Blood-based genome-wide DNA methylation correlations across body-fat- and adiposity-related biochemical traits

Authors

Alesha A. Hatton, Robert F. Hillary, Elena Bernabeu, Daniel L. McCartney, Riccardo E. Marioni, Allan F. McRae

Correspondence

a.mcrae@uq.edu.au

DNA methylation (DNAm) correlations evaluate the similarities in DNAm associations between traits, capturing both shared cause and consequence. DNAm correlations for six body-fat-related traits were of greater magnitude than genetic and phenotypic correlations and identified differences in associations between BMI and DNAm in males and females.

Hatton et al., 2023, The American Journal of Human Genetics *110*, 1–10
September 7, 2023
https://doi.org/10.1016/j.ajhg.2023.08.004

CellPress

Please cite this article in press as: Hatton et al., Blood-based genome-wide DNA methylation correlations across body-fat- and adiposity-related biochemical traits, The American Journal of Human Genetics (2023), https://doi.org/10.1016/j.ajhg.2023.08.004

ARTICLE

Blood-based genome-wide DNA methylation correlations across body-fat- and adiposity-related biochemical traits

Alesha A. Hatton,[1] Robert F. Hillary,[2] Elena Bernabeu,[2] Daniel L. McCartney,[2] Riccardo E. Marioni,[2] and Allan F. McRae[1,*]

Summary

The recent increase in obesity levels across many countries is likely to be driven by nongenetic factors. The epigenetic modification DNA methylation (DNAm) may help to explore this, as it is sensitive to both genetic and environmental exposures. While the relationship between DNAm and body-fat traits has been extensively studied, there is limited literature on the shared associations of DNAm variation across such traits. Akin to genetic correlation estimates, here, we introduce an approach to evaluate the similarities in DNAm associations between traits: DNAm correlations. As DNAm can be both a cause and consequence of complex traits, DNAm correlations have the potential to provide insights into trait relationships above that currently obtained from genetic and phenotypic correlations. Utilizing 7,519 unrelated individuals from Generation Scotland with DNAm from the EPIC array, we calculated DNAm correlations between body-fat- and adiposity-related traits by using the bivariate OREML framework in the OSCA software. For each trait, we also estimated the shared contribution of DNAm between sexes. We identified strong, positive DNAm correlations between each of the body-fat traits (BMI, body-fat percentage, and waist-to-hip ratio, ranging from 0.96 to 1.00), finding larger associations than those identified by genetic and phenotypic correlations. We identified a significant deviation from 1 in the DNAm correlations for BMI between males and females, with sex-specific DNAm changes associated with BMI identified at eight DNAm probes. Employing genome-wide DNAm correlations to evaluate the similarities in the associations of DNAm with complex traits has provided insight into obesity-related traits beyond that provided by genetic correlations.

Introduction

Obesity constitutes a growing healthcare burden and is a major risk factor for several chronic diseases including cardiovascular diseases and diabetes.[1,2] Body mass index (BMI), the most widely used measure of obesity, results from the complex interplay between genetic, environmental, and modifiable lifestyle factors. The increase in BMI levels in recent years[3] is likely driven by nongenetic factors. DNA methylation (DNAm) is a commonly studied epigenetic modification that is responsive to both genetics and the environment, making it an ideal target for studying the consequences of modifiable health factors, such as obesity. The relationship between DNAm and BMI, as well as other body-fat- and adiposity-related biochemical traits, has been extensively studied.[4–12] However, the shared associations of DNAm variation across such traits represents an important gap in our understanding of the biological processes pertaining to obesity.

Akin to genetic correlation estimates, which measure the degree of common genetic control between two traits, here, we introduce an approach to evaluate the similarities in DNAm associations between traits, DNAm correlations. In contrast with genetic variants, DNAm is responsive to a wide range of environmental exposures and may reflect the cumulative burden of adverse exposures throughout the life course. In addition, variation in DNAm has been implicated as arising from individual differences in traits such as BMI and smoking,[8,13] with some evidence suggesting BMI in childhood may be predictive of adolescent DNAm levels at sites throughout the genome.[4] Thus, while genetic correlations capture causal effects on the traits, DNAm correlations will capture consequence, too. Ascertaining effects from both directions may result in the detection of additional biological mechanisms underlying the relationship between these traits. We also recognize that with a large portion of the DNA methylome under genetic control,[14] DNAm correlations will likely capture part of the shared genetic contribution between these traits. However, recent work[7,8,13,15] has demonstrated that DNAm associated with BMI trait variance is independent of genetic variation. This indicates that DNAm correlations have the potential to provide insights into trait relationships as well as the molecular underpinnings and subsequent consequences of the traits beyond that currently obtained from genetic correlations.

We estimate DNAm correlations for six body-fat- and adiposity-related biochemical traits for 7,519 unrelated individuals from Generation Scotland (GS): Scottish Family Health Study. DNAm correlations are estimated by extending the omics restricted maximum likelihood (OREML) method in the omics-data-based complex trait analysis

[1]Institute for Molecular Bioscience, Brisbane, Australia; [2]Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
*Correspondence: a.mcrae@uq.edu.au
https://doi.org/10.1016/j.ajhg.2023.08.004.

(OSCA) software[16] to a bivariate model, akin to bivariate genome-based restricted maximum likelihood (GREML) as implemented in the genome-wide complex trait analysis (GCTA) software.[17,18] These DNAm correlation estimates provide a measure of the shared similarity of DNAm variation between phenotypes, noting that while SNPs explain the variation in traits, DNAm only captures this variation and reflects both cause and consequence. We compared these DNAm correlations with genetic and phenotypic correlations to investigate whether they provide insights into the shared genomic control of traits.

Material and methods

Study cohort

All data for the study came from GS. The family-based genetic epidemiological cohort consists of over 24,000 volunteers, which has previously been described.[19,20] Recruitment took place between 2006 and 2011, when individuals and their family members aged 18+ years were invited to a baseline clinic visit that included health questionnaires and blood or saliva sample donation for genomic analyses. All DNAm samples were measured in blood. This study uses phenotypic, DNAm, and genetic data from unrelated samples (n = 7,519, genetic relationship matrix [GRM] < 0.05), with DNAm levels quantified in three sets based on time of DNAm array processing.

Ethics approval

All components of GS received ethical approval from the NHS Tayside Committee on Medical Research Ethics (REC reference number: 05/S1401/89). GS has also been granted Research Tissue Bank status by the East of Scotland Research Ethics Service (REC reference number: 20-ES-0021), providing generic ethical approval for a wide range of uses within medical research.

