



# Role of DNA Methylation in Type 2 Diabetes Etiology: Using Genotype as a Causal Anchor

Hannah R. Elliott,<sup>1</sup> Hashem A. Shihab,<sup>1</sup> Gabrielle A. Lockett,<sup>2</sup> John W. Holloway,<sup>2,3</sup> Allan F. McRae,<sup>4,5</sup> George Davey Smith,<sup>1</sup> Susan M. Ring,<sup>1</sup> Tom R. Gaunt,<sup>1</sup> and Caroline L. Relton<sup>1</sup>

*Diabetes* 2017;66:1713–1722 | <https://doi.org/10.2337/db16-0874>

Several studies have investigated the relationship between genetic variation and DNA methylation with respect to type 2 diabetes, but it is unknown if DNA methylation is a mediator in the disease pathway or if it is altered in response to disease state. This study uses genotypic information as a causal anchor to help decipher the likely role of DNA methylation measured in peripheral blood in the etiology of type 2 diabetes. Illumina HumanMethylation450 BeadChip data were generated on 1,018 young individuals from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort. In stage 1, 118 unique associations between published type 2 diabetes single nucleotide polymorphisms (SNPs) and genome-wide methylation (methylation quantitative trait loci [mQTLs]) were identified. In stage 2, a further 226 mQTLs were identified between 202 additional independent non-type 2 diabetes SNPs and CpGs identified in stage 1. Where possible, associations were replicated in independent cohorts of similar age. We discovered that around half of known type 2 diabetes SNPs are associated with variation in DNA methylation and postulated that methylation could either be on a causal pathway to future disease or could be a noncausal biomarker. For one locus (*KCNQ1*), we were able to provide further evidence that methylation is likely to be on the causal pathway to disease in later life.

Type 2 diabetes is a major global health problem, affecting ~660 million people in Europe alone (1). Several large-scale genome-wide association studies (GWAS) have identified a major genetic contribution to type 2 diabetes in Europeans

(2,3) and other populations (4–7). Although many of these genetic variants have been linked to perturbed  $\beta$ -cell function (7,8), the molecular pathways through which they mediate their effects remain unclear. Increasing attention is being paid to the potential role of epigenetic mechanisms in mediating the influence of genetic variation on phenotype, including complex diseases (8,9).

Epigenetic mechanisms regulate gene expression in a variety of ways, for example via chromatin remodelling or the control of transcription factor binding by the addition of methyl groups to the DNA sequence. Genetic variants may directly influence DNA methylation marks, through *cis* or local effects, or by more distal *trans* effects, including chromosomal looping. Indeed, it is estimated that 24% of variance in DNA methylation in childhood and 21% of variance in middle age is due to genetic variation (10), and some of the genetic variants involved map to previously identified genetic risk factors for disease. Several loci with genetic variants predisposing to type 2 diabetes have been examined for differences in DNA methylation patterns. *HNF4A*, *IRS1*, *KCNQ1*, *PPARG*, *FTO*, and *TCF7L2* are examples of type 2 diabetes loci that show differences in methylation in type 2 diabetes case subjects compared with control subjects in various tissues (11–13). *FTO* has haplotype-specific methylation patterns, again observed when comparing type 2 diabetes case subjects to control subjects (14). These observations raise the possibility that DNA methylation is causally involved in the biological pathways contributing to type 2 diabetes. However, almost all studies to date have investigated case and control subjects, raising

<sup>1</sup>MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, U.K.

<sup>2</sup>Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, U.K.

<sup>3</sup>Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, U.K.

<sup>4</sup>Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia

<sup>5</sup>The University of Queensland Diamantina Institute, Translational Research Institute, University of Queensland, Brisbane, Queensland, Australia

Corresponding author: Hannah R. Elliott, [hannah.elliott@bristol.ac.uk](mailto:hannah.elliott@bristol.ac.uk).

Received 18 July 2016 and accepted 21 February 2017.

