of the other BMI traits.

Associations with gene-average methylation

The average methylation at *BHLHE40* and *SOCS3* were both negatively associated with current BMI (*BHLHE40*: coefficient[coef]=-1.1×10⁻³, 95% CI: -1.5, -0.7, *P*=3.8×10⁻⁷, FDR=0.01; *SOCS3*: coef=-1.6×10⁻³, 95% CI: -2.3, -1.0, *P*=1.4×10⁻⁶, FDR=0.01), and the average methylation at *PHGDH* was negatively associated with BMI change (coef=-2.2×10⁻³, 95% CI: -3.1, -1.3, *P*=2×10⁻⁶, FDR=0.04). No gene was found to be associated with BMI at age 18-21 years. The genomic inflation factor λ was 1.13, 0.95 and 1.04 for current BMI, BMI at age 18-21 years and BMI change, respectively. (Q-Q plots in Supplementary Figure 5). *SOCS3*¹²⁻¹⁵ and *PHGDH*^{9, 12-15} have been previously reported, and *BHLHE40* has not been reported before. Methylation at seven, five and five CpGs annotated to *BHLHE40*, *SOCS3* and *PHGDH* were associated with their respective BMI trait with nominal *P*<0.05, respectively (Supplementary Figures 6-8).

At nominal *P*<0.05, the average methylation at *BHLHE40* and *SOCS3* was also associated with BMI at age 18-21 years and BMI change. The average methylation at *PHGDH* was also associated with current BMI but not with BMI at age 18-21 years. The average methylation at *BHLHE40* and *SOCS3* was associated with current BMI independent of BMI at age 18-21 years and BMI change, and the average methylation at *PHGDH* was associated with BMI change independent of current BMI (data not shown).

Replication of previously reported associations

Altogether, 151 CpG associations were replicated with nominal *P*<0.05 and were in the same direction as that from previous studies, and the 59 most significant CpGs also had FDR<0.05. Our study reported the first replication for 104 CpGs (Supplementary Table 3). 41 gene associations were replicated with nominal *P*<0.05, and the six most significant genes also had FDR<0.05 (Supplementary Table 4).

Causal inference analyses results

The ICE FALCON results for methylation at CpGs reported by Wahl *et al.*¹⁴ are shown in Table 2. From the analyses with BMI was the predictor and the methylation score the outcome, a woman's methylation score was associated with her own BMI (Model 1; $\beta_{\text{self}} = 0.96$, 95% CI: 0.45, 1.47) and with her co-twin's BMI (Model 2; $\beta_{\text{co-twin}} = 0.57$, 95% CI: 0.06, 1.08). Conditioning on her co-twin's BMI (Model 3), β'_{self} remained unchanged (*P*=0.93), while conditioning on her own BMI (Model 3), $\beta'_{\text{co-twin}}$ attenuated by 95.5% (95CI: 7.0%, 181.2%; *P*=0.03) to be 0.03 (95% CI: -0.50, 0.55). From

the analyses with the methylation score was the predictor and BMI the outcome, a woman's BMI was associated with her own methylation score (Model 1; $\beta_{\text{self}} = 0.78$, 95% CI: 0.14, 1.41) but not with her co-twin's methylation score (Model 2; $\beta_{\text{co-twin}} = -0.02$, 95% CI: -0.33, 0.30). In Model 3, β'_{self} remained unchanged ($P=0.16$) compared with β_{self} in Model 1, and $\beta'_{\text{co-twin}}$ increased by 2500.0% (95%CI: 622.2%, 5044.3%; $P=0.03$) compared with $\beta_{\text{co-twin}}$ in Model 2 to be 0.38 (95% CI: -0.01, 0.77). These results are consistent with the findings by Wahl *et al.* that methylation at these CpGs are consequential to BMI.

Similarly, the results for methylation at CpGs identified in our study are consistent with that current BMI and BMI change have a causal effect on methylation at their related CpGs, respectively (Table 2). The same causation was inferred for gene-average methylation (Supplementary Table 5).