Phenotypic data

Three anthropometric measurements and three biochemical phenotypes were investigated: BMI (kg/m^2), body-fat percentage (%), waist-to-hip ratio (WHR), glucose (mmol/L), high-density lipoprotein (HDL) cholesterol (mmol/L), and total cholesterol (mmol/L). All phenotypes were trimmed for outliers (values that were ±4 SDs from the mean). In addition, BMI was trimmed for extreme values at <17 and >50 kg/m^2. For each trait, we stratified the samples by sex and then adjusted the phenotype for age and standardized the residuals by rank-based inverse normal transformation before recombining the data. There was no adjustment for set as there were minimal differences in the mean across sets. Residualized phenotypes were entered as dependent variables in the subsequent analysis. Smoking pack years were calculated by multiplying the number of packs of cigarettes smoked per day by the number of years the individual has smoked and used in the adjustment of DNAm data.

Genetic data

Genome-wide genotypic details have been previously described.[21] Briefly, GS participants were genotyped with either Illumina HumanOmniExpressExome8v1-2_A or HumanOmniExpressExome-8v1_A arrays. SNPs were excluded for missing genotype call rate (>2%), marked departure from Hardy-Weinberg equilibrium (HWE; $p < 1 \times 10^{-6}$), and low minor-allele frequency (MAF; <1%). Duplicate samples were removed alongside individuals with gender mismatch and missing genotype call rate (>2%). Genotype data were imputed against the Haplotype Reference Consortium (HRC) panel version 1.1.[22] Unrelated individuals were retained (GRM < 0.05) using the GCTA software.[18] All subsequent analyses were conducted on the unrelated individuals using HapMap3 SNPs only.

DNAm data

Genome-wide blood-based DNAm was profiled using the Illumina Methylation EPIC array and processed in three separate sets. DNAm quality control was performed as previously reported.[23] Briefly, outliers were excluded based on the visual inspection of methylated-to-unmethylated log intensities in addition to poorly performing probes and samples and sex mismatches. Further filtering was performed to exclude non-autosomal CpG sites, CpGs that were predicted to cross-hybridize, and those with polymorphisms at the target site that can alter probe binding.[24,25] Poor performing probes, X/Y chromosome probes, and participants with unreliable self-report data (self-reported "yes" for all diseases in the questionnaire), saliva samples (with no blood sample provided), and potential XXY genotype were excluded along with probes with almost invariable beta values across individuals (SD < 0.02). All 3 sets were normalized together with the final discovery dataset comprised of M values at 781,379 loci for 7,519 participants. Before analysis, DNAm was adjusted in the OSCA software for age, sex, batch (to correct for samples that were processed at the same time within each set), slide (to correct for samples included on the same array), cell-type proportions (estimated using the algorithm proposed by Houseman et al.[26]), smoking status, and pack years.

Variance component analyses

Utilizing 7,519 unrelated individuals from GS, we estimate the proportion of phenotypic variance captured by genome-wide DNAm across six body-fat- and adiposity-related biochemical traits using OREML framework in the OSCA software. This method fits all DNAm probes jointly in a random-effects model, meaning that each DNAm probe is fitted conditional on the joint effects of all other probes. This is performed through the construction of a DNAm relationship matrix (MRM) based on all DNAm probes and that is used to model the covariance between individuals. In doing so, the model is able to account for the lack of independence between DNAm probes, similar to how GREML accounts for linkage disequilibrium (LD) between SNPs.[27] This allows us to estimate the proportion of variation for each trait captured by all probes via restricted maximum likelihood, which is analogous to that of estimating SNP-based heritability based on genetic data.[18,27] Unlike SNP-based heritability, we note that the proportion of variance captured by all probes may be capturing both cause and consequence of the phenotype. The GCTA software[18] was used to calculate the GRM and similarly implemented in the OSCA software to estimate SNP-based heritability, referred to here as the proportion of phenotypic variance explained by all SNPs. We also estimated these quantities jointly in the OSCA software by using –multi-orm, which allows for multiple random effects.

DNAm Correlations

We estimate the DNAm correlation between phenotypes implemented using the bivariate OREML framework in OSCA utilizing DNAm relationship matrices rather than the standard GRM, where

Table 1. Cohort summary for Generation Scotland

Covariates	N	Mean	SD
Age	7,519	51.7	13.2
Sex	–	–	–
Female	4,261 (56.7%)	–	–
Male	3,258 (43.3%)	–	–
Set	–	–	–
Set 1	1,988 (26.4%)	–	–
Set 2	4,228 (56.2%)	–	–
Set 3	1,303 (17.3%)	–	–
Smoking status	–	–	–
Current smokers	1,192 (15.9%)	–	–
Quit <12 months ago	164 (2.2%)	–	–
Quit >12 months ago	2,244 (29.8%)	–	–
Never smokers	3,871 (51.5%)	–	–
Unknown	48 (0.6%)	–	–
Traits	–	–	–
BMI	7,452	26.9	5.0
Body fat %	7,324	30.4	9.3
Waist-to-hip ratio	7,403	0.9	0.1
Glucose	7,291	4.8	0.6
HDL cholesterol	7,438	1.5	0.4
Total cholesterol	7,455	5.2	1.1
Pack years	7,519	8.1	15.1
Cell-type proportions	–	–	–
CD8T	7,519	0.04	0.04
CD4T	7,519	0.18	0.06
NK	7,519	0.07	0.05
Bcell	7,519	0.06	0.03
Mono	7,519	0.06	0.02
Gran	7,519	0.58	0.09

Smoking pack years were calculated by multiplying the number of packs of cigarettes smoked per day by the number of years the individual has smoked and used in the adjustment of DNAm data. Cell type proportions estimated using the algorithm proposed by Houseman et al.[26]

the DNAm correlation is estimated from the one of the covariance components. Here, the phenotypic and DNAm information came from the same unrelated individuals. This approach estimates the shared contribution of DNAm based on the MRM between phenotypes. Likelihood ratio tests were performed to test the hypotheses of fixing the correlations at both zero and one. We additionally estimated genetic correlations using GCTA and phenotypic correlations using Pearson's correlation and compared these with DNAm correlations to investigate whether this metric provides insights into the molecular underpinnings of these traits. Joint estimation of genetic and DNAm correlations were not reported as the models were unable to converge.

Results

GS is a Scottish family-based study with over 24,000 participants recruited between 2006 and 2011.[19,20] We analyzed data from 7,519 unrelated individuals (GRM pruned at 0.05) from the larger GS dataset to avoid confounding between genetic relatedness and epigenetic similarity. Blood-based DNAm levels at 781,379 DNAm sites were quantified using the Illumina Methylation EPIC array in three sets based on time of generation of DNAm array processing. Three anthropometric measurements and three biochemical phenotypes were investigated: BMI (kg/m^2), body-fat percentage (%), WHR, glucose (mmol/L), HDL (mmol/L), and total cholesterol (mmol/L). Demographic and summary information from GS for the six phenotypes are presented in Table 1. We estimated the proportion of phenotypic variation captured by DNAm for each trait based on an MRM using OSCA,[16] the variation explained by SNPs based on a GRM, as well as that captured jointly by DNAm and SNPs. As demonstrated previously,[13] we observed non-zero estimates for the proportion of variance captured by DNAm when estimated jointly with SNPs, which demonstrates that some of the variation captured by DNAm is in addition to that being captured by SNPs (see supplemental results; Figure S1 and Table S1). The additional variation captured by DNAm indicates that there is a potential to gain insights into trait relationships with DNAm correlations that are not currently captured by genetic correlations based on common SNPs.

