Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk

Tom G. Richardson,[1,*] Jie Zheng,[1] George Davey Smith,[1] Nicholas J. Timpson,[1] Tom R. Gaunt,[1] Caroline L. Relton,[1] and Gibran Hemani[1]

The extent to which genetic influences on cardiovascular disease risk are mediated by changes in DNA methylation levels has not been systematically explored. We developed an analytical framework that integrates genetic fine mapping and Mendelian randomization with epigenome-wide association studies to evaluate the causal relationships between methylation levels and 14 cardiovascular disease traits. We identified ten genetic loci known to influence proximal DNA methylation which were also associated with cardiovascular traits after multiple-testing correction. Bivariate fine mapping provided evidence that the individual variants responsible for the observed effects on cardiovascular traits at the *ADCY3* and *ADIPOQ* loci were potentially mediated through changes in DNA methylation, although we highlight that we are unable to reliably separate causality from horizontal pleiotropy. Estimates of causal effects were replicated with results from large-scale consortia. Genetic variants and CpG sites identified in this study were enriched for histone mark peaks in relevant tissue types and gene promoter regions. Integrating our results with expression quantitative trait loci data, we provide evidence that variation at these regulatory regions is likely to also influence gene expression levels at these loci.

Introduction

Approximately 88% of trait-associated variants detected by genome-wide association studies (GWASs) reside in non-coding regions of the genome and might act through gene regulation.[1] Recent studies have incorporated data on genetic variants associated with gene expression (expression quantitative trait loci [eQTLs]) into results from GWASs of complex traits to help identify the putative causal variant in a genomic region, as well as provide evidence suggesting which genes might be influenced by this variant.[2–5] This direction of inquiry can be extended to other "-omic" data types to gain further insights into the mechanistic pathway between genetic variant and causally associated trait. In this study, we introduce an alternative analytical framework to integrate genetic predictors of DNA methylation levels with complex traits to evaluate bi-directional causal relationships.

DNA methylation is an epigenetic regulation mechanism that has been shown to play a key role in many biological processes and disease susceptibility.[6–8] Recent studies have had success in identifying genetic variants associated with DNA methylation (methylation quantitative trait loci [mQTLs]) and report that they appear to overlap with eQTLs at a large number of loci across the genome.[9,10] This suggests that both DNA methylation and gene expression could reside along the causal pathway between genetic variation and disease, although thus far, uncovering evidence of a mediated effect between mQTLs and traits has been more limited than using eQTLs.[11–14] Identifying epigenetic markers for disease risk should prove valuable in understanding the underlying biological mechanisms for trait-associated variants.[15] Indeed, the value of this approach was demonstrated in a recent study that applied the SMR[2] method to uncover pleiotropic effects between methylation levels and a range of complex traits.[16]

Mendelian randomization (MR) is a method by which genetic variants robustly associated with modifiable exposures can be used as instrumental variables to infer causality among correlated traits.[17,18] If DNA methylation resides along the causal pathway between genetic variant and trait, we would expect it to be correlated with our trait of interest. However, much like other traits analyzed in epidemiological studies, DNA methylation is prone to confounding and reverse causation. Using an MR framework, we can investigate whether DNA methylation has a causal relationship with a phenotypic outcome, suggesting that it might reside along the causal pathway to disease.[19] Effects such as this can be referred to as "mediation," as DNA methylation is mediating the effect from genetic variant to phenotype along the same biological pathway. As discussed in a recent review, MR has advantages over alternative approaches in mediation analysis (such as the causal inference test[20]), as it can detect the correct direction of effect in the presence of measurement error.[21] It is important to note that all current methods are faced with the challenge of distinguishing mediation from horizontal pleiotropy, defined as effects where genetic variation influences multiple phenotypes simultaneously[22] (such as DNA methylation and a complex trait) via independent biological pathways.

[1]MRC Integrative Epidemiology Unit, Bristol Medical School (Population Health Sciences), University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK
*Correspondence: tom.g.richardson@bristol.ac.uk
https://doi.org/10.1016/j.ajhg.2017.09.003.
© 2017 The Authors. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

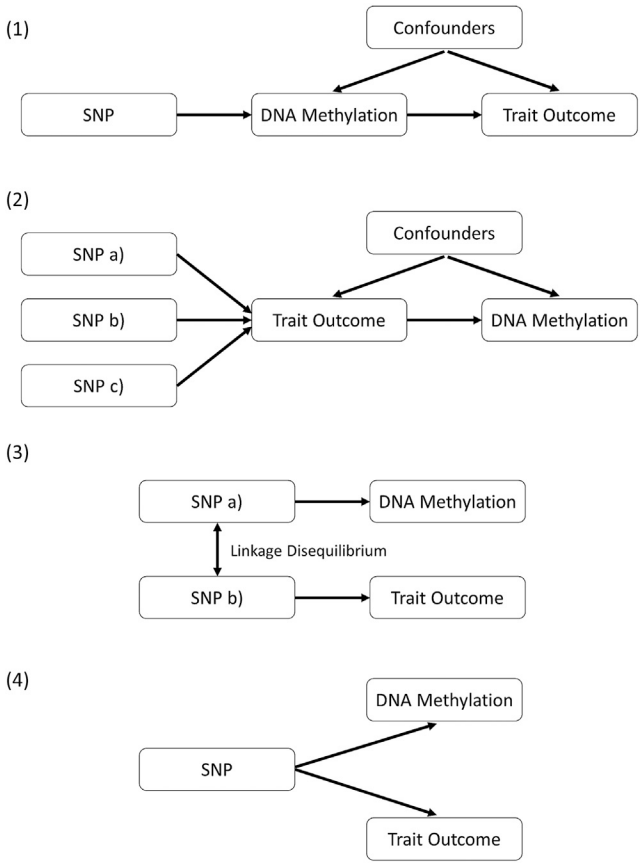

Figure 1. Explanations Evaluated to Explain Observed Associations between mQTLs and Trait Outcomes
(1) The genetic variant has an effect on the phenotype, mediated through DNA methylation.
(2) The genetic variant has an effect on the phenotype by alternative biological mechanisms, which then has a downstream effect on DNA methylation at this locus.
(3) The genetic variant that influences DNA methylation is simply in LD with another variant that is influencing the associated trait.
(4) The genetic variant influences both DNA methylation and phenotype by two independent biological pathways (also known as horizontal pleiotropy).

Recent approaches to MR have shown that the robustness of causal inference is improved if there are many instruments because one can evaluate whether the SNP effects on the causal trait are proportional to the SNP effects on the consequential trait.[17,23] We exploit this property to evaluate the causal influence of complex traits (which typically have many instruments) on DNA methylation (i.e., bi-directional MR[24]). A pitfall of evaluating the causal influence of DNA methylation on complex traits, however, is that DNA methylation is typically instrumented by only a single *cis*-acting variant. Hence, an unreliable MR estimate of causality could arise simply because the mQTL is in linkage disequilibrium (LD) with a variant that influences the cardiovascular trait through means other than the methylation level.

Together, the causal relationships between DNA methylation and cardiovascular traits are delineated into four potential categories (Figure 1).

1. The genetic variant has an effect on the phenotype, mediated by DNA methylation.
2. The genetic variant has an effect on the phenotype by alternative biological mechanisms, which then has a downstream effect on DNA methylation at this locus.
3. The genetic variant that influences DNA methylation is simply in LD with another variant that is influencing the associated trait.
4. The genetic variant influences both DNA methylation and phenotype by two independent biological pathways (also known as horizontal pleiotropy).

To address this issue, in this study we developed and implemented a framework that integrates MR with fine mapping to evaluate the likelihood that the mQTL is the same causal variant as the SNP influencing the cardiovascular trait. Other colocalization methods using intermediate traits have been devised for this purpose,[2,25,26] including the joint likelihood mapping (JLIM) method,[27] which we used to support our findings. We also undertook functional informatics and incorporated eQTL data because these can support findings suggesting that DNA methylation resides on the causal pathway between variant and disease. However, a limitation of using single-variant instruments in general is that it is not possible to reliably distinguish horizontal pleiotropy from mediation.[28]

In our discovery analysis, we used genotype and DNA methylation data from prepubertal individuals to discover causal pathways on early childhood phenotypes. Replication was then undertaken with GWAS summary statistics from large-scale consortia.

Material and Methods

The Avon Longitudinal Study of Parents and Children (ALSPAC)

ALSPAC is a population-based cohort study investigating genetic and environmental factors that affect the health and development of children. The study methods are described in detail elsewhere.[29,30] In brief, 14,541 pregnant women residents in the former region of Avon, UK, with an expected delivery date between April 1, 1991 and December 31, 1992, were eligible to take part in ALSPAC. Detailed information and biosamples have been collected on these women and their offspring at regular intervals, which are available through a searchable data dictionary.

Written informed consent was obtained for all study participants. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees.

Accessible Resource for Integrative Epigenomic Studies Project (ARIES)

Samples

Blood samples were obtained for 1,018 ALSPAC mother-offspring pairs (mothers at two time points and their offspring at three time

points) as part of the Accessible Resource for Integrative Epigenomic Studies project (ARIES).[31] The Illumina HumanMethylation450 (450K) BeadChip array was used to measure DNA methylation at over 480,000 sites across the epigenome.

Methylation Assays

DNA samples were treated with bisulfite with the Zymo EZ DNA Methylation Kit (Zymo). The Illumina HumanMethylation450 BeadChip (HM450k) was used to measure methylation across the genome, and the following arrays were scanned by Illumina iScan, as well as reviewed for quality by GenomeStudio. A purpose-built laboratory information management system (LIMS) was responsible for generating batch variables during data generation. LIMS also reported quality control (QC) metrics for the standard probes on the HM450k for all samples and excluded those that failed QC. We also excluded data points with a read count of 0 or with a low signal-to-noise ratio (p value > 0.01) on the basis of the QC report from Illumina to maintain the integrity of probe measurements. We then compared methylation measurements across time points for the same individual and with SNP-chip data (HM450k probes clustered by k-means) to identify and remove sample mismatches. All remaining data from probes were normalized with the Touleimat and Tost[32] algorithms in R with the wateRmelon package.[33] Then we rank-normalized the data to remove outliers. We removed potential batch effects by regressing data points on all covariates. These included the bisulfite-converted DNA (BCD) plate batch and white blood cell count, which was adjusted for with the "estimateCellCounts" function in the minfi Bioconductor package.[34]

Genotyping Assays

Genotype data were available for all ALSPAC individuals enrolled in the ARIES project, which had previously undergone quality control, cleaning, and imputation at the cohort level. ALSPAC offspring selected for this project had previously been genotyped with the Illumina HumanHap550 quad genome-wide SNP genotyping platform (Illumina) by the Wellcome Trust Sanger Institute (WTSI, Cambridge, UK) and the Laboratory Corporation of America (LCA). Samples were excluded on the basis of incorrect sex assignment, abnormal heterozygosity (<0.320 or >0.345 for WTSI data; <0.310 or >0.330 for LCA data), high missingness ($>3\%$), cryptic relatedness ($>10\%$ identity by descent), and non-European ancestry (detected by multidimensional scaling analysis). After QC, 500,527 SNP loci were available for the directly genotype dataset.

Imputation

Imputation was performed with a joint reference panel of variants discovered through whole-genome sequencing (WGS) in the UK10K project[35] along with known variants taken from the 1000 Genomes reference panel. We developed additional functionality in IMPUTE2[36] so we could use each reference panel to impute missing variants in their counterparts before ultimately combining them together. Following Gaunt et al.,[8] before imputation we performed strict filtering by using Hardy-Weinberg equilibrium $p > 5 \times 10^{-7}$ and minor allele frequency (MAF) > 0.01. After imputation, we converted the dosages to best-guess genotypes and filtered to keep only variants with an imputation quality score ≥ 0.8 and MAF > 0.01.