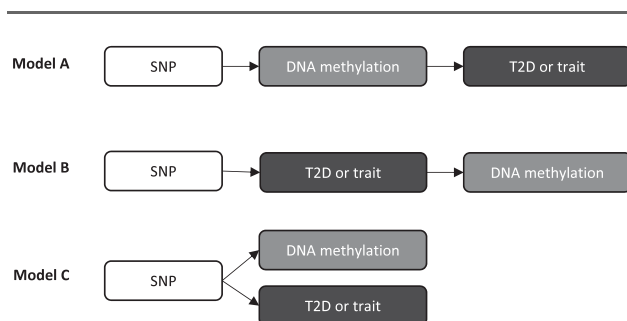
This article contains Supplementary Data online at <http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db16-0874/-/DC1>.

© 2017 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <http://www.diabetesjournals.org/content/license>.

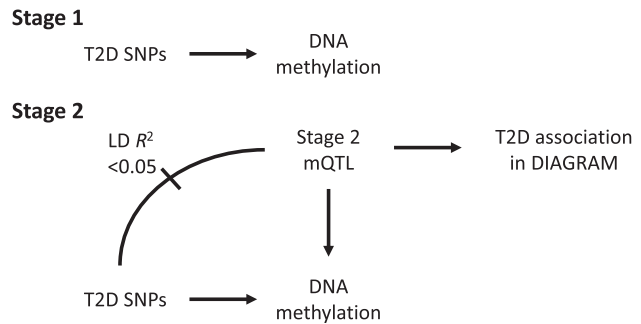
the concern that epigenetic processes may be altered in response to disease state, rather than vice versa.

We postulate that type 2 diabetes genetic risk variants exert their effects on disease (or diabetes-related traits) through perturbation of DNA methylation (Fig. 1, model A). However, genetic risk variants may be associated with DNA methylation through their influence on disease itself (Fig. 1, model B). Alternatively, type 2 diabetes genetic risk variants may be associated with both DNA methylation and disease independently and thus not be linked through a causal pathway (Fig. 1, model C). Genotypic information can provide a causal anchor to allow inferences to be made regarding the direction of the relationship between DNA methylation and type 2 diabetes, thus helping to decipher which of these models is most likely, an approach that forms the basis of Mendelian randomization (15). Mendelian randomization has previously been applied in the context of epigenetic mediation of cardiometabolic disease, such as in the exploration of the causal direction between BMI and *HIF3A* methylation (16) or, more recently, to interrogate causality with respect to many BMI-associated methylation-variable sites (17). The distinction here is that previous studies have applied Mendelian randomization after the identification of a methylation-variable locus. In the current study, Mendelian randomization is used to provide evidence of a mediating role of DNA methylation where the relationship between the causal anchor (type 2 diabetes GWAS single nucleotide polymorphisms [SNPs]) and disease outcome is already well established.

In the first stage of this study (Figs. 2 and 3), we investigated whether any known type 2 diabetes SNPs are associated with DNA methylation (i.e., identify type 2 diabetes SNPs that can be categorized as methylation quantitative trait loci [mQTLs]) in young individuals from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort (18–20). Because these subjects are young and do not have diabetes, such an association is indicative of a causal role of DNA methylation in mediating disease pathogenesis (Fig. 1, model A; although Fig. 1, model C cannot be discounted).



**Figure 1**—Potential pathways in which SNPs influence type 2 diabetes risk. In model A, type 2 diabetes risk variants exert their effects on disease (or disease-related traits) through perturbation of DNA methylation. In model B, genetic risk variants are associated with DNA methylation through their influence on disease itself. In model C, genetic risk variants are associated with DNA methylation and disease independently. T2D, type 2 diabetes.



**Figure 2**—The primary analyses conducted in ALSPAC/ARIES. In stage 1, 118 associations between published type 2 diabetes SNPs and genome-wide methylation were identified. In stage 2, a further 226 mQTLs were identified between 202 additional independent non-type 2 diabetes SNPs and the CpGs identified in stage 1. DIAGRAM data were then used to assess the relationship of these independent mQTLs with type 2 diabetes disease risk in order to strengthen causal inference that DNA methylation is acting as a mediating mechanism. T2D, type 2 diabetes.

To find further evidence for methylation being on a potential causal pathway to future disease, we undertook a second stage of analysis to identify further SNPs which were 1) associated with CpGs identified in stage 1 (i.e., were mQTLs) but 2) not in linkage disequilibrium (LD) with type 2 diabetes SNPs (i.e., were independent of known type 2 diabetes risk SNPs) (Fig. 2, stage 2, and Fig. 3). We then assessed the relationship of these “independent mQTLs” with type 2 diabetes disease risk to strengthen causal inference that DNA methylation is indeed acting as a mediating mechanism. This second step was undertaken using publicly available summary data from DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) (2).