We extended the OREML approach of the OSCA software to a bivariate model that simultaneously estimates the proportion of variance in the two traits captured by DNAm as well as quantifies the shared associations between DNAm and the two traits. We term this shared association as a DNAm correlation, reflecting the similarity of the approach to estimating genetic correlations via the GREML model.

DNAm correlations between sets

As a proof-of-principle illustration of the application of genetic correlation methods to DNAm, and to demonstrate strong concordance between the three sample sets with GS, we estimated the DNAm correlation of the six traits across the sample set. The underlying assumption is that there should be no inter-set variation in contribution of DNAm to each of the traits, and thus the DNAm correlation estimates should not be different from 1. We first test this assumption by performing an epigenome-wide association study (EWAS) within each set for each trait to determine whether there is concordance in probe effects by using simple linear regression in the OSCA software.[16] Due to differences in sample size between the sets, which impacts discovery, only those probes that were nominally significant across all sets ($p < 0.001$) were compared. The concordance in probe association coefficients was evaluated using Pearson's correlation and was found to be very high ($\rho \geq 0.95$)

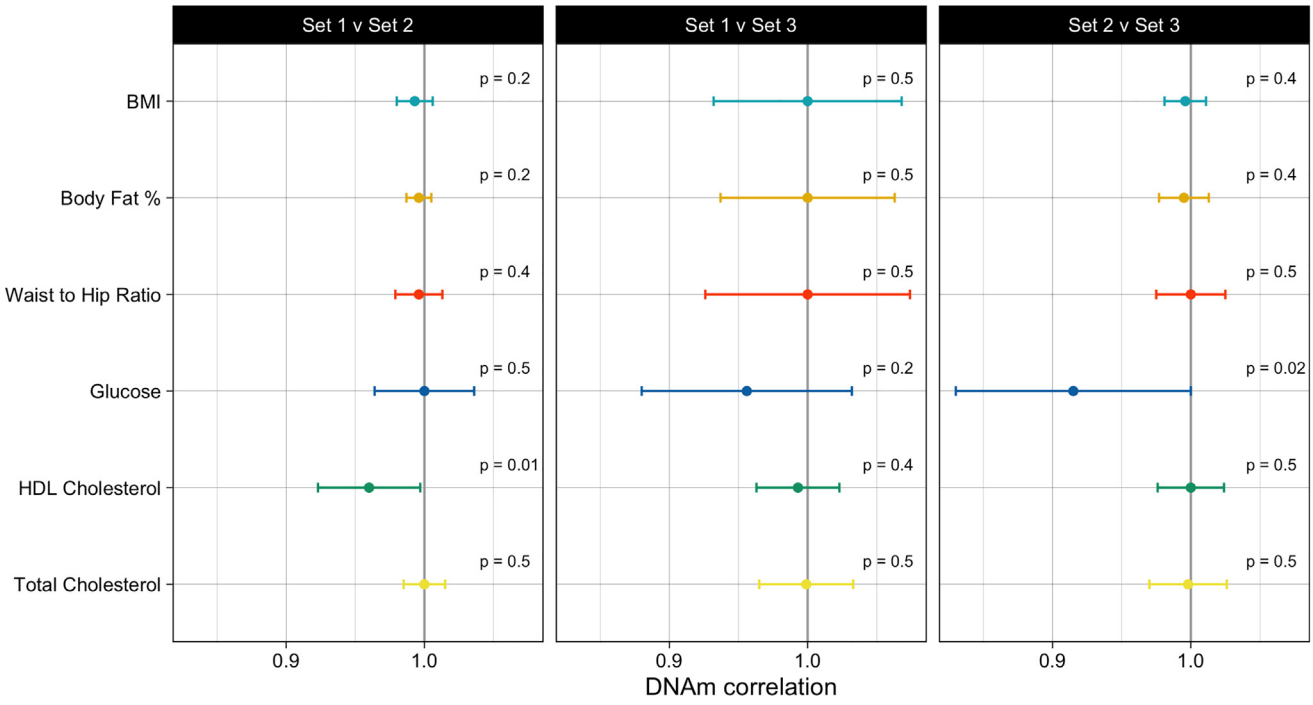

Figure 1. DNAm correlation between sets for each trait
The DNAm correlations between each of the set pairs is displayed on the x axis with standard errors indicated by error bars. p values from a log likelihood test for the hypothesis of fixing the DNAm correlation at 1 are presented in text below each estimate and in Table S2.

between all sets (Figure S2). This suggests that the estimated effect sizes between DNAm and each of the traits is consistent between sets. We subsequently calculated the DNAm correlation between sets for each trait. Most of the correlations were found to not significantly deviate from 1 ($p > 0.05$), consistent with our expectation (Figure 1; Table S2). We note that the slight deviations from 1 observed for glucose and HDL cholesterol, as well as large standard errors, while not significant after adjusting for multiple testing using a Bonferroni correction, may reflect deviations from sample collection protocols and measurement errors rather than a reflection of the method.

DNAm correlations between traits

We estimated the DNAm correlation between the six body-fat-related phenotypes using the bivariate OREML framework that estimates the similarity of DNAm associations between traits. The DNAm correlations are presented in Figure 2 (Table S3) alongside genetic correlations calculated using the bivariate GREML framework with a GRM implemented in the GCTA software[18] and phenotypic correlations. We identified strong, positive DNAm correlations between each of the body-fat traits (BMI, body-fat percentage, and WHR; ranging from 0.96 to 1.00), with correlation between BMI and WHR found to be not significant different from 1 ($r_{DNAm} = 1.00$, SE = 0.0005). These associations were observed to be of greater magnitude than both genetic (r_G ranging from 0.65 to 0.86) and phenotypic correlations (r_P ranging from 0.51 to 0.85). The body-fat traits demonstrated moderate DNAm correlations with glucose (r_{DNAm} ranging from 0.42 to 0.62), again of a greater magnitude than both genetic and phenotypic correlations. We observed negative DNAm correlations between each of the body-fat traits and HDL cholesterol, with a slightly stronger correlation observed for BMI. These correlations were in the same direction as genetic correlations with a similar magnitude, while phenotypic correlations were observed to be closer to zero. DNAm correlations for each of the body-fat traits with total cholesterol were found to not be significantly different from zero. This is consistent with genetic correlations between total cholesterol and both BMI and body-fat percentage, while the genetic correlations between WHR and total cholesterol was non-zero ($r_{DNAm} = 0.43$, SE = 0.22, p value = 0.02). Similarly, DNAm correlations between HDL cholesterol and glucose were observed to be similar to genetic correlations, although of slightly less magnitude. We observed moderate positive DNAm correlation between total cholesterol and both glucose and HDL cholesterol, while the genetic correlation was found to be not significantly different from zero between these trait pairs. Further, we demonstrate these results are independent of variance attributable to data structure by finding practically identical estimates for DNAm correlations when adjusting for the first 20 principal components of the DNAm levels and the first 20 principal components of the genetic data (Table S4).