Phenotypes

The 14 phenotypes analyzed in this study are as follows. At the ALSPAC clinic, subjects aged 7 years (mean age: 7.5, range: 7.1–8.8) were measured; height was measured to the nearest 0.1 cm with a Harpenden stadiometer (Holtain Crosswell), and weight was measured to the nearest 0.1 kg on Tanita electronic scales. Body mass index (BMI) was calculated as (weight [kg])/(height [m]).[2] Blood pressure was measured with a Dinamap 9301 vital monitor using the appropriate cuff size by trained staff. Two readings of both systolic and diastolic blood pressure (SBP and DBP, respectively) were taken when the study participants were at rest, and the mean of each was used as a measurement in our analysis.

Non-fasting blood samples were taken from participants who attended the clinic at age 10 years (mean age: 9.9, range: 8.9–11.5). Plasma lipid concentrations (total cholesterol [TC], triglycerides [TG], and high-density lipoprotein cholesterol [HDL]) were measured by modification of the standard Lipid Research Clinics Protocol with enzymatic reagents for lipid determination.[37] Low-density lipoprotein cholesterol (LDL) concentration was subsequently calculated with the Friedwald equation[38] as follows:

$$\mathrm{LDLc} = \mathrm{TC} - (\mathrm{HDLc} + \mathrm{TG} \times 0.45)$$

Very-low-density lipoprotein cholesterol (VLDL) concentration was calculated as follows:

$$\mathrm{VLDLc} = \mathrm{TC} - (\mathrm{HDLc} + \mathrm{LDLc})$$

Apolipoprotein A (Apo A1) and apolipoprotein b (Apo B) were measured by immunoturbidimetric assays (Roche). Interleukin 6 (IL-6) and adiponectin were measured by enzyme-linked immunosorbent assay (R&D Systems). High-sensitivity C-reactive protein (CRP) was measured by an automated particle-enhanced immunoturbidimetric assay (Roche). Leptin was measured in house by a linked immunosorbent assay that had been validated against commercial methods.[39] All assay coefficients of variation were $<5\%$.

Statistical Analysis

We undertook an mQTL-wide association study (MWAS) to evaluate the association between variants known to influence DNA methylation (referred to hereafter as mQTL) and each trait in turn. This was decided over a conventional epigenome-wide association study (EWAS) (i.e., evaluating the association between methylation levels at CpG sites and traits) given that ALSPAC had a larger proportion of individuals with genotype data than with 450K data after phenotypes were merged.

All mQTLs previously identified in ARIES were considered for this analysis, and the methods have been described in detail previously.[8] In brief, to discover mQTLs, Gaunt et al.[8] used a linear regression model adjusted for age, sex, bisulphite conversion batch, the top ten ancestry principal components, and cell counts to evaluate the associations of 8,074,398 imputed genetic variants against each of the 395,625 eligible methylation probes. We filtered methylation probes for exclusion on the basis of evaluations by Naeem et al.[40] by using their criteria of overlapping SNPs at CpG probes, probes that map to multiple locations and repeats on the 450K array. We applied a conservative multiple-testing correction to define mQTLs ($p < 1.0 \times 10^{-14}$). This threshold was selected because it equates to a false-positive rate of 0.2% after a Bonferroni correction is applied to account for the number of tests undertaken previously in ARIES. Furthermore, this strict threshold reduces the risk of MR analyses suffering from weak instrument bias. Full details on the mQTL analysis can be found in the study by Gaunt et al.[8]

The mQTL discovery study used the COJO-slct routine in GCTA to identify independent mQTLs, which was important to ensure that variants used as instruments were independent for

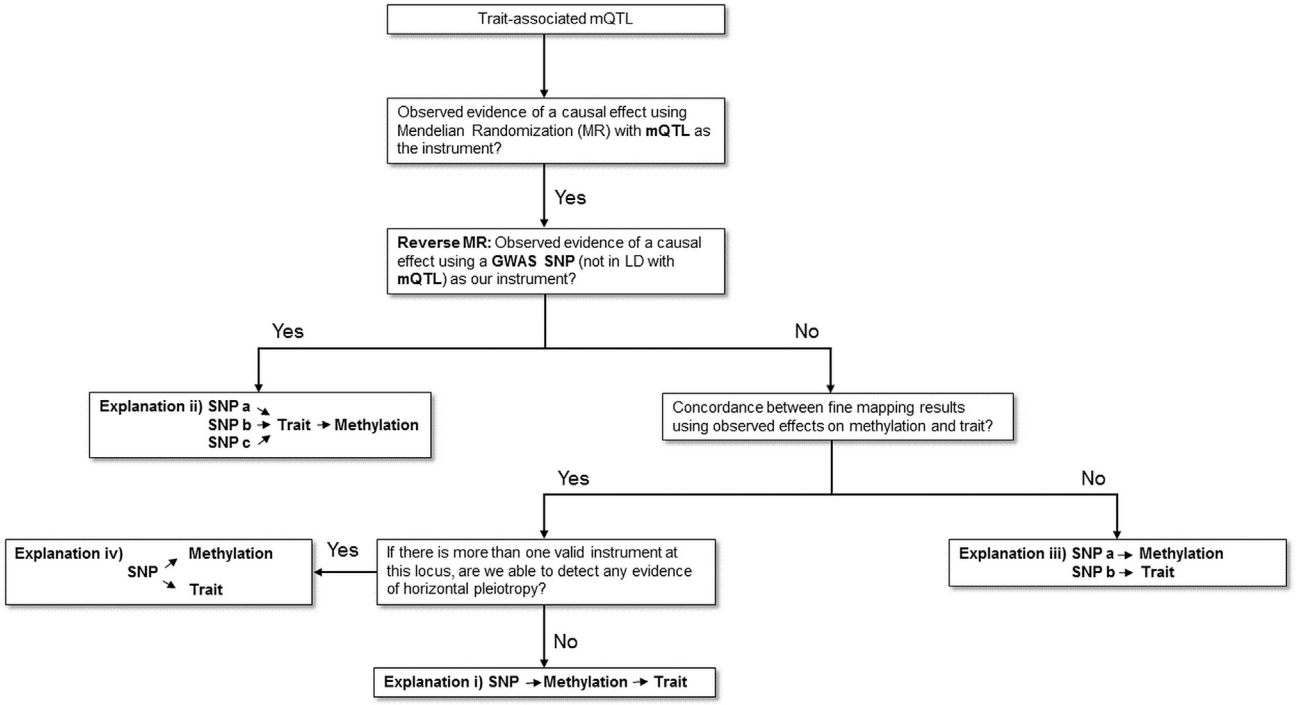

Figure 2. Analysis Pipeline to Evaluate Explanations for Observed Associations between mQTLs and Trait Outcomes
This flowchart provides an overview of the analysis plan in this study for evaluating four different explanations that might explain trait-associated mQTLs.

downstream MR analyses. We excluded mQTLs associated with a CpG site that was more than 1 Mb away (known as *trans*-mQTLs), therefore leaving mQTLs that were associated only with a nearby CpG site (known as *cis*-mQTLs). This was to reduce the possibility of pleiotropy in our analysis given that variants associating with methylation at multiple CpG sites across the epigenome might influence independent biological pathways simultaneously. This left 37,812 independent mQTL eligible for analysis.

The mQTLs were analyzed sequentially with each trait by linear regression with adjustment for age and sex. We also performed a sensitivity analysis adjusting for the first ten principal components to evaluate whether population stratification was influencing our results in this analysis, although we did not anticipate this given previous evaluations of population structure in the ALSPAC cohort.[41] Results were plotted on a Manhattan plot with code derived from the qqman R package.[42] Scripts to generate this plot are available at the location specified in the Web Resources.

Mendelian Randomization Analysis

Observed associations between genotype and traits that survived a stringent multiple-testing threshold (i.e., $p < 0.05$/number of tests undertaken) were then analyzed by MR. We performed this analysis to estimate the potential causal effect of DNA methylation on cardiovascular traits, given that we anticipated observing evidence of association after having already undertaken an MWAS. MR was undertaken by two-stage least-squares (2SLS) regression with DNA methylation as our exposure, phenotypic trait as our outcome, and the relevant mQTL as our instrumental variable. Measures of DNA methylation were initially taken from the childhood time point in ARIES (mean age, 7.5 years; standard deviation, 0.15) because this was the closest time point to phenotype measurements. Follow-up analyses were also undertaken with methylation data from the birth time point (with cord blood) and the adolescent time point (mean age, 17.1 years; standard deviation, 1.01). We used the R package "systemfit"[43] to obtain causal effect estimates with 2SLS.

We replicated observed effects by undertaking a two-sample MR analysis (2SMR)[44] with estimated effects between genetic variants and associated traits obtained from published studies. Moreover, a two-sample framework removes any potential bias encountered in the discovery analysis as a result of the existence of effects on both methylation and traits in the same sample. When observed effects for sentinel mQTL were not available from published studies, we used variants in LD with these SNPs instead ($r^2 > 0.8$).

Figure 1 illustrates the four possible explanations investigated where evidence of a causal effect was observed by MR. Figure 2 provides an overview of our approach to investigate these explanations. To robustly test explanation 2, we performed the reverse MR analysis, evaluating whether the cardiovascular trait influenced DNA methylation levels at the CpG site of interest. Instruments for this analysis were identified with the NHGRI-EBI GWAS Catalog.[45] Relevant GWASs for IL-6 were not available at the time of analysis and so we identified instruments on the basis of the findings from Naitza et al.[46] ($p < 5.0 \times 10^{-8}$). A p value greater than 0.05 indicated that explanation 2 was unlikely in each instance.

Bivariate Fine Mapping

Bivariate fine mapping was undertaken with FINEMAP[47] at each locus detected in the previous analysis. For each variant at a locus, FINEMAP generates a Bayes factor that reflects the likelihood that the variant is the underlying causal variant at this region. Bivariate fine mapping requires all variants at a locus to be fine mapped with two different effect estimates: (1) observed effects between SNPs

Table 1. Results of Linear Regression Analysis between Genetic Variants and Traits

SNP	Gene	CpG	Trait	Sample Size	Beta	SE	p Value	% Explained
rs266772	*ADIPOQ*	cg05578595	adiponectin (ng/mL)	4,248	−0.992	0.070	1.72×10^{-44}	4.51%
rs687621	*ABO*	cg21160290	IL-6 (pg/mL)	4,241	−0.265	0.022	1.15×10^{-31}	3.05%
rs13375019	*LEPR*	cg04111102	CRP (mg/L)	4,251	−0.213	0.022	2.65×10^{-22}	2.20%
rs7549250	*IL6R*	cg02856953	IL-6 (pg/mL)	4,241	−0.176	0.022	9.71×10^{-16}	1.40%
rs169109	*ADIPOQ*	cg05578595	adiponectin (ng/mL)	4,248	−0.167	0.022	1.44×10^{-14}	1.34%
rs541041	*APOB*	cg25035485	Apo B (g/L)	4,251	−0.209	0.028	3.76×10^{-14}	1.32%
rs7528419	*SORT1*	cg00908766	Apo B (g/L)	4,251	−0.196	0.026	4.63×10^{-14}	1.30%
rs625145	*APOA1*	cg04087571	Apo A1 (g/L)	4,251	0.200	0.027	9.78×10^{-14}	0.94%
rs174544	*FADS1*	cg19610905	total cholesterol (mmol/L)	4,250	−0.143	0.023	8.61×10^{-10}	0.86%
rs6749422	*ADCY3*	cg01884057	BMI	6,076	0.109	0.018	1.28×10^{-9}	0.55%

Abbreviations are as follows: SNP, single-nucleotide polymorphism; gene, most likely affected gene; CpG, 450K probe ID; trait, associated trait; sample size, sample size for this effect; beta, observed effect size (units in standard deviations); SE, standard error of the effect size; p value, p value for observed effect; and % explained, proportion of trait variance explained by mQTLs.

and DNA methylation and (2) observed effects between SNPs and outcome phenotypes. Given that we initially pruned all mQTL effects to identify independent loci, we included only variants that were in high LD ($r^2 \geq 0.8$) with the sentinel SNP for each association signal before applying FINEMAP with default settings. Interpretation of these results is therefore based on at least one underlying causal variant at each loci, given that follow-up analyses are necessary for evaluating whether multiple causal variants might be contributing to observed effects. Posterior probabilities to reflect the likelihood of multiple causal variants were calculated with FINEMAP.