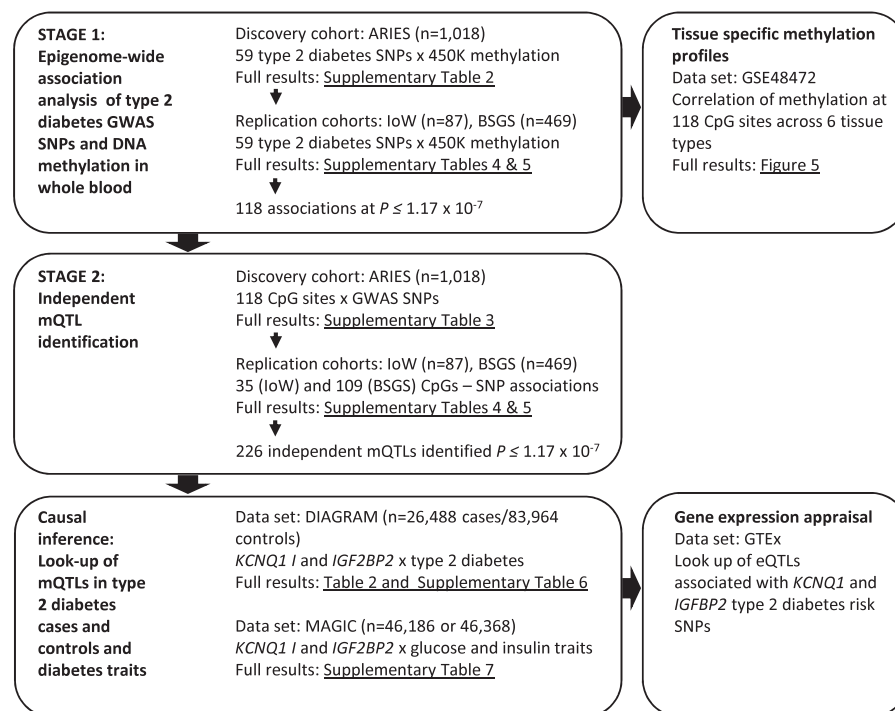
Where there was evidence for type 2 diabetes disease risk being mediated by DNA methylation, we further evaluated this in the context of publicly available gene expression data and phenotypic traits. All stages of analysis, including signposting to relevant results, are summarized in Fig. 3.

## RESEARCH DESIGN AND METHODS

### Samples

ALSPAC is a large prospective cohort study based in the South West of the U.K. ALSPAC recruited 14,541 pregnant women resident in Avon, U.K. with expected dates of delivery 1 April 1991 to 31 December 1992. Detailed information was collected during pregnancy and at regular intervals in the following years from both parents and offspring (18,19). The study website contains details of all the data that are available through a fully searchable data dictionary (<http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/>).

As part of the Accessible Resource for Integrated Epigenomic Studies (ARIES) project, Illumina HumanMethylation450 BeadChip data have been generated in 1,018 mother-offspring pairs from the ALSPAC cohort (20). The



**Figure 3**—A flow diagram showing the stages of analysis conducted with signposting to relevant results.

ARIES participants were selected based on availability of DNA samples at two time points for the mother (antenatal and at follow-up when the offspring were adolescents) and three time points for the offspring (neonatal, childhood [age 7 years], and adolescence [age 15–17 years]). Methylation data from the offspring at age 15–17 years are included in this analysis.

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

### Biological Measures and Anthropometry

Biological and anthropometric measures were collected at the same clinics at which samples for methylation were drawn. Fasting glucose and insulin levels were measured from blood samples in ALSPAC participants who agreed to give a sample and had fasted for a minimum of 4 h.

Height was measured using a Harpenden stadiometer while weight and bioelectrical impedance were measured using a Tanita body fat analyzer. BMI ( $\text{kg}/\text{m}^2$ ) was then calculated.

### Epigenetic Data

Epigenetic data were generated using the Illumina Human-Methylation450 BeadChip (Illumina, San Diego, CA). Detailed methods and normalization procedures have been described previously (20).