DNAm correlations between sexes

Given previously reported genetic[28–30] and DNAm[7,31] sex differences for body-fat-related traits, we investigated

Please cite this article in press as: Hatton et al., Blood-based genome-wide DNA methylation correlations across body-fat- and adiposity-related biochemical traits, The American Journal of Human Genetics (2023), https://doi.org/10.1016/j.ajhg.2023.08.004

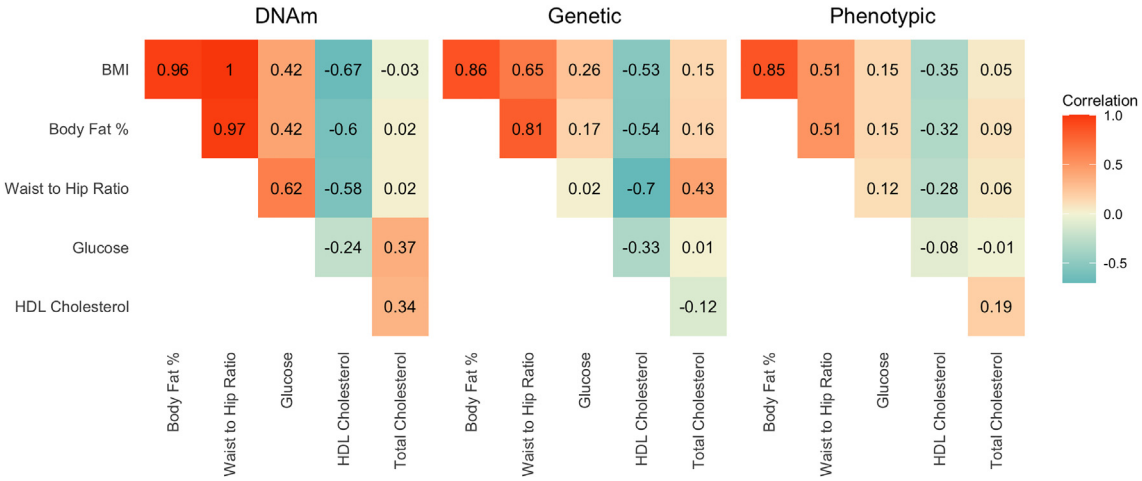

Figure 2. DNAm (left), genetic (middle), and phenotypic (right) correlations among six traits
Red, positive correlation; blue, negative correlation. DNAm correlations are estimated using bivariate OREML, genetic correlations using bivariate GREML, and phenotypic correlations using Pearson's correlation. Corresponding standard errors are in Table S3 along with p values from a log likelihood test for the hypothesis of fixing the DNAm correlation at both 1 and 0.

whether the contribution of DNAm for each trait was consistent across sex. First, we estimated the proportion of variance captured by DNAm in each sex. For BMI and body-fat percentage, the proportion of variance captured by DNAm was largely consistent across sexes, while for WHR, glucose, HDL cholesterol, and total cholesterol, the variance captured by DNAm in males was greater than that captured in females (Figure 3; Table S5). Next, DNAm correlations were calculated for each trait between sexes. Two traits were identified as having DNAm correlations significantly different from 1; however, only BMI survived multiple testing using a Bonferroni correction (BMI $r_{DNAm} = 0.79$, SE 0.07, p value $= 7.0 \times 10^{-4}$; WHR $r_{DNAm} = 0.95$, SE $= 0.04$, p value $= 0.016$; Figure 3; Table S6). Several previous studies have presented genetic correlations for BMI between the sexes that were significantly different from 1 (ranging from 0.93 to 0.96;[28–30]); however, the greater deviation between the sexes captured by DNAm correlation potentially suggests the presence of sex-differential biological consequences of BMI.

We further examined sex differences in the contribution of DNAm for BMI by investigating the presence of probe-by-sex interactions. We performed an EWAS for BMI including probe, sex, and the interaction between probe and sex as covariates in a linear model. We identified eight probes across four chromosomes with significant probe-by-sex interactions at $p < 6.4 \times 10^{-8}$, which is Bonferroni-corrected for the number of DNAm probes analyzed. We note that this set of probes represented six independent probes, with two pairs of probes that were closely located together likely co-methylated (correlation between DNAm M values > 0.8 between probe pairs cg16936953 and cg12054453, and cg18181703 and cg11047325). For all eight probes, DNAm was higher for females as BMI increased, with no significant association observed in males (Table 2), with trend plots provided in Figure S3. This may reflect a unique response in females to BMI levels.

Discussion

We investigated the shared associations of DNAm variation between body-fat- and adiposity-related biochemical traits by extending the OREML framework to a bivariate model, similar to the estimation of genetic correlation through GREML. For the majority of trait pairs, the DNAm correlations, while strongly concordant in direction, were observed to be greater in magnitude compared with both genetic and phenotypic correlations, particularly between body-fat traits. There are several potential explanations for this. DNAm is known to capture risk factors beyond genetics, suggesting DNAm correlations are likely capturing common environmental or lifestyle factors between traits such as dietary factors. DNAm correlations may also be capturing common consequence of these traits—that is, the consequence of both traits affecting downstream pathways, e.g., inflammation. This hypothesis is supported by previous studies that have demonstrated that while large amounts of the phenotypic variance can be captured by DNAm for some traits (e.g., BMI and smoking), for the most part, these have been implicated as arising from trait consequence.[8,13] In particular, Wahl et al. suggested that changes in DNAm (measured in blood and adiposity tissue) associated with BMI may be the consequence of changes in lipid and glucose metabolism associated with BMI.[8] This ascertainment of both causal and consequential effects may explain why DNAm correlations were observed to be of greater magnitude than their genetic counterparts. The strong positive DNAm correlations between each of the body-fat traits is consistent with DNAm derived from