We performed this analysis to evaluate explanation 3, that the mQTL analyzed might simply be in LD with the putative causal variant for the phenotypic trait. This was necessary because when the relationship between DNA methylation at a CpG site and the outcome trait is evaluated, there could be only one valid instrumental variable (i.e., the mQTL at this region). Bivariate fine mapping in this instance therefore evaluates whether the causal mQTL at a locus is likely to be the same causal variant for the observed effect on the outcome trait. However, it does not rule out the possibility that a single variant influences DNA methylation and an outcome trait through independent biological pathways (i.e., explanation 4).

Concordance between the top SNPs for the two sets of fine-mapping analyses would suggest that explanation 1 might be responsible for the observed effect and that DNA methylation resides on the causal pathway between variant and phenotypic trait. Bivariate fine mapping using effect estimates for both methylation and cardiovascular traits was advantageous in this study because we were able to obtain estimates for all SNPs in our dataset without having to rely on summary statistics. The concordance rate was defined as identifying the same variant from both analyses as causal after accounting for chance. We achieved this by identifying the rank of the top variant from the methylation-based analysis in the list of variants from the cardiovascular-trait analysis and then dividing that rank by the total number of variants in the region. A concordance rate < 0.05 suggested that explanation 3 was unlikely. To further evaluate explanation 3, we also used the JLIM approach.[27] Although JLIM doesn't specify the likely causal variant at a region, it can be used to examine whether the underlying causal variation is responsible for the observed effects on both methylation and cardiovascular traits in a two-sample framework. Prior probabilities were not integrated into these analyses with FINEMAP, which allowed for a more direct comparison with results of the JLIM method.

Impact of mQTLs on Gene Expression and Histone Modification

We applied 2SMR to evaluate the relationship between methylation and expression by using observed effects between SNPs and expression in relevant tissue types from the Genotype-Tissue Expression (GTEx) Consortium.[48] When observed effects for sentinel mQTLs were not available from GTEx, we identified a surrogate SNP instead ($r^2 > 0.8$).

We also assessed whether any mQTLs were in LD ($r^2 > 0.8$) with any previously reported histone quantitative trait loci (hQTLs).[49] When this was true, we applied 2SMR to evaluate the causal relationship between methylation and histone modification at these loci. This analysis was for exploratory purposes because some aspects of the relationship between DNA methylation and histone modification remain unexplored, despite progress by recent studies.[50,51]

Functional Informatics

We applied the Variant Effect Predictor (VEP)[52] to the top-ranked mQTLs from the bivariate fine-mapping analysis to calculate their predicted consequence. We obtained enhancer annotations from the Illumina 450K annotation file and additional regulatory data from Ensembl[53] to evaluate whether mQTLs and CpG sites were located within regulatory regions of the genome. Because we were interested in cardiovascular and lipid traits in this study, we used tissue-specific data from the Roadmap Epigenomics Project[54] to infer whether the potential causal variants and CpG sites at each locus resided within histone mark peaks and regions of DNase hypersensitivity. These tissues were adipose-derived mesenchymal stem cells, adipose nuclei, aorta, fetal heart, left ventricle, right atrium, and right ventricle, which we selected because of their biological relevance in cardiovascular etiology.

We performed enrichment analysis to test whether lead SNPs and associated CpG sites were located in regulatory regions more than can be accounted for by chance. To calibrate background

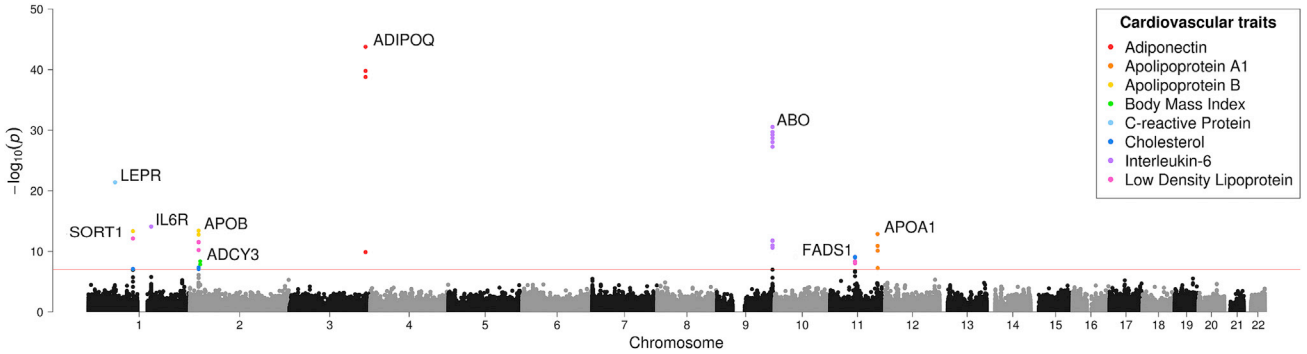

Figure 3. Manhattan Plot Illustrating Observed Association between mQTLs and Cardiovascular Traits
Points represent –log10 p values (y axis) for genetic variants according to their genomic location (x axis). Effects that survived the multiple-testing threshold in our analysis ($p < 9.45 \times 10^{-8}$ – represented by the red horizontal line) are colored according to their associated trait and annotated according to the most likely affected gene.

expectations, we obtained matched SNPs by using snpSNAP[55] and identified matched CpG sites by randomly sampling 450K array probes that were in similar regions across the genome (i.e., within CpG islands or first exons, etc.). We investigated enrichment by using the hypergeometric test and accounted for multiple testing for by randomly sampling control SNPs and probes and re-running analyses for 10,000 iterations.

Results

Mining for Putative Causal Influences of Methylation on Cardiovascular Traits

We undertook 529,368 tests to evaluate the association between previously identified mQTLs in ARIES with each trait in turn (37,812 unique variants × 14 traits). We identified ten independent association signals, which, after multiple-testing correction, provided strong evidence of association ($p < 9.45 \times 10^{-8}$ [i.e., 0.05/529,368]); these can be found in Table 1 and Figure 3. Two of these effects were observed at the same CpG site near *ADIPOQ* (MIM: 612556), although they were identified with two independent mQTLs ($r^2 = 0.02$).

The ten sentinel mQTLs identified in this analysis were strongly associated with DNA methylation only at a proximal CpG site and not any other CpG sites in the epigenome, according to our findings in ARIES. A summary of these mQTLs can be found in Table S1. We repeated our analysis with adjustment for the first ten principal components, although results did not suggest that population stratification was an issue in this analysis (Table S2).

Inferring Putative Causal Relationships

We obtained estimates of putative causal effects between methylation and cardiovascular traits at each locus in the MR analysis by using mQTLs as our instrumental variables (Table 2). Effect estimates suggested a direct relationship between methylation and cardiovascular traits at the *IL6R* (MIM: 147880), *APOB* (MIM: 107730), *SORT1* (MIM: 602458), and *ADCY3* (MIM: 600291) loci (i.e., increased methylation results in an observed increase in the cardiovascular trait), whereas an inverse relationship was observed at the *ADIPOQ*, *ABO* (MIM: 110300), *LEPR* (MIM: 601007), *APOA1* (MIM: 107680), and *FADS1* (MIM: 606148) loci (i.e., increased methylation causes a decrease in cardiovascular-trait levels). Because two independent mQTLs were contributing to methylation at *ADIPOQ*, we undertook multivariate MR, which provided strong evidence of an inverse relationship between methylation and adiponectin at this locus (−0.548 ng/mL per standard deviation change in methylation levels, standard error = 0.107, $p = 3.79 \times 10^{-7}$).

Taking these putative associations forward, we evaluated the potential for reverse causal relationships by performing MR of the cardiovascular traits against the DNA methylation levels by using SNPs from GWASs as our instruments. There was no evidence to suggest that the putative associations were due to the cardiovascular traits influencing the methylation levels (Table S3), and therefore these effects cannot be attributed to explanation 2. We note, however, that statistical power to detect an effect in this direction is low.

Using methylation data from two other time points across the life course (at birth and adolescence [mean age: 17.1 years]), we observed directions of effect consistent with those observed with data from the childhood time point (mean age: 7.5 years) (Tables S4 and S5). We observed evidence of association at each locus in this analysis except when we used cord data for the *ABO* and *IL6R* loci. We reproduced similar effects for nine of the ten mQTLs on cardiovascular traits by using effect estimates from published studies (Table 3). The only locus for which we were not able to find a replication effect estimate was the mQTL at *IL6R*, because it was not in LD ($r^2 > 0.8$) with any previously published findings for IL-6.

Evaluating Putative Causal Variants to Infer Mediated Effects

There was concordance among the top SNPs in the bivariate fine-mapping analyses for IL-6 (*ABO* locus), BMI

Table 2. Results of MR Analysis between DNA Methylation and Traits

SNP	Gene	CpG	Trait	Sample Size	Beta	SE	p Value
rs266772	*ADIPOQ*	cg05578595	adiponectin (ng/mL)	646	−0.846	0.168	5.93×10^{-7}
rs687621	*ABO*	cg21160290	IL-6 (pg/mL)	646	−0.293	0.061	1.77×10^{-6}
rs13375019	*LEPR*	cg04111102	CRP (mg/L)	646	−0.265	0.076	0.001
rs7549250	*IL6R*	cg02856953	IL-6 (pg/mL)	646	0.468	0.175	0.008
rs169109	*ADIPOQ*	cg05578595	adiponectin (ng/mL)	646	−0.363	0.121	0.003
rs541041	*APOB*	cg25035485	Apo B (g/L)	646	0.298	0.114	0.009
rs7528419	*SORT1*	cg00908766	Apo B (g/L)	646	0.271	0.064	2.74×10^{-5}
rs625145	*APOA1*	cg04087571	Apo A1 (g/L)	646	−0.301	0.082	2.68×10^{-4}
rs174544	*FADS1*	cg19610905	total cholesterol (mmol/L)	646	−0.363	0.121	0.003
rs6749422	*ADCY3*	cg01884057	BMI	846	0.106	0.048	0.028

Abbreviations are as follows: SNP, single-nucleotide polymorphism; gene, most likely affected gene; CpG, 450K probe ID; trait, associated trait; sample size, sample size for this effect; beta, observed effect size (units in standard deviations); SE, standard error of the effect size; and p value, p value for observed effect.

(*ADCY3* locus), and adiponectin (*ADIPOQ* common locus), given that the variant with the largest Bayes factor was the same for the effect on DNA methylation and outcome trait (Tables S6). These results lend support to the hypothesis that DNA methylation resides on the causal pathway between genetic variants and outcome traits (i.e., explanation 1). There was a lack of concordance for the results for adiponectin (*ADIPOQ* low-frequency locus), Apo B (*SORT1* locus), total cholesterol (*FADS1* locus), and CRP (*LEPR* locus), suggesting that the mQTLs might be in LD with the putative causal variant for the phenotypic trait (i.e., explanation 3). Results of the JLIM method supported evidence at the *ADIPOQ* and *ADCY3* loci, although we were unable to further evaluate signals at the *ABO* and *IL6R* regions because GWAS summary results were unavailable for IL-6 (Table S7). Posterior probabilities from FINEMAP suggested that there was most likely only a single variant influencing trait variation for each observed effect (Table S7).