Discussion

We performed EWASs between BMI at two life points and blood DNA methylation, and found that methylation at several loci was associated with BMI at middle age, BMI at age 18-21 years and BMI change. Methylation at some of these loci, *ZPLDI*^{10, 14}, *SOCS3*¹²⁻¹⁶, *CRELD2*^{10, 13, 14}, *NOD2*^{13, 14, 16} and *PHGDH*^{9, 13, 14, 16}, has been previously reported to be associated with BMI. Associations at the other loci, such as *LY9*, *COL6A2*, *ZBBX*, *MAP3K13*, *CRH*, etc., do not appear to have been previously reported. We replicated the associations for methylation at 151 CpGs and 41 genes reported by previous studies, and reported the first replication for 104 of the CpGs. The investigation of causation suggests that BMI and BMI change have a causal effect on methylation, but not *vice versa*.

Several of the identified loci are involved in inflammatory pathways, which are biologically relevant to obesity. *LY9* (lymphocyte antigen 9) is involved in regulation of the development of thymic innate memory-like CD8+ T cell³⁵ and interleukin-17 production³⁶. *MAP3K13* (mitogen-activated protein kinase kinase kinase 13) is involved in activation of mitogen-activated protein kinases (MAPK)³⁷. It has also been found to regulate NF- κ B transcription factor activity³⁸. *NOD2* (nucleotide binding oligomerization domain containing 2) is involved in immune response by playing roles in several biological pathways, such as MAPK activation³⁹, cytokine production⁴⁰⁻⁴² and NF- κ B activation^{43, 44}. *SOCS3* (suppressor of cytokine signalling 3) plays a role in regulation of inflammatory response by negatively regulating cytokine signalling. It is also involved in the Gene Ontology (GO) pathway for negative regulation of insulin receptor signalling (GO:0046627). *CRH* (corticotropin releasing hormone) is a member of the corticotropin-releasing factor family. It is included in the GO pathway for inflammatory response (GO:0006954). Previous studies have also reported methylation at loci involved in inflammatory pathways to be associated with BMI^{9, 10, 12-16}.

Some of the identified loci are related to other pathways. Three loci are involved in oxidation reduction process. *NADK* (NAD kinase) is a NAD⁺ kinase, which catalyses the reduction process from NAD to NADP. *CRYZLI* (crystallin zeta like 1) encodes quinone oxidoreductase and is involved in oxidation reduction process. *PHGDH* (phosphoglycerate dehydrogenase) is a member of oxidoreductase family. Two loci are involved in serine metabolism. *ULK4* (unc-51 like kinase 4) is a serine/threonine kinase. *PHGDH* is involved in the early steps of L-serine synthesis. Associations between methylation at loci related to serine metabolism and BMI have also been reported by previous studies^{9, 13, 16}.

Our study implies that BMI at different life stages influences DNA methylation through different pathways. For example, we found evidence that the loci at which methylation was associated with BMI at age 18-21 years differed from those with current BMI, and especially with BMI change. Though current BMI was associated with methylation at the loci of BMI change and *vice versa*, the associations with current BMI and BMI change are independent of each other, suggesting roles for different pathways.

Our study provides evidence for causation underlying the BMI-methylation association. We replicated that BMI has a causal effect on methylation at the CpGs reported by Wahl *et al.*¹⁴. We found evidence consistent with that the differential methylation at our identified CpGs is secondary to BMI, but not *vice versa*. Findings from previous studies^{8, 13, 45} also support such a causal relationship. Inflammation is proposed to play a role in the development of some obesity-related diseases, and several of the previously reported and our discovered loci are involved in inflammatory pathways, so our findings imply that methylation mediates the effects of obesity on obesity-related diseases. More research on this mediating role of methylation is warranted.