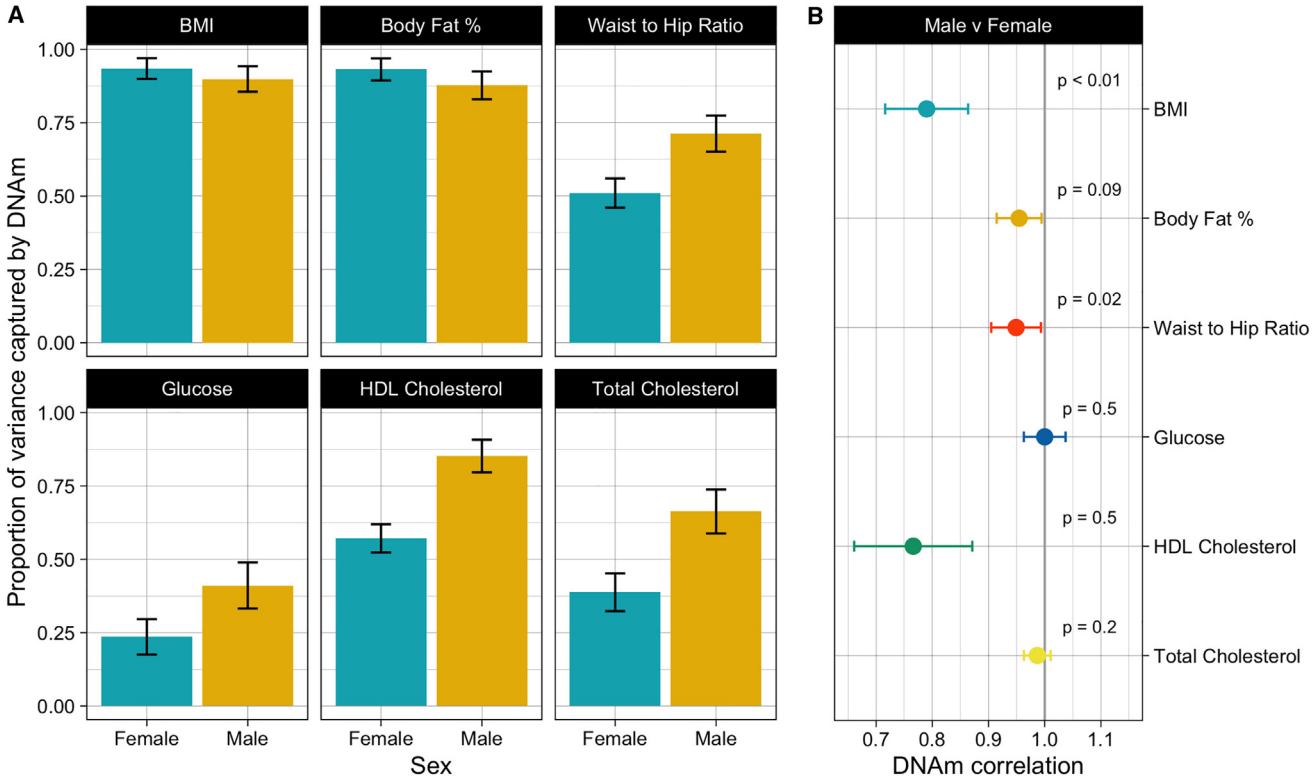

Figure 3. Differences in associations between DNAm and each of the traits in males and females
(A) The proportion of phenotypic variance captured by DNAm by sex for each trait.
(B) The DNAm correlation between sexes for each trait. Standard errors are indicated by error bars. p values from a log likelihood test for the hypothesis of fixing the DNAm correlation at 1 are presented in text below each estimate and in Table S6.

whole blood, reflecting a general response to adiposity, while genetic correlations are capturing differences in the genetic control of specific fat distribution. Support for such a conclusion in literature is conflicting. A recent study of DNAm in adipose tissue in women identified associations with body-fat distribution, of which 50% of sites replicated whole blood-derived DNAm.[32] Several other studies have demonstrated strong overlap between CpG sites associated with BMI, waist circumference, and body-fat percentage, indicating that common methylation sites are similarly influenced by both general and abdominal obesity.[33–35] However, Crocker et al.[35] found a low degree of overlap between waist circumference and body-fat percentage from subsequent Gene Ontology enrichment and differentially methylated region analyses, suggesting these measurements represent biologically distinct concepts. We note the inconsistency in conclusions from Crocker et al. may have been impacted by the investigation of overlap in significant results rather than formally testing for differences and additionally limited by sample size (n = 2,325).

We also recognize that, given that a large portion of the DNAm is under genetic control,[14] DNAm correlations are likely capturing part of the shared genetic contribution between these traits. We demonstrated that a large portion of the phenotypic variation captured by DNAm is separate from that being explained by SNPs, a conclusion that is supported in the literature.[7,8,13,15] Further, we identified independence between the MRM and GRM when fit as random effects in the univariate GREML framework using the covariance between random effects (CORE) GREML approach. Despite this, we were unable to formally determine whether the contribution of DNAm, which was shared between traits, was similarly separate of the shared genetic influence. This limitation in our study was likely due to sample size, with joint MRM and GRM bivariate REML models unable to converge and therefore unable to estimate DNAm correlations conditional on SNPs.

Given previously reported genetic[28–30] and DNAm[7,31] sex differences for body-fat-related traits, we investigated whether these are also captured by DNAm correlations. We identified a significant deviation from 1 in the DNAm correlation for BMI between males and females. Several previous studies have presented r_Gs for BMI between the sexes that were significantly different from 1 (ranging from 0.93 to 0.96[28–30]). The greater deviation captured by the DNAm correlation, however, potentially suggests the presence of sex-differential biological consequences of BMI. We further identified eight DNAm probe-by-sex interactions for BMI (which represent six independent DNAm sites), observing hypermethylation in females as BMI increased, with no association observed in males. Of note, all but one of these probes has been previously shown to be associated with BMI.[7,8,34,36,37] In particular, probe cg18181703 is located

Please cite this article in press as: Hatton et al., Blood-based genome-wide DNA methylation correlations across body-fat- and adiposity-related biochemical traits, The American Journal of Human Genetics (2023), https://doi.org/10.1016/j.ajhg.2023.08.004

Table 2. DNAm probes identified with probe-by-sex interactions ($p < 6.4 \times 10^{-8}$) with BMI

					In females			In males			Interaction		
Probe ID	CHR	Probe BP (CRCh37)	CpG island	Related gene	Probe effect	Probe SE	Probe p value	Probe effect	Probe SE	Probe p value	Effect	SE	p value
cg12269535	6	43142014	shore	*SRF*	−0.66	0.07	6.0×10^{-20}	−0.06	0.08	0.44	0.60	0.11	4.0×10^{-8}
cg16936953	17	57915665	open sea	*VMP1*	−0.37	0.05	2.0×10^{-14}	0.03	0.05	0.56	0.40	0.07	3.3×10^{-8}
cg12054453	17	57915717	open sea	*VMP1*	−0.27	0.04	9.4×10^{-14}	0.04	0.04	0.28	0.31	0.05	9.0×10^{-9}
cg19748455	17	76274856	open sea	–	−0.75	0.06	2.4×10^{-30}	−0.12	0.08	0.12	0.63	0.10	2.2×10^{-10}
cg18181703	17	76354621	shore	*SOCS3*	−0.79	0.07	1.2×10^{-27}	−0.18	0.08	0.02	0.60	0.11	1.3×10^{-8}
cg11047325	17	76354934	island	*SOCS3*	−0.43	0.04	4.8×10^{-25}	−0.08	0.04	0.06	0.35	0.06	1.1×10^{-8}
cg00840791	19	16453259	intergenic	–	−0.45	0.03	2.1×10^{-46}	−0.16	0.03	2.4×10^{-6}	0.29	0.05	7.4×10^{-10}
cg09349128	22	50327986	shore	*CRELD2*	−1.05	0.08	5.2×10^{-39}	−0.26	0.09	0.01	0.79	0.12	1.3×10^{-10}