Investigating the Role of DNA Methylation in Gene Expression and Histone Modification

To further dissect the relationship between DNA methylation and complex traits, we sought to evaluate the influence of the methylation levels on local gene expression. Using data from the GTEx Consortium, we observed evidence of a causal relationship between methylation and expression at eight of the ten loci (Table 4). Effect estimates suggest an inverse relationship (i.e., increased methylation results in decreased gene expression) at the *ADIPOQ* (low-frequency signal) and *APOA1* loci, whereas a direct relationship was observed at the other six loci (i.e., increased methylation results in increased gene expression). We were unable to identify a surrogate variant ($r^2 > 0.8$) to obtain a suitable effect estimate at the *LEPR* and *ADIPOQ* (common signal) loci.

mQTLs at the *APOA1* and *IL6R* loci were also in high LD with previously reported hQTLs according to findings by Grubert et al.[49] Results from our 2SMR analyses to evaluate the influence of methylation levels on histone modification provided strong evidence of a causal effect as well as an inverse relationship in each instance (Table S8).

Functional Informatics

To better understand the functional role underlying these putative causal associations, we evaluated variants and CpG sites to discern whether they reside within regulatory regions across the genome. An overview of the regulatory data used can be found in Table S9. In this analysis, we used the lead variants based on the bivariate fine-mapping analysis (using effect estimates on DNA methylation) and used the VEP to predict their functional consequences (Table S10).

Every associated CpG site identified in this study resides within multiple histone mark peaks according to tissue data from the Roadmap Epigenomics Project (Table S11). All sites, with the exception of the CpG site near *ADIPOQ*, also reside in either enhancer, promoter, or promoter flanking regions. There was strong evidence of enrichment of regulatory annotations for both SNPs and CpG sites, which supports previous evidence that they are likely to have a causal downstream effect on phenotypic variation (Table S12).

Discussion

We have designed a framework for evaluating the putative causal influences of DNA methylation on complex traits and disease via MR. For observed effects on cardiovascular traits that appear to be caused by methylation, we used bivariate fine mapping and JLIM to evaluate whether the putative causal variant influencing methylation was the same causal variant responsible for influencing the trait. The bivariate fine mapping suggested that cardiovascular traits might be influenced by altered DNA methylation levels at the *ABO*, *ADCY3*, *ADIPOQ*, *APOA1*, *APOB*, and

Table 3. Results of Replication Analysis via Two-Sample MR

SNP	Gene	Trait	CpG	CpG Effect (SE)	Trait Effect (SE)	2SMR Effect (SE)	p Value	Study
rs266772	*ADIPOQ*	Adiponectin (ng/mL)	cg05578595	0.982 (0.103)	−0.629 (0.143)	−0.641 (0.160)	6.50×10^{-5}	UK10K Consortium (TwinsUK individuals only)[35]
rs687621	*ABO*	IL-6 (pg/mL)	cg21160290	0.912 (0.036)	−0.245 (0.026)	−0.269 (0.03)	9.16×10^{-19}	Naitza et al.[46]
rs2211651*	*LEPR*	CRP (mg/L)	cg04111102	0.682 (0.036)	−0.170 (0.022)	−0.249 (0.035)	3.09×10^{-13}	Reiner et al.[56]
rs541041	*APOB*	Apo B (g/L)	cg25035485	0.627 (0.053)	0.098 (0.013)	0.156 (0.025)	2.05×10^{-10}	Kettunen et al.[57]
rs169109	*ADIPOQ*	Adiponectin (ng/mL)	cg05578595	0.383 (0.036)	−0.052 (0.005)	−0.136 (0.017)	2.58×10^{-15}	Dastani et al.[58]
rs7528419	*SORT1*	Apo B (g/L)	cg00908766	−0.980 (0.037)	−0.089 (0.012)	0.091 (0.013)	9.20×10^{-13}	Kettunen et al.[57]
rs625145	*APOA1*	Apo A1 (g/L)	cg04087571	−0.884 (0.044)	0.057 (0.013)	−0.064 (0.015)	1.84×10^{-5}	Kettunen et al.[57]
rs174544	*FADS1*	total cholesterol (mmol/L)	cg19610905	−0.655 (0.031)	0.047 (0.004)	−0.072 (0.007)	9.73×10^{-25}	Willer et al.[59]
rs6749422	*ADCY3*	BMI	cg01884057	0.908 (0.026)	0.068 (0.007)	0.075 (0.008)	8.05×10^{-21}	Felix et al.[60]

Abbreviations are as follows: SNP, single-nucleotide polymorphism; gene, most likely affected gene; trait, associated trait; CpG, 450K probe ID; CpG effect, effect estimate of SNP on methylation; trait effect, effect estimate of SNP on trait; 2SMR effect, effect estimates from two-sample MR analysis; p value, p value for observed effect; study, published study where effect estimates for traits were obtained; and SE, standard error. The asterisk indicates that a surrogate variant was used ($r^2 > 0.8$).

IL6R regions. However, JLIM supported findings only at the *ADCY3* and *ADIPOQ* loci. This provides compelling evidence that DNA methylation might play a mediatory role for the effects at these loci. 2SMR analyses provided evidence that DNA methylation levels influenced gene expression at these loci, suggesting that functional effects for the causal variants induce a coordinated system of effects. This was important to demonstrate, given that having only single valid instruments available for CpGs meant that we were unable to robustly show that variants were not influencing methylation and traits through horizontal pleiotropy. This limitation has also been encountered by other attempts to evaluate the relationship between DNA methylation and complex traits.[16] Nevertheless, the ability to indicate putative mediating molecular phenotypes between genetic factors and complex traits is particularly attractive for therapeutic evaluation of drug targets.

The *ABO* locus identified in this study has been associated with many different traits and diseases by previous studies,[25,61,62] and there is also evidence implicating eQTLs as putative causal SNPs for this effect.[63] Here, we provide evidence that DNA methylation might reside along the causal pathway to these observed effects (MR effect estimate: 0.29 [standard error = 0.06] change in trait per standard deviation change in methylation), although its widespread effect also raises the possibility of horizontal pleiotropy. A deletion (rs200533593) was found to be the putative causal variant for both the observed effect on DNA methylation and phenotypic variation.

The observed effect of genetic variation at *ADCY3* on BMI is a relatively recent finding.[60,64,65] In this study, our bivariate fine-mapping analysis suggests that an intergenic variant (rs6737082) might be responsible for the observed signal that is mediated through DNA methylation at this locus (MR effect estimate: 0.11 [0.05]). Furthermore, a variant in LD with rs6737082 (rs713586, $r^2 = 0.80$) has been previously reported to regulate DNA methylation at this location in adipose tissue.[7]

Two independent effects associated with adiponectin were detected near *ADIPOQ* in our study. The common variant signal was located upstream of *ADIPOQ* within *RFC4* but associated with DNA methylation levels proximal to *ADIPOQ*, which can help explain this variant's observed effect on adiponectin (MR effect estimate: −0.36 [0.12]). Concordance in the bivariate fine-mapping analysis suggested that a non-coding transcript variant (rs169109) was responsible. The lead SNP from the ADIPOGen Consortium[66] at this locus (rs6810075) is neither an mQTL nor in high LD with rs169109 ($r^2 = 0.20$), suggesting that these two association signals influence adiponectin levels by alternative biological mechanisms. The low-frequency variant signal was previously detected by the UK10K project,[35] although bivariate fine-mapping results at this locus suggest that the causal mQTL was in LD with the trait-associated variant.

The CpG site associated with Apo A1resides between *APOA1* and *APOA1-antisense* (*APOA1-AS*), a negative transcriptional regulator of *APOA1* that has been shown to increase *APOA1* expression both *in vitro* and *in vivo*.[67] The highest ranked mQTL according to our bivariate fine mapping using estimates with DNA methylation is in a promoter region upstream of *APOA1*, suggesting that it might be more likely to influence *APOA1* than *APOA1-AS*. GWAS association signals for lipid traits have been previously reported at this locus.[68,69] However, given the evidence in this study of a causal effect with DNA methylation (MR effect estimate: −0.30 [0.08] g/L per SD methylation level), it is possible that these are downstream effects of the observed effect on Apo A1 variation.

Table 4. Results of Analysis Investigating Causal Relationship between Methylation and Expression via Two-Sample MR

SNP	Gene	CpG	CpG Effect (SE)	eQTL Effect (SE)	eQTL p Value	eQTL Tissue	2SMR (SE)	p Value
rs116552240*	*ABO*	cg21160290	0.912 (0.036)	0.548 (0.069)	1.316×10^{-13}	adipose	0.601 (0.079)	3.28×10^{-14}
rs6737082	*ADCY3*	cg01884057	0.908 (0.026)	0.208 (0.047)	1.456×10^{-5}	adipose	0.229 (0.052)	1.13×10^{-5}
rs266772	*ADIPOQ*	cg05578595	0.982 (0.103)	−0.339 (0.078)	1.893×10^{-5}	adipose	−0.345 (0.087)	7.67×10^{-5}
rs688456	*APOA1*	cg04087571	−0.884 (0.044)	0.420 (0.095)	1.789×10^{-5}	heart	−0.475 (0.11)	1.58×10^{-5}
rs541041	*APOB*	cg25035485	−0.627 (0.053)	−0.370 (0.066)	6.326×10^{-8}	heart	0.590 (0.116)	4.06×10^{-7}
rs646776	*SORT1*	cg00908766	−0.980 (0.037)	−1.240 (0.105)	1.556×10^{-20}	liver	1.265 (0.117)	4.01×10^{-27}
rs174559	*FADS1*	cg19610905	−0.655 (0.031)	−0.707 (0.089)	5.629×10^{-13}	pancreas	1.079 (0.145)	1.04×10^{-13}
rs10908837	*IL6R*	cg02856953	−0.303 (0.039)	−0.120 (0.020)	4.171×10^{-9}	whole blood	0.396 (0.083)	2.05×10^{-6}

Abbreviations are as follows: SNP, single-nucleotide polymorphism; gene, most likely affected gene; CpG, 450K probe ID; CpG effect, effect estimate of SNP on methylation; eQTL effect, effect estimate of SNP on expression according to GTEx data; eQTL p, p value for eQTL from GTEx; eQTL tissue, tissue type for observed effect according to GTEx; 2SMR effect, effect estimates from two-sample MR analysis (standard deviation units per standard deviation units); p value, p value for 2SMR effect; and SE, standard error. The asterisk indicates that a surrogate variant was used ($r^2 > 0.8$).

The signal at the *IL6R* locus influencing IL-6 has been previously associated with a range of traits related to respiratory and cardiovascular health.[70–72] Our results suggest that genetic variation at *IL6R* influences DNA methylation at this region, which in turn could have a downstream effect on the amount of IL-6 (MR effect estimate: 0.47 [0.18] pg/mL per standard deviation methylation level). Furthermore, this association signal was not in LD with a previously reported missense variant at this locus (rs2228145, $r^2 = 0.47$ in ALSPAC), which was also supported by findings from an in-depth functional study of this variant.[73]

Evidence from the GTEx Consortium suggests that mQTLs at all eight of the loci with available expression data overlap eQTL effects, which serves as a form of independent replication of the relationships discovered through DNA methylation levels. It is biologically plausible that a variant's impact on DNA methylation levels might have a downstream effect on gene expression along the causal pathway to disease,[74,75] which could help explain these observations. Effects at four loci in particular appear to be biologically plausible in this regard, as the likely genes influenced by these variants are involved in the protein synthesis of the associated trait (i.e., *ADIPOQ* with adiponectin, *APOB* with Apo B, *APOA-I* with Apo A1, and *IL6R* with IL-6). Furthermore, each CpG site identified in this study resides within histone mark peaks in adipose tissue according to data from the Roadmap Epigenomics project. There was evidence of enrichment for these observations in comparison to background CpG sites which are located in similar genomic regions.