Our study demonstrates how ICE FALCON can be useful for the investigation of causation underlying epigenetic associations. ICE FALCON shares similarities with a bidirectional MR approach. In the scenario represented by the Figure 1b, we essentially use S_X as an instrumental variable for X_{self} . However, S_X is not measured, so we use $X_{\text{co-twin}}$ as a surrogate instrumental variable (Similarly, $Y_{\text{co-twin}}$ is used as a surrogate for S_Y , the instrumental variable for Y_{self} , when the roles of Y and X are swapped). S_X includes all causes of familial correlation in X that are specific to X, which is stronger than a finite number of genetic variants that are assumed to be specific to X, thus the results from ICE FALCON are less biased by the strength of the instrumental variable. S_X , by definition, is not associated with confounders of the relationship between X and Y within an individual and has no directional pleiotropy (any association or directional pleiotropy is captured in S_{XY}). Even though the surrogate instrumental variable $X_{\text{co-twin}}$ is associated with the confounders or has directional pleiotropy, ICE FALCON still works. For example, in a scenario where the association between $X_{\text{co-twin}}$ and Y_{self} is mediated not only through S_X and X_{self} but also through other unmeasured factors such as familial

confounders, i.e., a mixture of Figures 1a and 1b, even though $X_{\text{co-twin}}$ will still be associated with Y_{self} after adjusting for X_{self} , $\beta_{\text{co-twin}}$ is still expected to attenuate towards the null, given that adjusting for X_{self} blocks the pathway $X_{\text{co-twin}} \leftarrow S_X \rightarrow X_{\text{self}} \rightarrow Y_{\text{self}}$. ICE FALCON can be an alternative method for investigating causation when genetic variants data are unavailable, or the assumption underlying MR analyses are suspect. ICE FALCON can also be used as an independent method to replicate the findings from MR analyses, and *vice versa*. However, requirements such as the need for substantial within-pair correlations may limit the application of ICE FALCON, which was the case for BMI at age 18-21 years in our study.

Our gene-average methylation approach is of potential utility for analysing EWAS. Such utility is supported to some extent by our replication of previously reported genes including *SOCS3* and *PHGDH*. A CpG-by-CpG approach is typically used in EWAS. However, this approach requires a stringent genome-wide significance level, e.g., 10^{-7} when half a million of CpGs are analysed and the Bonferroni adjustment is applied, which may not be achievable. The gene-based approach aimed to address this challenge by reducing the number of tests to 20 000. This approach can potentially find novel associations missed by the CpG-by-CpG approach, such as the association for *BHLHE40*.

Our study has several strengths. We studied three BMI traits at different life points. We used a novel family-based analytical approach to investigate the causation underlying the cross-sectional association and reached the same conclusion as the MR approach. We investigated gene-average methylation, which can potentially provide evidence for novel associations.

One limitation of our study is that BMI traits were self-reported, thus are subject to bias. Another limitation is that we studied methylation in peripheral blood, not in tissues more relevant to adiposity such as adipose tissue. However, cross-tissue investigations suggest moderate to good overall consistency between blood and adiposity-relevant tissues in the context of BMI-methylation association. In the study by Demerath *et al.*, associations for 64% of CpGs identified in blood were replicated in adipose tissue¹⁰. In the study by Wahl *et al.*, blood methylation levels at BMI-related CpGs had moderate to high correlations with methylation levels in adiposity-relevant tissues¹⁴. More importantly, given that peripheral blood samples are easily accessible, any association with BMI detectable in this sample is potentially valuable for large-scale public health translation on population level.

In conclusion, we found evidence that in peripheral blood from middle aged women, DNA methylation at several loci is associated with current BMI, BMI at age 18-21 years and BMI change between the two ages. Additionally, our study found evidence consistent with BMI having a causal effect on peripheral blood DNA methylation.

Acknowledgements

We would like to thank all women participating in this study. The data analysis was facilitated by Spartan, the High Performance Computer and Cloud hybrid system of the University of Melbourne.

The AMDTSS was facilitated through the Australian Twin Registry, a national research resource in part supported by a Centre for Research Excellence Grant from the National Health and Medical Research Council (NHMRC) APP 1079102. The AMDTSS was supported by NHMRC (grant numbers 1050561 and 1079102), Cancer Australia and National Breast Cancer Foundation (grant number 509307).

SL is supported by the Australian Government Research Training Program Scholarship from the University of Melbourne. TLN is supported by a NHMRC Post-Graduate Scholarship and the Richard Lovell Travelling Scholarship from the University of Melbourne. MCS is a NHMRC Senior Research Fellow. JLH is a NHMRC Senior Principal Research Fellow.

Conflict of Interest

The authors declare no conflicts of interest.