Probe-association statistics (effect, SE, and p value) were estimated from female- and male-only models for each DNAm probe. Interaction (between sex and probe)-association statistics were estimated from a model containing both sexes. Effect sizes are in M values.

in *SOCS3*, a suppressor of the cytokine signaling pathway, and has been found to be inversely associated with BMI, WHR, triglycerides, and metabolic syndrome and positively associated with HDL.[7] It has also been shown to moderate the effect of cumulative stress on obesity.[38] DNAm of cg09349128 located in *PIM3*, a gene involved in energy metabolism, has been found to mediate the association between famine exposure and BMI. Additionally, probes cg16936953 and cg12054453 are located in *VMP1*, which has been implicated broadly in lipid homeostasis and regulation in the formation of lipid droplets and lipoproteins, for which dysregulation is involved in a variety of diseases including obesity, fatty liver disease, and cholesterol ester storage.[39,40] We also found evidence for a probe-by-sex interaction with DNAm at probe cg12269535 located in *SRF*, which is associated with insulin resistance and may contribute to the pathogenesis of type 2 diabetes.[41] We note that probe-by-sex interactions have been previously investigated in the context of BMI,[7,42] with each study identifying only a single CpG; however, we were unable to replicate any previous findings (Figure S4).

We recognize there are some caveats and further considerations for this work. The EPIC array captures only a small proportion of the methylome, with Hillary et al. previously demonstrating that decreasing the number of methylation sites reduces estimates of variance captured by DNAm and prediction metrics.[43] This impacts the interpretability of our analyses as a low variance captured by DNAm doesn't necessarily indicate a lack of correlation between DNAm and traits as DNAm sites, which are unmeasured, may contribute to the association. As such, greater coverage may resultingly influence DNAm correlation estimates. Similarly, while variance component estimation based on DNAm requires smaller samples sizes than needed for accurate estimation of genetic correlation due to the MRM capturing more variance, there is value in increasing sample sizes as well. In these analyses, we were unable to report on DNAm correlations conditional on SNPs as joint MRM and GRM bivariate REML models were unable to converge. We attempt to address this by adjusting univariate and bivariate OREML models based on DNAm with covariate adjustment for first 20 principal components of the genetic data. We find models with and without these adjustments yield practically identical estimates for both proportion of variance captured and DNAm correlations.

While it has been previously shown that much of the genetic control of DNAm is shared across populations,[44–47] as DNAm is also responsive to the environment, it would not be unexpected for such estimates to vary by ancestry or geography. While we suspect our results will be generalizable across comparable samples, replication in similar populations as well as populations of different ancestry, ethnicity, or geography would provide greater insight into these results. We also acknowledge the DNAm correlations represent the shared associations in DNAm variation between traits at a snapshot in time and within a particular tissue. DNAm levels are known to change over time, both as a progressive response to aging as well as due to environmental and stochastic influence, and resultingly, there may be variation in the estimated DNAm correlations within a population over time. Further, while this study examines DNAm correlations in whole blood, whether these samples accurately reflect the degree of shared DNAm associations in other tissues needs further validation.

Lastly, while DNAm correlations demonstrate the degree of similarity in epigenetic processes underlying these traits,

as this metric is quantified at the genome-wide level, it does not provide direct insights into specific processes (i.e., identifying individual loci, genes, or molecular pathways that are shared between traits). As such, the estimation of DNAm correlations can serve as an initial step in identification of the underlying epigenetic processes shared between traits and subsequently guide further investigation to uncover the specific molecular pathways involved. Longitudinal studies may elucidate timings of the shared response at individual points in the genome, provide insight into whether the shared epigenetic processes are causal or a consequence of disease, and direct potential follow-up approaches that could be undertaken to gain insights into the shared molecular pathways between traits.

Overall, we present an approach to investigating shared biology across traits using DNAm correlations. This has provided insight into obesity-related traits, showing the shared associations of DNAm between BMI, WHR, and body-fat percentage beyond that recognized through genetic correlation analysis and has identified sex-specific DNAm changes associated with BMI.

Data and code availability

According to the terms of consent for GS participants, access to data must be reviewed by the Generation Scotland Access Committee. Applications should be made to access@generationscotland.org.

The datasets generated during the current study are available in the supplemental tables.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2023.08.004.

Acknowledgments

GS received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006). Genotyping and DNAm profiling of the GS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, Edinburgh, Scotland, and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award Stratifying Resilience and Depression Longitudinally [STRADL; ref. 104036/Z/14/Z]). The DNAm data assayed for GS was partially funded by a 2018 NARSAD Young Investigator grant from the Brain & Behavior Research Foundation (ref. 27404; awardee: Dr. David M. Howard) and by a JMAS SIM fellowship from the Royal College of Physicians of Edinburgh (Awardee: Dr. Heather C. Whalley).

A.A.H. is supported by an Australian Government Research Training Program (RTP) scholarship. A.F.M. is supported by an Australian Research Council Future Fellowship (FT200100837).

Author contributions

A.A.H. contributed to the conception and design, analysis, interpretation, drafting, production, and revision of the manuscript. R.F.H. contributed to the data analysis, interpretation, and revision of the manuscript. E.B. contributed to the interpretation and revision of the manuscript. D.L.M. contributed to the data generation, interpretation, and revision of the manuscript. R.E.M. contributed to the conception and design, interpretation, drafting, production, and revision of the manuscript. A.F.M. contributed to the conception and design, drafting, production, and revision of the manuscript. All authors read and approved the final manuscript.

Declaration of interests

R.E.M. is a scientific advisor to the Epigenetic Clock Development Foundation and Optima Partners. R.F.H. has received consultant fees from Illumina and acts as a scientific advisor to Optima Partners.