As with any study that applies single-instrument MR to investigate causal relationships in epidemiology, an important limitation is the inability to disentangle potential horizontal pleiotropic effects, where the same causal variant influences both exposure (i.e., DNA methylation) and outcome (i.e., cardiovascular trait) through independent pathways. To reduce the possibility of this, we selected mQTLs that were influencing only proximal CpG sites and not others in the epigenome, given that *trans*-mQTLs would be more prone to influence traits via alternative biological mechanisms. Although ARIES includes CpG sites that have two or three independent instruments (such as the CpG site at *ADIPOQ* in this study), distinguishing mediation from pleiotropy at these loci remains a challenging endeavor. Future studies that continue to uncover multiple mQTLs per CpG across the genome (as well as across various tissue types) should facilitate analyses that are able to more reliably address concerns of pleiotropy by using methods such as MR-Egger[76] and median- and mode-based MR estimators.[77,78] These findings should also facilitate analyses that model the joint effects of multiple causal mQTLs at loci across the genome rather than evaluate mQTL effects independently of each other, as we did in this study.

Weak instrumental variables and reverse causation are other factors that can bias MR analyses. Our analysis is unlikely to have suffered from the former because each mQTL had a large effect on DNA methylation in *cis* ($p < 1.0 \times 10^{-14}$) and was robustly associated with traits that we were able to replicate by using results from studies with large population samples. We conducted analyses to evaluate whether reverse causation was an issue in our study (i.e., trait variation caused changes in DNA methylation at each locus). Although our results suggest that this was not the case, it is important to note that the statistical power to detect causal effects in this direction is low because the sample size available for the SNP effects on CpG levels was small.

In this study, we demonstrated the value of 2SMR to MR analyses using summary statistics.[44,79] This allowed us to provide evidence of replication for the observed effects in our study as well as investigate the relationship between DNA methylation and expression along the causal pathway to disease. This approach has the attractive advantage of enabling the interrogation of the potential epigenetic-complex trait interplay on a much wider scale by foregoing the requirement that "omic" data and phenotypes are measured in the same sample.

Supplemental Data

Supplemental Data include 12 tables and can be found with this article online at https://doi.org/10.1016/j.ajhg.2017.09.003.

Acknowledgments

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole Avon Longitudinal Study of Parents and Children (ALSPAC) team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council and the Wellcome Trust (grant 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data were generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) with support from 23andMe. Methylation data in the ALSPAC cohort were generated as part of the UK BBSRC-funded (BB/I025751/1 and BB/I025263/1) Accessible Resource for Integrated Epigenomic Studies (ARIES). This publication is the work of the authors, and T.G.R. will serve as guarantor for the contents of this paper. This work was supported by the UK Medical Research Council (MRC Integrative Epidemiology Unit) (MC UU 12013/1, MC UU 12013/2, MC UU 12013/3, and MC UU 12013/8). T.G.R. is supported by the Elizabeth Blackwell Institute Proximity to Discovery award (EBI 424).

Received: April 29, 2017
Accepted: September 6, 2017
Published: October 5, 2017

Web Resources

ALSPAC, http://www.bristol.ac.uk/alspac/
ALSPAC data dictionary, http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/
ARIES Explorer, http://www.ariesepigenomics.org.uk/ariesexplorer
Ensembl, http://www.ensembl.org/index.html
GTEx Portal, http://www.gtexportal.org/home/
GWAS Catalog, http://www.ebi.ac.uk/gwas/
Multicolored Manhattan plot script, https://github.com/MRCIEU/qqman_multiple_colours
mQTL Database, http://www.mqtldb.org/
OMIM, http://www.omim.org/
Roadmap Epigenomics, http://www.roadmapepigenomics.org/
SNPsnap, http://www.broadinstitute.org/mpg/snpsnap/index.html

References

1. Edwards, S.L., Beesley, J., French, J.D., and Dunning, A.M. (2013). Beyond GWASs: illuminating the dark road from association to function. Am. J. Hum. Genet. *93*, 779–797.
2. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487.
3. Burkhardt, R., Kirsten, H., Beutner, F., Holdt, L.M., Gross, A., Teren, A., Tönjes, A., Becker, S., Krohn, K., Kovacs, P., et al. (2015). Integration of genome-wide SNP data and gene-expression profiles reveals six novel loci and regulatory mechanisms for amino acids and acylcarnitines in whole blood. PLoS Genet. *11*, e1005510.
4. Pavlides, J.M., Zhu, Z., Gratten, J., McRae, A.F., Wray, N.R., and Yang, J. (2016). Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. Genome Med. *8*, 84.
5. Mancuso, N., Shi, H., Goddard, P., Kichaev, G., Gusev, A., and Pasaniuc, B. (2017). Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. Am. J. Hum. Genet. *100*, 473–487.
6. Kulis, M., Heath, S., Bibikova, M., Queirós, A.C., Navarro, A., Clot, G., Martínez-Trillos, A., Castellano, G., Brun-Heath, I., Pinyol, M., et al. (2012). Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat. Genet. *44*, 1236–1242.
7. Grundberg, E., Meduri, E., Sandling, J.K., Hedman, A.K., Keildson, S., Buil, A., Busche, S., Yuan, W., Nisbet, J., Sekowska, M., et al.; Multiple Tissue Human Expression Resource Consortium (2013). Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am. J. Hum. Genet. *93*, 876–890.
8. Gaunt, T.R., Shihab, H.A., Hemani, G., Min, J.L., Woodward, G., Lyttleton, O., Zheng, J., Duggirala, A., McArdle, W.L., Ho, K., et al. (2016). Systematic identification of genetic influences on methylation across the human life course. Genome Biol. *17*, 61.
9. Shi, J., Marconett, C.N., Duan, J., Hyland, P.L., Li, P., Wang, Z., Wheeler, W., Zhou, B., Campan, M., Lee, D.S., et al. (2014). Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. Nat. Commun. *5*, 3365.
10. Bell, J.T., Tsai, P.C., Yang, T.P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., et al.; MuTHER Consortium (2012). Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS Genet. *8*, e1002629.
11. 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.
12. Liang, L., Willis-Owen, S.A.G., Laprise, C., Wong, K.C.C., Davies, G.A., Hudson, T.J., Binia, A., Hopkin, J.M., Yang, I.V., Grundberg, E., et al. (2015). An epigenome-wide association study of total serum immunoglobulin E concentration. Nature *520*, 670–674.
13. Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W., Jansen, R., de Geus, E.J., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.
14. Powell, J.E., Fung, J.N., Shakhbazov, K., Sapkota, Y., Cloonan, N., Hemani, G., Hillman, K.M., Kaufmann, S., Luong, H.T., Bowdler, L., et al. (2016). Endometriosis risk alleles at 1p36.12 act through inverse regulation of CDC42 and LINC00339. Hum. Mol. Genet. *25*, 5046–5058.
15. Rawlik, K., Rowlatt, A., and Tenesa, A. (2016). Imputation of DNA methylation levels in the brain implicates a risk factor for Parkinson's disease. Genetics *204*, 771–781.
16. Hannon, E., Weedon, M., Bray, N., O'Donovan, M., and Mill, J. (2017). Pleiotropic effects of trait-associated genetic variation on DNA methylation: utility for refining GWAS loci. Am. J. Hum. Genet. *100*, 954–959.

17. Davey Smith, G., and Hemani, G. (2014). Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum. Mol. Genet. *23* (R1), R89–R98.
18. Davey Smith, G., and Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int. J. Epidemiol. *32*, 1–22.
19. Relton, C.L., and Davey Smith, G. (2012). Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int. J. Epidemiol. *41*, 161–176.
20. Millstein, J., Zhang, B., Zhu, J., and Schadt, E.E. (2009). Disentangling molecular relationships with a causal inference test. BMC Genet. *10*, 23.
21. Richmond, R.C., Hemani, G., Tilling, K., Davey Smith, G., and Relton, C.L. (2016). Challenges and novel approaches for investigating molecular mediation. Hum. Mol. Genet. *25* (R2), R149–R156.
22. Hodgkin, J. (1998). Seven types of pleiotropy. Int. J. Dev. Biol. *42*, 501–505.
23. Ference, B.A., Yoo, W., Alesh, I., Mahajan, N., Mirowska, K.K., Mewada, A., Kahn, J., Afonso, L., Williams, K.A., Sr., and Flack, J.M. (2012). Effect of long-term exposure to lower low-density lipoprotein cholesterol beginning early in life on the risk of coronary heart disease: a Mendelian randomization analysis. J. Am. Coll. Cardiol. *60*, 2631–2639.
24. Vimaleswaran, K.S., Berry, D.J., Lu, C., Tikkanen, E., Pilz, S., Hiraki, L.T., Cooper, J.D., Dastani, Z., Li, R., Houston, D.K., et al.; Genetic Investigation of Anthropometric Traits-GIANT Consortium (2013). Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. PLoS Med. *10*, e1001383.
25. Pickrell, J.K., Berisa, T., Liu, J.Z., Ségurel, L., Tung, J.Y., and Hinds, D.A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. *48*, 709–717.
26. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383.
27. Chun, S., Casparino, A., Patsopoulos, N.A., Croteau-Chonka, D.C., Raby, B.A., De Jager, P.L., Sunyaev, S.R., and Cotsapas, C. (2017). Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. Nat. Genet. *49*, 600–605.
28. Hemani, G., Tilling, K., and Davey Smith, G. (2017). Orienting the causal relationship between imprecisely measured traits using genetic instruments. bioRxiv. http://dx.doi.org/10.1101/117101.
29. Boyd, A., Golding, J., Macleod, J., Lawlor, D.A., Fraser, A., Henderson, J., Molloy, L., Ness, A., Ring, S., and Davey Smith, G. (2013). Cohort Profile: the 'children of the 90s'–the index offspring of the Avon Longitudinal Study of Parents and Children. Int. J. Epidemiol. *42*, 111–127.
30. Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., Henderson, J., Macleod, J., Molloy, L., Ness, A., et al. (2013). Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int. J. Epidemiol. *42*, 97–110.
31. Relton, C.L., Gaunt, T., McArdle, W., Ho, K., Duggirala, A., Shihab, H., Woodward, G., Lyttleton, O., Evans, D.M., Reik, W., et al. (2015). Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). Int. J. Epidemiol. *44*, 1181–1190.
32. Touleimat, N., and Tost, J. (2012). Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. Epigenomics *4*, 325–341.
33. Pidsley, R., Y Wong, C.C., Volta, M., Lunnon, K., Mill, J., and Schalkwyk, L.C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics *14*, 293.
34. Jaffe, A.E., and Irizarry, R.A. (2014). Accounting for cellular heterogeneity is critical in epigenome-wide association studies. Genome Biol. *15*, R31.
35. Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R., Xu, C., Futema, M., Lawson, D., et al.; UK10K Consortium (2015). The UK10K project identifies rare variants in health and disease. Nature *526*, 82–90.
36. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
37. Myers, G.L., Cooper, G.R., Greenberg, N., et al. (2000). Standardization of lipid and lipoprotein measurements. In Handbook of Lipoprotein Testing, Second Edition, N. Rifai, G.R. Warnick, and M.H. Dominiczak, eds. (AACC Press), pp. 717–748.
38. Warnick, G.R., Knopp, R.H., Fitzpatrick, V., and Branson, L. (1990). Estimating low-density lipoprotein cholesterol by the Friedewald equation is adequate for classifying patients on the basis of nationally recommended cutpoints. Clin. Chem. *36*, 15–19.
39. Wallace, A.M., McMahon, A.D., Packard, C.J., Kelly, A., Shepherd, J., Gaw, A., and Sattar, N. (2001). Plasma leptin and the risk of cardiovascular disease in the west of Scotland coronary prevention study (WOSCOPS). Circulation *104*, 3052–3056.
40. Naeem, H., Wong, N.C., Chatterton, Z., Hong, M.K., Pedersen, J.S., Corcoran, N.M., Hovens, C.M., and Macintyre, G. (2014). Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. BMC Genomics *15*, 51.
41. St Pourcain, B., Whitehouse, A.J., Ang, W.Q., Warrington, N.M., Glessner, J.T., Wang, K., Timpson, N.J., Evans, D.M., Kemp, J.P., Ring, S.M., et al. (2013). Common variation contributes to the genetic architecture of social communication traits. Mol. Autism *4*, 34.
42. Turner, S.D. (2014). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. bioRxiv. http://dx.doi.org/10.1101/005165.
43. Henningsen, A., and Hamann, J.D. (2007). systemfit: a package for estimating systems of simultaneous equations in R. J. Stat. Softw. *23*, 1–40.
44. Burgess, S., Scott, R.A., Timpson, N.J., Davey Smith, G., Thompson, S.G.; and EPIC- InterAct Consortium (2015). Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. Eur. J. Epidemiol. *30*, 543–552.
45. MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. *45* (D1), D896–D901.