Figure legend

Figure 1 Some possible directed acyclic graphs for the causation underlying the cross-trait cross-twin correlation

References

1. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature* 2010; **465**(7299): 721-7.
2. Esteller M. Epigenetics in cancer. *N Engl J Med* 2008; **358**(11): 1148-59.
3. Wang X, Zhu H, Snieder H, Su S, Munn D, Harshfield G *et al.* Obesity related methylation changes in DNA of peripheral blood leukocytes. *BMC Med* 2010; **8**: 87.
4. Kuehnen P, Mischke M, Wiegand S, Sers C, Horsthemke B, Lau S *et al.* An Alu element-associated hypermethylation variant of the POMC gene is associated with childhood obesity. *PLoS Genet* 2012; **8**(3): e1002543.
5. Ding X, Zheng D, Fan C, Liu Z, Dong H, Lu Y *et al.* Genome-wide screen of DNA methylation identifies novel markers in childhood obesity. *Gene* 2015; **566**(1): 74-83.
6. Almen MS, Nilsson EK, Jacobsson JA, Kalnina I, Klovins J, Fredriksson R *et al.* Genome-wide analysis reveals DNA methylation markers that vary with both age and obesity. *Gene* 2014; **548**(1): 61-7.
7. Su S, Zhu H, Xu X, Wang X, Dong Y, Kapuku G *et al.* DNA methylation of the LY86 gene is associated with obesity, insulin resistance, and inflammation. *Twin Res Hum Genet* 2014; **17**(3): 183-91.
8. Dick KJ, Nelson CP, Tsaprouni L, Sandling JK, Aissi D, Wahl S *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* 2014; **383**(9933): 1990-8.
9. Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L *et al.* Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity (Silver Spring)* 2015; **23**(7): 1493-501.
10. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet* 2015; **24**(15): 4464-79.

11. Al Muftah WA, Al-Shafai M, Zaghlool SB, Visconti A, Tsai PC, Kumar P *et al.* Epigenetic associations of type 2 diabetes and BMI in an Arab population. *Clin Epigenetics* 2016; **8**: 13.
12. Ali O, Cerjak D, Kent JW, Jr., James R, Blangero J, Carless MA *et al.* Methylation of SOCS3 is inversely associated with metabolic syndrome in an epigenome-wide association study of obesity. *Epigenetics* 2016; **11**(9): 699-707.
13. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman AK, 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. *PLoS Med* 2017; **14**(1): e1002215.
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* 2017; **541**(7635): 81-86.
15. Wilson LE, Harlid S, Xu Z, Sandler DP, Taylor JA. An epigenome-wide study of body mass index and DNA methylation in blood using participants from the Sister Study cohort. *Int J Obes (Lond)* 2017; **41**(1): 194-199.
16. Geurts Y, Dugue PA, Joo EJ, Makalic E, Jung CH, Guan W *et al.* Novel associations between blood DNA methylation and body mass index in middle-aged and older adults. *Internal Journal of Obesity (in press)*.
17. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V *et al.* Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol* 2015; **3**(7): 526-534.
18. Li S, Wong EM, Southey MC, Hopper JL. Association between DNA methylation at SOCS3 gene and body mass index might be due to familial confounding. *Int J Obes (Lond)* 2017; **41**(6): 995-996.
19. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 2012; **41**(1): 161-76.

20. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. *Epidemiology* 2014; **25**(3): 427-35.
21. Odefrey F, Stone J, Gurrin LC, Byrnes GB, Apicella C, Dite GS *et al.* Common genetic variants associated with breast cancer and mammographic density measures that predict disease. *Cancer Res* 2010; **70**(4): 1449-58.
22. Li S, Wong EM, Joo JE, Jung CH, Chung J, Apicella C *et al.* Genetic and Environmental Causes of Variation in the Difference Between Biological Age Based on DNA Methylation and Chronological Age for Middle-Aged Women. *Twin Res Hum Genet* 2015; **18**(6): 720-6.
23. Joo JE, Wong EM, Baglietto L, Jung CH, Tsimiklis H, Park DJ *et al.* The use of DNA from archival dried blood spots with the Infinium HumanMethylation450 array. *BMC biotechnology* 2013; **13**(1): 23.
24. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014; **30**(10): 1363-9.
25. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. *Genome Biology* 2012; **13**(6): R44.
26. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**(1): 118-127.
27. Price ME, Cotton AM, Lam LL, Farre P, Emberly E, Brown CJ *et al.* Additional annotation enhances potential for biologically-relevant analysis of the Illumina Infinium HumanMethylation450 BeadChip array. *Epigenetics Chromatin* 2013; **6**(1): 4.
28. 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.
29. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* 1995; **57**(1): 289-300.