Received: March 28, 2023
Accepted: August 3, 2023
Published: August 30, 2023

References

1. Williams, E.P., Mesidor, M., Winters, K., Dubbert, P.M., and Wyatt, S.B. (2015). Overweight and Obesity: Prevalence, Consequences, and Causes of a Growing Public Health Problem. Curr. Obes. Rep. *4*, 363–370. https://doi.org/10.1007/s13679-015-0169-4.
2. Guh, D.P., Zhang, W., Bansback, N., Amarsi, Z., Birmingham, C.L., and Anis, A.H. (2009). The incidence of co-morbidities related to obesity and overweight: A systematic review and meta-analysis. BMC Publ. Health *9*, 88. https://doi.org/10.1186/1471-2458-9-88.
3. GBD 2015 Obesity Collaborators, Afshin, A., Forouzanfar, M.H., Reitsma, M.B., Sur, P., Estep, K., Lee, A., Marczak, L., Mokdad, A.H., Moradi-Lakeh, M., et al. (2017). Health Effects of Overweight and Obesity in 195 Countries over 25 Years. N. Engl. J. Med. *377*, 13–27. https://doi.org/10.1056/NEJMoa1614362.
4. Reed, Z.E., Suderman, M.J., Relton, C.L., Davis, O.S.P., and Hemani, G. (2020). The association of DNA methylation with body mass index: distinguishing between predictors and biomarkers. Clin. Epigenetics *12*, 50. https://doi.org/10.1186/s13148-020-00841-5.
5. Do, W.L., Nguyen, S., Yao, J., Guo, X., Whitsel, E.A., Demerath, E., Rotter, J.I., Rich, S.S., Lange, L., Ding, J., et al. (2021). Associations between DNA methylation and BMI vary by metabolic health status: a potential link to disparate cardiovascular outcomes. Clin. Epigenetics *13*, 230. https://doi.org/10.1186/s13148-021-01194-3.
6. Bell, C.G. (2017). The Epigenomic Analysis of Human Obesity. Obesity *25*, 1471–1481. https://doi.org/10.1002/oby.21909.
7. Mendelson, M.M., Marioni, R.E., Joehanes, R., Liu, C., Hedman, Å.K., Aslibekyan, S., Demerath, E.W., Guan, W., Zhi, D., Yao, C., et al. (2017). 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. *14*, e1002215. https://doi.org/10.1371/journal.pmed.1002215.
8. Wahl, S., Drong, A., Lehne, B., Loh, M., Scott, W.R., Kunze, S., Tsai, P.C., Ried, J.S., Zhang, W., Yang, Y., et al. (2017). Epigenome-wide association study of body mass index, and the

adverse outcomes of adiposity. Nature *541*, 81–86. https://doi.org/10.1038/nature20784.

9. Klemp, I., Hoffmann, A., Müller, L., Hagemann, T., Horn, K., Rohde-Zimmermann, K., Tönjes, A., Thiery, J., Löffler, M., Burkhardt, R., et al. (2022). DNA methylation patterns reflect individual's lifestyle independent of obesity. Clin. Transl. Med. *12*, e851. https://doi.org/10.1002/ctm2.851.
10. Cao, V.T., Lea, R.A., Sutherland, H.G., Benton, M.C., Pishva, R.S., Haupt, L.M., and Griffiths, L.R. (2021). A genome-wide methylation study of body fat traits in the Norfolk Island isolate. Nutr. Metab. Cardiovasc. Dis. *31*, 1556–1563. https://doi.org/10.1016/j.numecd.2021.01.027.
11. Wu, Y., Tian, H., Wang, W., Li, W., Duan, H., and Zhang, D. (2022). DNA methylation and waist-to-hip ratio: an epigenome-wide association study in Chinese monozygotic twins. J. Endocrinol. Invest. *45*, 2365–2376. https://doi.org/10.1007/s40618-022-01878-4.
12. Jones, A.C., Irvin, M.R., Claas, S.A., and Arnett, D.K. (2021). Lipid Phenotypes and DNA Methylation: a Review of the Literature. Curr. Atheroscler. Rep. *23*, 71. https://doi.org/10.1007/s11883-021-00965-w.
13. Trejo Banos, D., McCartney, D.L., Patxot, M., Anchieri, L., Battram, T., Christiansen, C., Costeira, R., Walker, R.M., Morris, S.W., Campbell, A., et al. (2020). Bayesian reassessment of the epigenetic architecture of complex traits. Nat. Commun. *11*, 2865. https://doi.org/10.1038/s41467-020-16520-1.
14. Min, J.L., Hemani, G., Hannon, E., Dekkers, K.F., Castillo-Fernandez, J., Luijk, R., Carnero-Montoro, E., Lawson, D.J., Burrows, K., Suderman, M., et al. (2021). Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat. Genet. *53*, 1311–1321. https://doi.org/10.1038/s41588-021-00923-x.
15. Shah, S., Bonder, M.J., Marioni, R.E., Zhu, Z., McRae, A.F., Zhernakova, A., Harris, S.E., Liewald, D., Henders, A.K., Mendelson, M.M., et al. (2015). Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. Am. J. Hum. Genet. *97*, 75–85. https://doi.org/10.1016/j.ajhg.2015.05.014.
16. Zhang, F., Chen, W., Zhu, Z., Zhang, Q., Nabais, M.F., Qi, T., Deary, I.J., Wray, N.R., Visscher, P.M., McRae, A.F., and Yang, J. (2019). OSCA: a tool for omic-data-based complex trait analysis. Genome Biol. *20*, 107. https://doi.org/10.1186/s13059-019-1718-z.
17. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M., and Wray, N.R. (2012). Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics *28*, 2540–2542. https://doi.org/10.1093/bioinformatics/bts474.
18. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.
19. Smith, B.H., Campbell, A., Linksted, P., Fitzpatrick, B., Jackson, C., Kerr, S.M., Deary, I.J., MacIntyre, D.J., Campbell, H., McGilchrist, M., et al. (2013). Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int. J. Epidemiol. *42*, 689–700. https://doi.org/10.1093/ije/dys084.
20. Smith, B.H., Campbell, H., Blackwood, D., Connell, J., Connor, M., Deary, I.J., Dominiczak, A.F., Fitzpatrick, B., Ford, I., Jackson, C., et al. (2006). Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. BMC Med. Genet. *7*, 74. https://doi.org/10.1186/1471-2350-7-74.
21. Marioni, R.E., Campbell, A., Scotland, G., Hayward, C., Porteous, D.J., and Deary, I.J. (2016). Differential effects of the APOE e4 allele on different domains of cognitive ability across the life-course. Eur. J. Hum. Genet. *24*, 919–923. https://doi.org/10.1038/ejhg.2015.210.
22. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nat. Genet. *48*, 1279–1283. https://doi.org/10.1038/ng.3643.
23. McCartney, D.L., Hillary, R.F., Stevenson, A.J., Ritchie, S.J., Walker, R.M., Zhang, Q., Morris, S.W., Bermingham, M.L., Campbell, A., Murray, A.D., et al. (2018). Epigenetic prediction of complex traits and death. Genome Biol. *19*, 136. https://doi.org/10.1186/s13059-018-1514-1.
24. Chen, Y.A., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J., and Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics *8*, 203–209. https://doi.org/10.4161/epi.23470.
25. McCartney, D.L., Walker, R.M., Morris, S.W., McIntosh, A.M., Porteous, D.J., and Evans, K.L. (2016). Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genom. Data *9*, 22–24. https://doi.org/10.1016/j.gdata.2016.05.012.
26. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinf. *13*, 86. https://doi.org/10.1186/1471-2105-13-86.
27. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569. https://doi.org/10.1038/ng.608.
28. Bernabeu, E., Canela-Xandri, O., Rawlik, K., Talenti, A., Prendergast, J., and Tenesa, A. (2021). Sex differences in genetic architecture in the UK Biobank. Nat. Genet. *53*, 1283–1289. https://doi.org/10.1038/s41588-021-00912-0.
29. Traglia, M., Bseiso, D., Gusev, A., Adviento, B., Park, D.S., Mefford, J.A., Zaitlen, N., and Weiss, L.A. (2017). Genetic Mechanisms Leading to Sex Differences Across Common Diseases and Anthropometric Traits. Genetics *205*, 979–992. https://doi.org/10.1534/genetics.116.193623.
30. Rawlik, K., Canela-Xandri, O., and Tenesa, A. (2016). Evidence for sex-specific genetic architectures across a spectrum of human complex traits. Genome Biol. *17*, 166. https://doi.org/10.1186/s13059-016-1025-x.
31. Shungin, D., Winkler, T.W., Croteau-Chonka, D.C., Ferreira, T., Locke, A.E., Mägi, R., Strawbridge, R.J., Pers, T.H., Fischer, K., Justice, A.E., et al. (2015). New genetic loci link adipose and insulin biology to body fat distribution. Nature *518*, 187–196. https://doi.org/10.1038/nature14132.
32. Divoux, A., Eroshkin, A., Erdos, E., Sandor, K., Osborne, T.F., and Smith, S.R. (2021). DNA Methylation as a Marker of Body Shape in Premenopausal Women. Front. Genet. *12*, 709342. https://doi.org/10.3389/fgene.2021.709342.