46. Naitza, S., Porcu, E., Steri, M., Taub, D.D., Mulas, A., Xiao, X., Strait, J., Dei, M., Lai, S., Busonero, F., et al. (2012). A genome-wide association scan on the levels of markers of inflammation in Sardinians reveals associations that underpin its complex regulation. PLoS Genet. *8*, e1002480.
47. Benner, C., Spencer, C.C., Havulinna, A.S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-wide association studies. Bioinformatics *32*, 1493–1501.
48. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.
49. Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A., et al. (2015). Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell *162*, 1051–1065.
50. Rose, N.R., and Klose, R.J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. Biochim. Biophys. Acta *1839*, 1362–1372.
51. Liu, L., Jin, G., and Zhou, X. (2015). Modeling the relationship of epigenetic modifications to transcription factor binding. Nucleic Acids Res. *43*, 3873–3885.
52. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl variant effect predictor. Genome Biol. *17*, 122.
53. Yates, A., Akanni, W., Amode, M.R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., et al. (2016). Ensembl 2016. Nucleic Acids Res. *44* (D1), D710–D716.
54. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.
55. Pers, T.H., Timshel, P., and Hirschhorn, J.N. (2015). SNPsnap: a Web-based tool for identification and annotation of matched SNPs. Bioinformatics *31*, 418–420.
56. Reiner, A.P., Beleza, S., Franceschini, N., Auer, P.L., Robinson, J.G., Kooperberg, C., Peters, U., and Tang, H. (2012). Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. Am. J. Hum. Genet. *91*, 502–512.
57. Kettunen, J., Demirkan, A., Würtz, P., Draisma, H.H., Haller, T., Rawal, R., Vaarhorst, A., Kangas, A.J., Lyytikäinen, L.P., Pirinen, M., et al. (2016). Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat. Commun. *7*, 11122.
58. Dastani, Z., Johnson, T., Kronenberg, F., Nelson, C.P., Assimes, T.L., März, W., Richards, J.B.; CARDIoGRAM Consortium; and ADIPOGen Consortium (2013). The shared allelic architecture of adiponectin levels and coronary artery disease. Atherosclerosis *229*, 145–148.
59. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al.; Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.
60. Felix, J.F., Bradfield, J.P., Monnereau, C., van der Valk, R.J., Stergiakouli, E., Chesi, A., Gaillard, R., Feenstra, B., Thiering, E., Kreiner-Møller, E., et al.; Bone Mineral Density in Childhood Study (BMDCS); Early Genetics and Lifecourse Epidemiology (EAGLE) consortium; Early Growth Genetics (EGG) Consortium; and Bone Mineral Density in Childhood Study BMDCS (2016). Genome-wide association analysis identifies three new susceptibility loci for childhood body mass index. Hum. Mol. Genet. *25*, 389–403.
61. Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat. Genet. *47*, 1121–1130.
62. Hinds, D.A., Buil, A., Ziemek, D., Martinez-Perez, A., Malik, R., Folkersen, L., Germain, M., Mälarstig, A., Brown, A., Soria, J.M., et al.; METASTROKE Consortium, INVENT Consortium (2016). Genome-wide association analysis of self-reported events in 6135 individuals and 252 827 controls identifies 8 loci associated with thrombosis. Hum. Mol. Genet. *25*, 1867–1874.
63. Wessel, J., Chu, A.Y., Willems, S.M., Wang, S., Yaghootkar, H., Brody, J.A., Dauriz, M., Hivert, M.F., Raghavan, S., Lipovich, L., et al.; EPIC-InterAct Consortium (2015). Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. Nat. Commun. *6*, 5897.
64. Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., et al.; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; and International Endogene Consortium (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197–206.
65. Warrington, N.M., Howe, L.D., Paternoster, L., Kaakinen, M., Herrala, S., Huikari, V., Wu, Y.Y., Kemp, J.P., Timpson, N.J., St Pourcain, B., et al. (2015). A genome-wide association study of body mass index across early life and childhood. Int. J. Epidemiol. *44*, 700–712.
66. Dastani, Z., Hivert, M.F., Timpson, N., Perry, J.R., Yuan, X., Scott, R.A., Henneman, P., Heid, I.M., Kizer, J.R., Lyytikäinen, L.P., et al.; DIAGRAM+ Consortium; MAGIC Consortium; GLGC Investigators; MuTHER Consortium; DIAGRAM Consortium; GIANT Consortium; Global B Pgen Consortium; Procardis Consortium; MAGIC investigators; and GLGC Consortium (2012). Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multi-ethnic meta-analysis of 45,891 individuals. PLoS Genet. *8*, e1002607.
67. Halley, P., Kadakkuzha, B.M., Faghihi, M.A., Magistri, M., Zeier, Z., Khorkova, O., Coito, C., Hsiao, J., Lawrence, M., and Wahlestedt, C. (2014). Regulation of the apolipoprotein gene cluster by a long noncoding RNA. Cell Rep. *6*, 222–230.
68. Lu, X., Huang, J., Mo, Z., He, J., Wang, L., Yang, X., Tan, A., Chen, S., Chen, J., Gu, C.C., et al. (2016). Genetic susceptibility to lipid levels and lipid change over time and risk of incident hyperlipidemia in Chinese populations. Circ Cardiovasc Genet *9*, 37–44.
69. Kurano, M., Tsukamoto, K., Kamitsuji, S., Kamatani, N., Hara, M., Ishikawa, T., Kim, B.J., Moon, S., Jin Kim, Y., and Teramoto, T. (2016). Genome-wide association study of serum lipids confirms previously reported associations as well as new associations of common SNPs within PCSK7 gene with triglyceride. J. Hum. Genet. *61*, 427–433.

70. Ferreira, M.A., Matheson, M.C., Duffy, D.L., Marks, G.B., Hui, J., Le Souëf, P., Danoy, P., Baltic, S., Nyholt, D.R., Jenkins, M., et al.; Australian Asthma Genetics Consortium (2011). Identification of IL6R and chromosome 11q13.5 as risk loci for asthma. Lancet *378*, 1006–1014.
71. Dehghan, A., Dupuis, J., Barbalic, M., Bis, J.C., Eiriksdottir, G., Lu, C., Pellikka, N., Wallaschofski, H., Kettunen, J., Henneman, P., et al. (2011). Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation *123*, 731–738.
72. Khera, A.V., Emdin, C.A., Drake, I., Natarajan, P., Bick, A.G., Cook, N.R., Chasman, D.I., Baber, U., Mehran, R., Rader, D.J., et al. (2016). Genetic risk, adherence to a healthy lifestyle, and coronary disease. N. Engl. J. Med. *375*, 2349–2358.
73. van Dongen, J., Jansen, R., Smit, D., Hottenga, J.J., Mbarek, H., Willemsen, G., Kluft, C., Penninx, B.W., Ferreira, M.A., Boomsma, D.I., de Geus, E.J.; and AAGC Collaborators (2014). The contribution of the functional IL6R polymorphism rs2228145, eQTLs and other genome-wide SNPs to the heritability of plasma sIL-6R levels. Behav. Genet. *44*, 368–382.
74. Jones, P.A., and Takai, D. (2001). The role of DNA methylation in mammalian epigenetics. Science *293*, 1068–1070.
75. Baylin, S.B., Esteller, M., Rountree, M.R., Bachman, K.E., Schuebel, K., and Herman, J.G. (2001). Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Hum. Mol. Genet. *10*, 687–692.
76. Bowden, J., Davey Smith, G., and Burgess, S. (2015). Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int. J. Epidemiol. *44*, 512–525.
77. Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet. Epidemiol. *40*, 304–314.
78. Pires Hartwig, F., Davey Smith, G., and Bowden, J. (2017). Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. International Journal of Epidemiology. Published online July 12, 2017. http://dx.doi.org/10.1093/ije/dyx102.
79. Hemani, G., Zheng, J., Wade, K.H., Laurin, C., Elsworth, E., Burgess, S., Bowden, J., Langdon, R., Tan, V., Yarmolinsky, J., et al. (2016). MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. bioRxiv. http://dx.doi.org/10.1101/078972.

The American Journal of Human Genetics, Volume *101*

Supplemental Data

Mendelian Randomization Analysis Identifies CpG Sites as Putative Mediators for Genetic Influences on Cardiovascular Disease Risk

Tom G. Richardson, Jie Zheng, George Davey Smith, Nicholas J. Timpson, Tom R. Gaunt, Caroline L. Relton, and Gibran Hemani

SNP	Gene	Chr	BP	MAF	Other	Effect
rs266772	*ADIPOQ*	*3*	186546540	0.027	T	C
rs687621	*ABO*	*9*	136137065	0.328	G	A
rs13375019	*LEPR*	*1*	66123136	0.367	C	G
rs7549250	*IL6R*	*1*	154404336	0.416	C	T
rs169109	*ADIPOQ*	*3*	186525798	0.496	G	C
rs541041	*APOB*	*2*	21294975	0.181	G	A
rs7528419	*SORT1*	*1*	109817192	0.212	G	A
rs625145	*APOA1*	*11*	116727936	0.198	T	A
rs174544	*FADS1*	*11*	61567753	0.298	A	C
rs6749422	*ADCY3*	*2*	25150011	0.487	G	C

Table S1: Summary of methylation quantitative trait loci used as instruments in Mendelian randomization analysis SNP – Single Nucleotide Polymorphism, Chr – Chromosome, BP – Base position, MAF – minor allele frequency, Effect – effect allele coded additively for each SNP in all analyses, Other – other allele for this SNP

SNP	Gene	CpG	Trait	Sample Size	Beta	SE	P-value
rs266772	*ADIPOQ*	cg05578595	Adiponectin	4248	-0.990	0.070	5.97×10^{-44}
rs687621	*ABO*	cg21160290	Interleukin-6	4241	-0.260	0.023	9.74×10^{-30}
rs13375019	*LEPR*	cg04111102	C-reactive protein	4251	-0.220	0.022	1.14×10^{-22}
rs7549250	*IL6R*	cg02856953	Interleukin-6	4241	-0.169	0.022	1.87×10^{-14}
rs169109	*ADIPOQ*	cg05578595	Adiponectin	4248	-0.138	0.023	2.98×10^{-10}
rs541041	*APOB*	cg25035485	Apo B	4251	-0.211	0.028	4.98×10^{-14}
rs7528419	*SORT1*	cg00908766	Apo B	4251	-0.199	0.026	5.31×10^{-14}
rs625145	*APOA1*	cg04087571	Apo A1	4251	0.204	0.027	8.00×10^{-14}
rs174544	*FADS1*	cg19610905	Cholesterol	4250	-0.141	0.023	1.96×10^{-09}
rs6749422	*ADCY3*	cg01884057	Body mass index	6076	0.107	0.018	3.36×10^{-09}

Table S2: Results of linear regression analysis between genetic variants and traits adjusting for the top 10 principal components