30. Hopper JL, Bui QM, Erbas B, Matheson MC, Gurrin LC, Burgess JA *et al.* Does eczema in infancy cause hay fever, asthma, or both in childhood? Insights from a novel regression model of sibling data. *J Allergy Clin Immunol* 2012; **130**(5): 1117-1122 e1.
31. Stone J, Dite GS, Giles GG, Cawson J, English DR, Hopper JL. Inference about causation from examination of familial confounding: application to longitudinal twin data on mammographic density measures that predict breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2012; **21**(7): 1149-55.
32. Bui M, Bjornerem A, Ghasem-Zadeh A, Dite GS, Hopper JL, Seeman E. Architecture of cortical bone determines in part its remodelling and structural decay. *Bone* 2013; **55**(2): 353-8.
33. Dite GS, Gurrin LC, Byrnes GB, Stone J, Gunasekara A, McCredie MR *et al.* Predictors of mammographic density: insights gained from a novel regression analysis of a twin study. *Cancer Epidemiol Biomarkers Prev* 2008; **17**(12): 3474-81.
34. Davey CG, Lopez-Sola C, Bui M, Hopper JL, Pantelis C, Fontenelle LF *et al.* The effects of stress-tension on depression and anxiety symptoms: evidence from a novel twin modelling analysis. *Psychol Med* 2016; **46**(15): 3213-3218.
35. Sintes J, Cuenca M, Romero X, Bastos R, Terhorst C, Angulo A *et al.* Cutting edge: Ly9 (CD229), a SLAM family receptor, negatively regulates the development of thymic innate memory-like CD8⁺ T and invariant NKT cells. *J Immunol* 2013; **190**(1): 21-6.
36. Chatterjee M, Rauen T, Kis-Toth K, Kyttaris VC, Hedrich CM, Terhorst C *et al.* Increased expression of SLAM receptors SLAMF3 and SLAMF6 in systemic lupus erythematosus T lymphocytes promotes Th17 differentiation. *J Immunol* 2012; **188**(3): 1206-12.
37. Ikeda A, Hasegawa K, Masaki M, Moriguchi T, Nishida E, Kozutsumi Y *et al.* Mixed lineage kinase LZK forms a functional signaling complex with JIP-1, a scaffold protein of the c-Jun NH(2)-terminal kinase pathway. *J Biochem* 2001; **130**(6): 773-81.
38. Masaki M, Ikeda A, Shiraki E, Oka S, Kawasaki T. Mixed lineage kinase LZK and antioxidant protein-1 activate NF-kappaB synergistically. *Eur J Biochem* 2003; **270**(1): 76-83.

39. Hsu YM, Zhang Y, You Y, Wang D, Li H, Duramad O *et al.* The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens. *Nat Immunol* 2007; **8**(2): 198-205.
40. Netea MG, Azam T, Ferwerda G, Girardin SE, Walsh M, Park JS *et al.* IL-32 synergizes with nucleotide oligomerization domain (NOD) 1 and NOD2 ligands for IL-1beta and IL-6 production through a caspase 1-dependent mechanism. *Proc Natl Acad Sci U S A* 2005; **102**(45): 16309-14.
41. Davey MP, Martin TM, Planck SR, Lee J, Zamora D, Rosenbaum JT. Human endothelial cells express NOD2/CARD15 and increase IL-6 secretion in response to muramyl dipeptide. *Microvasc Res* 2006; **71**(2): 103-7.
42. van Beelen AJ, Zelinkova Z, Taanman-Kueter EW, Muller FJ, Hommes DW, Zaat SA *et al.* Stimulation of the intracellular bacterial sensor NOD2 programs dendritic cells to promote interleukin-17 production in human memory T cells. *Immunity* 2007; **27**(4): 660-9.
43. McDonald C, Chen FF, Ollendorff V, Ogura Y, Marchetto S, Lecine P *et al.* A role for Erbin in the regulation of Nod2-dependent NF-kappaB signaling. *J Biol Chem* 2005; **280**(48): 40301-9.
44. Zhao Y, Alonso C, Ballester I, Song JH, Chang SY, Guleng B *et al.* Control of NOD2 and Rip2-dependent innate immune activation by GEF-H1. *Inflamm Bowel Dis* 2012; **18**(4): 603-12.
45. Richmond RC, Sharp GC, Ward ME, Fraser A, Lyttleton O, McArdle WL *et al.* DNA Methylation and BMI: Investigating Identified Methylation Sites at HIF3A in a Causal Framework. *Diabetes* 2016; **65**(5): 1231-44.