33. Dhana, K., Braun, K.V.E., Nano, J., Voortman, T., Demerath, E.W., Guan, W., Fornage, M., van Meurs, J.B.J., Uitterlinden, A.G., Hofman, A., et al. (2018). An Epigenome-Wide Association Study of Obesity-Related Traits. Am. J. Epidemiol. *187*, 1662–1669. https://doi.org/10.1093/aje/kwy025.
34. Chen, Y., Kassam, I., Lau, S.H., Kooner, J.S., Wilson, R., Peters, A., Winkelmann, J., Chambers, J.C., Chow, V.T., Khor, C.C., et al. (2021). Impact of BMI and waist circumference on epigenome-wide DNA methylation and identification of epigenetic biomarkers in blood: an EWAS in multi-ethnic Asian individuals. Clin. Epigenetics *13*, 195. https://doi.org/10.1186/s13148-021-01162-x.
35. Crocker, K.C., Domingo-Relloso, A., Haack, K., Fretts, A.M., Tang, W.Y., Herreros, M., Tellez-Plaza, M., Daniele Fallin, M., Cole, S.A., and Navas-Acien, A. (2020). DNA methylation and adiposity phenotypes: an epigenome-wide association study among adults in the Strong Heart Study. Int. J. Obes. *44*, 2313–2322. https://doi.org/10.1038/s41366-020-0646-z.
36. Demerath, E.W., Guan, W., Grove, M.L., Aslibekyan, S., Mendelson, M., Zhou, Y.H., Hedman, Å.K., Sandling, J.K., Li, L.A., Irvin, M.R., et al. (2015). Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. Hum. Mol. Genet. *24*, 4464–4479. https://doi.org/10.1093/hmg/ddv161.
37. Sun, D., Zhang, T., Su, S., Hao, G., Chen, T., Li, Q.Z., Bazzano, L., He, J., Wang, X., Li, S., and Chen, W. (2019). Body Mass Index Drives Changes in DNA Methylation: A Longitudinal Study. Circ. Res. *125*, 824–833. https://doi.org/10.1161/circresaha.119.315397.
38. Xu, K., Zhang, X., Wang, Z., Hu, Y., and Sinha, R. (2018). Epigenome-wide association analysis revealed that SOCS3 methylation influences the effect of cumulative stress on obesity. Biol. Psychol. *131*, 63–71. https://doi.org/10.1016/j.biopsycho.2016.11.001.
39. Li, Y.E., Wang, Y., Du, X., Zhang, T., Mak, H.Y., Hancock, S.E., McEwen, H., Pandzic, E., Whan, R.M., Aw, Y.C., et al. (2021). TMEM41B and VMP1 are scramblases and regulate the distribution of cholesterol and phosphatidylserine. J. Cell Biol. *220*, e202103105. https://doi.org/10.1083/jcb.202103105.
40. Ghanbarpour, A., Valverde, D.P., Melia, T.J., and Reinisch, K.M. (2021). A model for a partnership of lipid transfer proteins and scramblases in membrane expansion and organelle biogenesis. Proc. Natl. Acad. Sci. USA *118*, e2101562118. https://doi.org/10.1073/pnas.2101562118.
41. Jin, W., Goldfine, A.B., Boes, T., Henry, R.R., Ciaraldi, T.P., Kim, E.Y., Emecan, M., Fitzpatrick, C., Sen, A., Shah, A., et al. (2011). Increased SRF transcriptional activity in human and mouse skeletal muscle is a signature of insulin resistance. J. Clin. Invest. *121*, 918–929. https://doi.org/10.1172/jci41940.
42. Wang, J., Zhang, H., Rezwan, F.I., Relton, C., Arshad, S.H., and Holloway, J.W. (2021). Pre-adolescence DNA methylation is associated with BMI status change from pre- to post-adolescence. Clin. Epigenetics *13*, 64. https://doi.org/10.1186/s13148-021-01042-4.
43. Hillary, R.F., McCartney, D.L., McRae, A.F., Campbell, A., Walker, R.M., Hayward, C., Horvath, S., Porteous, D.J., Evans, K.L., and Marioni, R.E. (2022). Identification of influential probe types in epigenetic predictions of human traits: implications for microarray design. Clin. Epigenetics *14*, 100. https://doi.org/10.1186/s13148-022-01320-9.
44. Smith, A.K., Kilaru, V., Kocak, M., Almli, L.M., Mercer, K.B., Ressler, K.J., Tylavsky, F.A., and Conneely, K.N. (2014). Methylation quantitative trait loci (meQTLs) are consistently detected across ancestry, developmental stage, and tissue type. BMC Genom. *15*, 145. https://doi.org/10.1186/1471-2164-15-145.
45. Husquin, L.T., Rotival, M., Fagny, M., Quach, H., Zidane, N., McEwen, L.M., MacIsaac, J.L., Kobor, M.S., Aschard, H., Patin, E., and Quintana-Murci, L. (2018). Exploring the genetic basis of human population differences in DNA methylation and their causal impact on immune gene regulation. Genome Biol. *19*, 222. https://doi.org/10.1186/s13059-018-1601-3.
46. Kassam, I., Tan, S., Gan, F.F., Saw, W.-Y., Tan, L.W.-L., Moong, D.K.N., Soong, R., Teo, Y.-Y., and Loh, M. (2021). Genome-wide identification of cis DNA methylation quantitative trait loci in three Southeast Asian Populations. Hum. Mol. Genet. *30*, 603–618. https://doi.org/10.1093/hmg/ddab038.
47. Hawe, J.S., Wilson, R., Schmid, K.T., Zhou, L., Lakshmanan, L.N., Lehne, B.C., Kühnel, B., Scott, W.R., Wielscher, M., Yew, Y.W., et al. (2022). Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function. Nat. Genet. *54*, 18–29. https://doi.org/10.1038/s41588-021-00969-x.