Tables S3

SNP	CpG site	Trait	Beta	Standard Error	P-value
rs7531118_T	cg01884057	BMI	-3.350	4.758	0.482
rs7550711_T	cg01884057	BMI	2.203	2.142	0.304
rs543874_G	cg01884057	BMI	0.668	1.394	0.632
rs13021737_A	cg01884057	BMI	0.077	0.749	0.918
rs10182181_G	cg01884057	BMI	0.106	0.048	0.028
rs13078960_G	cg01884057	BMI	-0.664	1.020	0.516
rs1516725_T	cg01884057	BMI	-1.972	2.343	0.400
rs13130484_T	cg01884057	BMI	1.152	2.586	0.656
rs2112347_G	cg01884057	BMI	0.665	1.117	0.552
rs943005_T	cg01884057	BMI	-1.696	2.152	0.431
rs2183825_C	cg01884057	BMI	-10.916	235.015	0.963
rs11030104_G	cg01884057	BMI	0.353	1.745	0.840
rs3817334_T	cg01884057	BMI	0.988	1.377	0.473
rs7138803_A	cg01884057	BMI	-3.670	5.625	0.514
rs7144011_T	cg01884057	BMI	-0.089	2.330	0.969
rs13329567_T	cg01884057	BMI	-1.057	1.076	0.326
rs3888190_A	cg01884057	BMI	1.067	2.579	0.679
rs6567160_C	cg01884057	BMI	-0.486	0.781	0.534
rs11672660_T	cg01884057	BMI	5.258	21.226	0.804

Table S3a: Mendelian randomization analysis using SNPs with evidence of association with BMI from GWAS as instrumental variables to investigate reverse causation at *ADCY3*

SNP	CpG site	Trait	Beta	Standard Error	P-value
rs1108842_C	cg05578595	Adiponectin	-0.195	0.534	0.715
rs1597466_T	cg05578595	Adiponectin	3.774	9.589	0.694
rs2062632_C	cg05578595	Adiponectin	-1.319	1.352	0.330
rs6810075_C	cg05578595	Adiponectin	-1.234	0.876	0.159
rs7615090_G	cg05578595	Adiponectin	0.231	0.533	0.665
rs2980879_A	cg05578595	Adiponectin	-5.060	19.255	0.793
rs7955516_C	cg05578595	Adiponectin	1.255	2.705	0.643
rs601339_G	cg05578595	Adiponectin	0.943	1.059	0.373
rs7964945_A	cg05578595	Adiponectin	-0.600	1.779	0.736
rs8042532_G	cg05578595	Adiponectin	4.044	9.280	0.663
rs2927324_C	cg05578595	Adiponectin	5.878	31.560	0.852
rs12051272_T	cg05578595	Adiponectin	-1.967	2.146	0.360
rs731839_G	cg05578595	Adiponectin	-0.609	0.770	0.429

Table S3b: Mendelian randomization analysis using SNPs with evidence of association with adiponectin from GWAS as instrumental variables to investigate reverse causation at *ADIPOQ*

SNP	CpG site	Trait	Beta	Standard Error	P-value
rs190934192_A	cg25035485	Apo B	4.202	23.047	0.855
rs629301_G	cg25035485	Apo B	14.004	42.476	0.742
rs1260326_T	cg25035485	Apo B	2.269	2.880	0.431
rs6756629_A	cg25035485	Apo B	1.178	0.676	0.082
rs144064722_G	cg25035485	Apo B	10.567	88.540	0.905
rs182695896_C	cg25035485	Apo B	-3.517	12.705	0.782
rs10056811_A	cg25035485	Apo B	-0.525	4.150	0.899
rs4722043_C	cg25035485	Apo B	0.078	1.094	0.943
rs115849089_A	cg25035485	Apo B	-9.912	64.961	0.879
rs2980875_G	cg25035485	Apo B	-0.420	0.816	0.607
rs635634_T	cg25035485	Apo B	-44.950	890.435	0.960
rs964184_G	cg25035485	Apo B	-1.430	0.936	0.127
rs142130958_A	cg25035485	Apo B	-5.004	19.134	0.794
rs150617279_A	cg25035485	Apo B	0.962	1.203	0.424
rs1081105_C	cg25035485	Apo B	-2.328	4.902	0.635
rs1883711_C	cg25035485	Apo B	0.099	0.674	0.883

Table S3c: Mendelian randomization analysis using SNPs with evidence of association with apolipoprotein B from GWAS as instrumental variables to investigate reverse causation at *APOB*

SNP	CpG site	Trait	Beta	Standard Error	P-value
rs190934192_A	cg00908766	Apo B	-0.343	0.515	0.506
rs1260326_T	cg00908766	Apo B	5.730	16.260	0.725
rs6756629_A	cg00908766	Apo B	3.700	4.976	0.457
rs144064722_G	cg00908766	Apo B	-21.624	355.829	0.952
rs182695896_C	cg00908766	Apo B	-0.658	0.824	0.425
rs10056811_A	cg00908766	Apo B	0.111	0.737	0.881
rs4722043_C	cg00908766	Apo B	-0.057	0.802	0.944
rs115849089_A	cg00908766	Apo B	1.003	0.835	0.230
rs2980875_G	cg00908766	Apo B	0.473	0.831	0.569
rs635634_T	cg00908766	Apo B	22.436	206.885	0.914
rs964184_G	cg00908766	Apo B	-15.376	78.864	0.845
rs142130958_A	cg00908766	Apo B	2.558	4.578	0.577
rs150617279_A	cg00908766	Apo B	4.708	19.372	0.808
rs1081105_C	cg00908766	Apo B	-28.185	594.353	0.962
rs1883711_C	cg00908766	Apo B	0.157	1.056	0.882

Table S3d: Mendelian randomization analysis using SNPs with evidence of association with apolipoprotein B from GWAS as instrumental variables to investigate reverse causation at *SORT1*

SNP	CpG site	Trait	Beta	Standard Error	P-value
rs7553007_A	cg04111102	C-reactive protein	-6.406	7.637	0.402
rs6734238_G	cg04111102	C-reactive protein	-0.473	0.935	0.613
rs7748513_G	cg04111102	C-reactive protein	2.352	3.575	0.511
rs7979473_A	cg04111102	C-reactive protein	-1.538	1.883	0.414
rs1183910_A	cg04111102	C-reactive protein	-11.801	70.720	0.868
rs2259816_T	cg04111102	C-reactive protein	-0.507	1.846	0.784
rs4420638_G	cg04111102	C-reactive protein	-14.441	88.345	0.870

Table S3e: Mendelian randomization analysis using SNPs with evidence of association with C-reactive protein from GWAS as instrumental variables to investigate reverse causation at *LEPR*

SNP	CpG site	Trait	Beta	Standard Error	P-value
rs646776_C	cg19610905	Cholesterol	-8.414	13.767	0.541
rs558971_A	cg19610905	Cholesterol	0.260	0.700	0.710
rs9306897_T	cg19610905	Cholesterol	1.047	1.266	0.409
rs515135_T	cg19610905	Cholesterol	-2.715	2.897	0.349
rs780093_T	cg19610905	Cholesterol	2.606	8.452	0.758
rs6544713_T	cg19610905	Cholesterol	-1.870	2.846	0.511
rs6882076_T	cg19610905	Cholesterol	0.497	1.436	0.730
rs9987289_A	cg19610905	Cholesterol	0.294	5.115	0.954
rs1883025_T	cg19610905	Cholesterol	9.927	41.901	0.813
rs579459_C	cg19610905	Cholesterol	-3.494	4.691	0.457
rs1535_G	cg19610905	Cholesterol	0.197	0.077	0.011
rs10468017_T	cg19610905	Cholesterol	-1.610	2.099	0.443
rs2000999_A	cg19610905	Cholesterol	-2.258	2.097	0.282
rs2156552_A	cg19610905	Cholesterol	1.762	1.360	0.196
rs6511720_T	cg19610905	Cholesterol	-1.389	1.847	0.452
rs2228603_T	cg19610905	Cholesterol	-2.052	2.096	0.328

Table S3f: Mendelian randomization analysis using SNPs with evidence of association with cholesterol from GWAS as instrumental variables to investigate reverse causation at *FADS1*

SNP	CpG site	Trait	Beta	Standard Error	P-value
rs144064722_G	cg04087571	Apo AI	-0.028	2.809	0.992
rs1461729_A	cg04087571	Apo AI	8.999	78.078	0.908
rs75835816_C	cg04087571	Apo AI	-2.291	3.653	0.531
rs1883025_T	cg04087571	Apo AI	15.825	54.031	0.770
rs174594_C	cg04087571	Apo AI	-2.513	2.835	0.376
rs261291_C	cg04087571	Apo AI	1.106	0.840	0.188
rs73424577_G	cg04087571	Apo AI	0.491	0.934	0.599
rs6507939_A	cg04087571	Apo AI	6.219	8.531	0.466

Table S3g: Mendelian randomization analysis using SNPs with evidence of association with apolipoprotein A1 from GWAS as instrumental variables to investigate reverse causation at *APOA1*

SNP	CpG site	Trait	Beta	Standard Error	P-value
rs2228145_C	cg21160290	Interleukin-6	0.013	0.183	0.942

Table S3h: Mendelian randomization analysis using SNPs with evidence of association with interleukin-6 from GWAS as instrumental variables to investigate reverse causation at *ABO*

SNP	CpG site	Trait	Beta	Standard Error	P-value
rs643434_A	cg02856953	Interleukin-6	0.173	0.194	0.372

Table S3i: Mendelian randomization analysis using SNPs with evidence of association with interleukin-6 from GWAS as instrumental variables to investigate reverse causation at *IL6R*

SNP	Gene	CpG	Trait	Sample Size	Beta	SE	P-value
rs266772	*ADIPOQ*	cg05578595	Adiponectin	605	-0.901	0.210	2.01E-05
rs687621	*ABO*	cg21160290	Interleukin-6	605	-2.823	1.940	0.146
rs13375019	*LEPR*	cg04111102	C-reactive protein	605	-0.406	0.142	0.005
rs7549250	*IL6R*	cg02856953	Interleukin-6	605	3.123	3.854	0.418
rs169109	*ADIPOQ*	cg05578595	Adiponectin	605	-0.435	0.169	0.009
rs541041	*APOB*	cg25035485	Apo B	605	0.427	0.165	0.010
rs7528419	*SORT1*	cg00908766	Apo B	605	0.264	0.071	2.04E-04
rs625145	*APOA1*	cg04087571	Apo A1	605	-0.316	0.112	0.005
rs174544	*FADS1*	cg19610905	Cholesterol	605	-0.245	0.109	0.024
rs6749422	*ADCY3*	cg01884057	Body mass index	792	0.144	0.054	0.008

Table S4: Results of Mendelian randomization analysis between DNA methylation (from cord blood) and traits

SNP	Gene	CpG	Trait	Sample Size	Beta	SE	P-value
rs266772	*ADIPOQ*	cg05578595	Adiponectin	647	-0.900	0.190	2.75E-06
rs687621	*ABO*	cg21160290	Interleukin-6	647	-0.278	0.059	2.55E-06
rs13375019	*LEPR*	cg04111102	C-reactive protein	647	-0.394	0.108	2.89E-04
rs7549250	*IL6R*	cg02856953	Interleukin-6	647	0.708	0.319	0.027
rs169109	*ADIPOQ*	cg05578595	Adiponectin	647	-0.445	0.172	0.010
rs541041	*APOB*	cg25035485	Apo B	647	0.245	0.100	0.015
rs7528419	*SORT1*	cg00908766	Apo B	647	0.235	0.058	5.99E-05
rs625145	*APOA1*	cg04087571	Apo A1	647	-0.255	0.076	8.95E-04
rs174544	*FADS1*	cg19610905	Cholesterol	647	-0.248	0.089	0.005
rs6749422	*ADCY3*	cg01884057	Body mass index	846	0.099	0.05	0.043

Table S5: Results of Mendelian randomization analysis between DNA methylation (from blood at age 15.5) and traits