Table 1 CpGs at which methylation was found to be associated with the three BMI traits

| BMI trait | CpG | CHR | Coordinate | Gene | Coef | SE | P | FDR |
|------------------------|------------|-----|------------|--------------|-------|-----|---------|------|
| Current BMI | | | | | | | | |
| | cg12992827 | 3 | 101901234 | ZPLD1 | -12.6 | 2.4 | 1.4E-07 | 0.03 |
| | cg18865207 | 1 | 160765919 | LY9 | 5.8 | 1.1 | 2.8E-07 | 0.03 |
| | cg11529924 | 21 | 47545200 | COL6A2 | -7.4 | 1.5 | 3.1E-07 | 0.03 |
| | cg10542883 | 3 | 185000726 | MAP3K13 | -20.7 | 4.1 | 4.3E-07 | 0.03 |
| | cg24197445 | 3 | 167045873 | ZBBX | -10.4 | 2.1 | 4.4E-07 | 0.03 |
| | cg00636368 | 3 | 185000586 | MAP3K13 | -26.2 | 5.2 | 5.2E-07 | 0.03 |
| | cg18181703 | 17 | 76354621 | SOCS3 | -9.3 | 1.9 | 6.1E-07 | 0.03 |
| | cg06181567 | 16 | 50731115 | NOD2 | 9.0 | 1.8 | 6.2E-07 | 0.03 |
| | cg18640030 | 8 | 67090752 | CRH | -21.9 | 4.4 | 7.5E-07 | 0.03 |
| BMI at age 18-21 years | | | | | | | | |
| | cg14990430 | 22 | 50277562 | ZBED4 | 33.1 | 5.9 | 2.0E-08 | 0.01 |
| | cg06045576 | 22 | 19948957 | COMT | 13.2 | 2.4 | 5.5E-08 | 0.01 |
| | cg00819078 | 5 | 16615081 | FAM134B | 14.1 | 2.7 | 2.2E-07 | 0.03 |
| | cg03357727 | 3 | 41360280 | ULK4 | -19.1 | 3.8 | 3.7E-07 | 0.04 |
| | cg19309463 | 8 | 145806752 | KIAA1688 | 23.6 | 4.7 | 5.2E-07 | 0.04 |
| | cg09446247 | 7 | 151102752 | WDR86 | 11.6 | 2.3 | 7.0E-07 | 0.05 |
| BMI change | | | | | | | | |
| | cg12992827 | 3 | 101901234 | ZPLD1 | -15.1 | 2.4 | 5.1E-10 | 0.00 |
| | cg25498514 | 11 | 3190676 | OSBPL5 | -22.6 | 4.0 | 2.5E-08 | 0.01 |
| | cg01776926 | 12 | 116560359 | MED13L | -24.1 | 4.5 | 8.4E-08 | 0.01 |
| | cg26342907 | 1 | 1687532 | NADK | -15.7 | 3.1 | 5.3E-07 | 0.03 |
| | cg12644285 | 15 | 93570953 | CHD2 | -10.2 | 2.0 | 6.1E-07 | 0.03 |
| | cg07698196 | 7 | 150485638 | LOC100128542 | -10.4 | 2.1 | 7.1E-07 | 0.03 |
| | cg11336535 | 1 | 230472614 | PGBD5 | 9.6 | 1.9 | 7.4E-07 | 0.03 |
| | cg20432448 | 21 | 35014337 | CRYZL1 | -14.9 | 3.0 | 7.5E-07 | 0.03 |
| | cg06983052 | 1 | 90288099 | LRRC8D | -8.1 | 1.6 | 7.6E-07 | 0.03 |
| | cg09349128 | 22 | 50327986 | CRELD2 | -8.5 | 1.7 | 7.8E-07 | 0.03 |
| | cg01593552 | 13 | 22185680 | EFHA1 | 4.5 | 0.9 | 9.9E-07 | 0.03 |
| | cg00636368 | 3 | 185000586 | MAP3K13 | -28.3 | 5.8 | 9.9E-07 | 0.03 |