Tables S6

CpG effect			IL6 effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs200533593_A	0.994	3.628	rs200533593_A	0.232	0.892
rs116552240_A	0.569	1.533	rs687621_G	0.154	0.670
rs543040_T	0.312	1.069	rs116552240_A	0.123	0.558
rs644234_G	0.298	1.039	rs687289_A	0.109	0.499
rs493246_A	0.199	0.807	rs657152_A	0.106	0.487
rs8176646_C	0.084	0.376	rs8176646_C	0.058	0.200
rs200700167_A	0.071	0.293	rs544873_A	0.057	0.191
rs676457_T	0.067	0.266	rs582094_T	0.055	0.177
rs612169_G	0.067	0.266	rs582118_G	0.054	0.167
rs491626_T	0.062	0.233	rs494242_T	0.053	0.159

Table S6a: Bivariate fine mapping results at the *ABO* locus

CpG effect			BMI effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs6737082_C	0.886	2.595	rs6737082_C	0.067	0.561
rs59086897_A	0.815	2.346	rs6746013_G	0.062	0.526
rs10865321_C	0.521	1.740	rs10182181_G	0.062	0.520
rs10865322_G	0.479	1.667	rs10182458_G	0.060	0.512
rs6746013_G	0.175	1.030	rs6752378_A	0.058	0.493
rs58048722_C	0.099	0.742	rs6749422_G	0.053	0.452
rs2118826_A	0.092	0.709	rs59086897_A	0.050	0.422
rs1470039_G	0.090	0.701	rs10203482_C	0.035	0.268
rs4665736_C	0.088	0.690	rs10185143_C	0.033	0.239
rs1172294_G	0.081	0.650	rs58048722_C	0.031	0.213

Table S6b: Bivariate fine mapping results at the *ADCY3* locus

CpG effect			Adiponectin effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs266772_T	0.367	0.732	rs115527175_T	0.821	1.631
rs150411458_C	0.192	0.346	rs74577862_A	0.126	0.129
rs78800820_C	0.156	0.234	rs266772_T	0.124	0.12
rs146920076_A	0.156	0.234	rs201071850_G	0.081	-0.088
rs201071850_G	0.140	0.182	rs150411458_C	0.072	-0.142
rs74577862_A	0.126	0.128	rs78800820_C	0.054	-0.275
rs76786086_T	0.123	0.115	rs146920076_A	0.054	-0.275
rs141134215_A	0.120	0.106	rs76786086_T	0.029	-0.559
rs189214085_G	0.120	0.106	rs202214354_G	0.027	-0.580
rs143257534_T	0.120	0.106	rs1011551_T	0.026	-0.608

Table S6c: Bivariate fine mapping results at the *ADIPOQ* (low freq) locus

CpG effect			Adiponectin effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs169109_G	0.472	1.001	rs169109_G	0.421	0.911
rs67654560_A	0.374	0.826	rs34587333_A	0.306	0.695
rs34587333_A	0.294	0.669	rs266728_G	0.282	0.645
rs266728_G	0.279	0.638	rs58575091_T	0.161	0.333
rs58575091_T	0.266	0.609	rs864264_C	0.043	-0.297
rs266720_A	0.183	0.400	rs67654560_A	0.034	-0.399
rs864264_C	0.167	0.353	rs266750_C	0.027	-0.502
rs843991_T	0.161	0.333	rs266759_T	0.026	-0.528
rs16861153_A	0.111	0.145	rs201648331_T	0.026	-0.531
rs266754_T	0.085	0.018	rs843991_T	0.025	-0.541

Table S6d: Bivariate fine mapping results at the *ADIPOQ* (common) locus

CpG effect			APOAI effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs688456_T	0.957	2.519	rs625145_T	0.225	0.639
rs12225230_C	0.494	1.166	rs688456_T	0.167	0.478
rs11216162_A	0.335	0.879	rs11216162_A	0.059	-0.028
rs61903423_A	0.153	0.435	rs12225230_C	0.053	-0.078
rs61905689_T	0.105	0.244	rs7932655_A	0.052	-0.082
rs7928320_T	0.077	0.095	rs34999185_A	0.051	-0.098
rs59511712_T	0.070	0.054	rs61905717_A	0.050	-0.103
rs17120244_T	0.069	0.049	rs7928320_T	0.049	-0.110
rs7946729_T	0.063	0.049	rs17120244_T	0.047	-0.127
rs7932655_A	0.059	-0.027	rs7930783_A	0.044	-0.166

Table S6e: Bivariate fine mapping results at the *APOA1* locus

CpG effect			APOB effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs541041_G	0.148	1.285	rs580889_C	0.108	1.128
rs581411_G	0.127	1.211	rs201027918_C	0.100	1.091
rs562338_A	0.122	1.188	rs10692845_T	0.095	1.068
rs515135_T	0.111	1.144	rs548145_T	0.080	0.984
rs563290_G	0.104	1.111	rs541041_G	0.073	0.942
rs10692845_T	0.100	1.093	rs515135_T	0.072	0.939
rs668948_G	0.100	1.091	rs562338_A	0.067	0.901
rs580889_C	0.098	1.080	rs563290_G	0.062	0.865
rs548145_T	0.075	0.958	rs668948_G	0.060	0.855
rs1652416_A	0.016	0.250	rs581411_G	0.060	0.848

Table S6f: Bivariate fine mapping results at the *APOB* locus

CpG effect			APOB effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs646776_C	0.771	1.316	rs4970836_G	0.164	0.080
rs7528419_G	0.207	0.206	rs629301_G	0.159	0.066
rs629301_G	0.020	-0.897	rs599839_G	0.123	-0.064
rs12740374_T	0.002	-1.924	rs646776_C	0.116	-0.092
rs4970836_G	0	-6.704	rs7528419_G	0.108	-0.130
rs583104_G	0	-7.790	rs583104_G	0.107	-0.132
rs1277930_G	0	-9.292	rs1277930_G	0.080	-0.272
rs599839_G	0	-9.556	rs12740374_T	0.055	-0.446
rs660240_T	0	-9.783	rs602633_T	0.033	-0.676
rs57677983_C	0	-9.783	rs57677983_C	0.029	-0.743

Table S6g: Bivariate fine mapping results at the *SORT1* locus

CpG effect			Cholesterol effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs174559_A	0.986	3.274	rs174548_G	0.134	0.612
rs1535_G	0.527	1.469	rs174560_C	0.116	0.540
rs5792235_C	0.226	0.887	rs174555_C	0.106	0.494
rs174561_C	0.221	0.875	rs199977718_C	0.102	0.479
rs174555_C	0.206	0.836	rs174549_A	0.102	0.477
rs174556_T	0.197	0.812	rs174561_C	0.079	0.356
rs174544_A	0.179	0.760	rs174556_T	0.076	0.338
rs174549_A	0.154	0.682	rs28456_G	0.043	0.076
rs174568_T	0.039	0.031	rs174544_A	0.042	0.059
rs174537_T	0.038	0.014	rs174574_A	0.035	-0.020

Table S6h: Bivariate fine mapping results at the *FADS1* locus

CpG effect			IL6 effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs10908837_G	0.187	0.607	rs12118721_T	0.130	0.420
rs4845618_G	0.164	0.538	rs10908837_G	0.124	0.397
rs6687726_A	0.163	0.535	rs6687726_A	0.113	0.349
rs12117832_A	0.150	0.490	rs12117832_A	0.079	0.179
rs7536152_A	0.141	0.459	rs4845618_G	0.076	0.161
rs12118721_T	0.136	0.440	rs10908836_C	0.056	0.020
rs12129500_T	0.126	0.403	rs4553185_C	0.047	-0.059
rs6686750_A	0.125	0.400	rs6694817_T	0.043	-0.102
rs6689393_A	0.124	0.394	rs10908838_T	0.042	-0.111
rs7526131_G	0.123	0.393	rs34926346_A	0.039	-0.143

Table S6i: Bivariate fine mapping results at the *IL6R* locus

CpG effect			IL6 effect		
SNP	**snp_prob**	**snp_log10bf**	**SNP**	**snp_prob**	**snp_log10bf**
rs6693842_C	0.754	2.232	rs6700896_T	0.122	0.887
rs4655764_T	0.583	1.890	rs200641814_TA	0.078	0.670
rs59508186_A	0.353	1.481	rs7535218_A	0.067	0.602
rs4655582_G	0.173	1.066	rs7524581_T	0.051	0.476
rs10889574_A	0.141	0.959	rs10443261_C	0.039	0.349
rs11208715_T	0.139	0.951	rs7518710_G	0.037	0.328
rs79653480_A	0.139	0.951	rs7541434_A	0.034	0.289
rs12021623_C	0.139	0.951	rs12067936_A	0.032	0.267
rs78052673_C	0.139	0.951	rs7515766_G	0.030	0.241
rs6664374_T	0.129	0.916	rs12753193_G	0.029	0.218

Table S6j: Bivariate fine mapping results at the *LEPR* locus

SNP	Gene	CpG	Pr(single var | trait)	Concordance rate	P_{JLIM}
rs266772	*ADIPOQ*	cg05578595	0.746	0.188	$<10^{-6}$
rs200533593	*ABO*	cg21160290	0.699	0.024	NA
rs6693842	*LEPR*	cg04111102	0.785	0.213	0.95
rs10908837	*IL6R*	cg02856953	0.929	0.069	NA
rs541041	*APOB*	cg25035485	0.736	0.028	1
rs169109	*ADIPOQ*	cg05578595	0.725	0.048	$<10^{-6}$
rs646776	*SORT1*	cg00908766	0.984	0.360	1
rs688456	*APOA1*	cg04087571	0.841	0.077	1
rs174559	*FADS1*	cg19610905	0.809	0.488	0.99
rs6737082	*ADCY3*	cg01884057	0.761	0.012	$<10^{-6}$

Table S7: Results of bivariate fine mapping concordance and joint likelihood mapping (JLIM) SNP – Single Nucleotide Polymorphism, Gene – likely implicated gene, CpG – 450K probe ID, Pr(single var | trait) – the posterior probability that a single causal variant is responsible for the observed effect on complex traits (based on a maximum of 5 causal variants), Concordance rate – concordance rate generated from bivariate fine mapping results, P_{JLIM} – P value reported by the JLIM method

mQTL	Gene	CpG	CpG effect	hQTL	Histone Mark	hQTL effect	hQTL P	2SMR	P-value
rs625145	*APOA1*	cg04087571	-0.884 (0.044)	rs688456	H3K4ME3	0.913 (0.166)	5.58×10^{-7}	-1.033 (0.195)	1.13×10^{-7}
rs625145	*APOA1*	cg04087571	-0.884 (0.044)	rs688456	H3K27AC	0.758 (0.176)	4.95×10^{-5}	-0.857 (0.204)	2.54×10^{-5}
rs625145	*APOA1*	cg04087571	-0.884 (0.044)	rs688456	H3K4ME1	0.930 (0.165)	3.04×10^{-7}	-1.052 (0.194)	5.74×10^{-8}
rs10908837	*IL6R*	cg02856953	-0.303 (0.039)	rs59632925	H3K4ME1	0.620 (0.138)	2.53×10^{-5}	-2.045 (0.526)	1.01×10^{-4}
rs10908837	*IL6R*	cg02856953	-0.303 (0.039)	rs59632925	H3K27AC	0.658 (0.124)	1.14×10^{-6}	-2.172 (0.496)	1.18×10^{-5}

Table S8: Results of analysis investigating causal relationship between methylation and histone modification using Two-Sample Mendelian randomization mQTL – methylation quantitative trait loci, Gene – likely implicated gene, CpG – 450K probe ID, CpG effect – effect estimate of SNP on methylation, hQTL – histone quantitative trait loci,, hQTL effect - effect estimate of LD SNP on histone mark from Grubert et al (2015), hQTL P – P-value for hQTL effect, 2SMR effect – effect estimates from 2-Sample MR analysis, P-value – P-value for observed 2SMR effect