Coefficient was reported as coefficients $\times 10^3$, so as well for standard error. BMI, body mass index; CpG, cytosine-guanine dinucleotide; CHR, chromosome; Coef, coefficient; SE, standard error; FDR, false discovery rate.

Table 2 Results from the ICE FALCON analyses for CpG-specific methylation*

| CpG | Predictor | Coefficient | Model 1 | | Model 2 | | Model 3 | | Change | |
|-------------------------------------|-------------------|----------------------|-------------|----------|--------------|----------|--------------|----------|--------------|----------|
| | | | Est (SE) | <i>P</i> | Est (SE) | <i>P</i> | Est (SE) | <i>P</i> | Est (SE) | <i>P</i> |
| CpGs reported by Wahl <i>et al.</i> | | | | | | | | | | |
| Current BMI CpGs | BMI | β _{self} | 0.96 (0.26) | 2.3E-04 | | | 0.95 (0.31) | 2.3E-03 | -0.01 (0.16) | 9.3E-01 |
| | | β _{co-twin} | | | 0.57 (0.26) | 2.8E-02 | 0.03 (0.27) | 9.2E-01 | -0.54 (0.26) | 3.3E-02 |
| | Methylation score | β _{self} | 0.78 (0.32) | 1.6E-02 | | | 1.00 (0.32) | 1.6E-03 | 0.22 (0.16) | 1.6E-01 |
| | | β _{co-twin} | | | -0.02 (0.16) | 9.2E-01 | 0.38 (0.20) | 5.4E-02 | 0.40 (0.18) | 3.0E-02 |
| | BMI | β _{self} | 0.40 (0.06) | 7.5E-13 | | | 0.45 (0.07) | 2.3E-09 | 0.05 (0.05) | 3.6E-01 |
| | | β _{co-twin} | | | 0.28 (0.07) | 9.4E-05 | -0.07 (0.07) | 3.3E-01 | -0.35 (0.08) | 3.6E-05 |
| | Methylation score | β _{self} | 4.57 (1.10) | 3.3E-05 | | | 5.91 (1.12) | 1.2E-07 | 1.33 (0.67) | 4.6E-02 |
| | | β _{co-twin} | | | -0.79 (0.78) | 3.1E-01 | 2.20 (0.78) | 5.0E-03 | 3.00 (0.98) | 2.2E-03 |
| BMI change CpGs | | | | | | | | | | |
| | BMI change | β _{self} | 0.57 (0.07) | 2.0E-16 | | | 0.49 (0.09) | 1.2E-07 | -0.08 (0.08) | 3.2E-01 |
| | | β _{co-twin} | | | 0.47 (0.09) | 6.3E-08 | 0.12 (0.10) | 2.0E-01 | -0.35 (0.09) | 3.9E-05 |
| | Methylation score | β _{self} | 3.83 (0.78) | 8.5E-07 | | | 4.25 (0.70) | 1.5E-09 | 0.43 (0.44) | 3.3E-01 |
| | | β _{co-twin} | | | 0.26 (0.61) | 6.7E-01 | 1.41 (0.62) | 2.3E-02 | 1.15 (0.56) | 3.9E-02 |

ICE FALCON, inference about causation through examination of familial confounding; CpG, cytosine-guanine dinucleotide; BMI, body mass index; Est, estimate; SE, standard error.

*Regression coefficient from the analyses in which methylation score as the predictor was reported as coefficient $\times 10$, so as well for standard error.