### Genetic Data

GWAS data were generated using Illumina HumanHap550-Quad chips by Sample Logistics and Genotyping Facilities

at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. The resulting raw genome-wide genotype data were subjected to standard quality control methods. In brief, individuals were removed if there was evidence of sex mismatches, minimal or excessive heterozygosity, or >3% missingness. SNPs with a minor allele frequency of <1%, a call rate of <95%, or evidence of violations of Hardy-Weinberg equilibrium ( $P < 5 \times 10^{-7}$ ) were removed. Imputation was performed using Impute v2.2.2 software using 1000 Genomes phase 1 version 3 as a reference panel (21,22). For imputed genotypes, dosages were converted to “best guess” genotypes in binary plink format, filtered to include only SNPs with minor allele frequency >1% and imputation info score >0.8.

Sixty-two SNPs associated with type 2 diabetes were selected for analysis based on a large recent GWAS of type 2 diabetes (2). Sixty-one variants (excluding rs3132524) were available in ALSPAC. Two SNPs, rs9502570 and rs2284219, had minor allele frequencies of 0% in the ALSPAC population and were discarded from the analysis. Full details including allele frequencies are shown in Supplementary Table 1.

### Data From Additional Resources

DIAGRAM data were used to investigate the associations between mQTLs identified in stage 2 and type 2 diabetes. These data included 26,488 type 2 diabetes case subjects and 83,964 control subjects (2). Data are freely available from the consortium website [www.diagram-consortium.org](http://www.diagram-consortium.org). Data used on glycemic traits have been contributed

by Meta-analyses of Glucose and Insulin-Related Traits Consortium (MAGIC) investigators and have been downloaded from [www.magicinvestigators.org](http://www.magicinvestigators.org) (23,24). Gene expression data were derived from the Genotype-Tissue Expression (GTEx) Portal (release v6), [www.gtexportal.org](http://www.gtexportal.org) (25).

### Statistical Analysis

For identification of epigenome-wide associations between SNPs and DNA methylation, the Matrix expression quantitative trait loci (eQTL) package was implemented (26). Methylation M-values (27) were first rank-transformed so they followed a normal distribution. Covariates age, sex, batch (defined as bisulphite conversion plate), cell counts (28), and the first 10 principal components from genetic data were regressed out. The resultant residuals were then regressed against genotype for each CpG site on the array.

All analyses were conducted in R, version 3.2.1 (<http://www.r-project.org>). The following R packages were used: base, stats, MatrixEQTL, plyr, snpStats, xlsx, pwr, RCircos, Biobase, and GEOquery.

### Replication Studies

#### Isle of Wight Birth Cohort

In 1989, a whole population birth cohort was recruited on the Isle of Wight (IoW) to assess the impact of heredity and environment on the development of allergic disorders and allergen sensitization. The IoW 1989 birth cohort has been described in detail previously (29). Exact age at 18-year follow-up was calculated from the date of blood sample collection for the 18-year follow-up and the date of birth. BMI was calculated based on height and weight at the 18-year follow-up. DNA methylation was profiled in peripheral blood samples collected at the 18-year follow-up, using Illumina's HumanMethylation450 array in a subset ( $n = 367$ ) of subjects. DNA methylation data were preprocessed using IMA (30) and batch corrected using ComBat (31) as described previously (32). Genotyping was performed in a subset of cohort subjects with DNA methylation data ( $n = 87$ ) using Illumina's OmniExpressExome BeadChip (v1.2). Potential mQTLs were modeled using generalized linear models for the effect of genotype (additive model) on logit-transformed DNA methylation, adjusting for sex and exact age at 18-year follow-up. All analyses used SPSS (v22.0).

#### Brisbane System Genetics Study

A subset of 469 individuals from the Brisbane System Genetics Study (BSGS) (33,34) aged <20 years was used. This consisted of monozygotic and dizygotic twin pairs and their adolescent siblings. DNA methylation was measured using HumanMethylation450 BeadChips, which was cleaned as described in detail elsewhere (34). Genotype data were imputed from Illumina 610-Quad BeadChip arrays against 1000 Genomes phase I version 3 using Impute V2 and filtered to have  $R^2 > 0.8$ . Associations were tested using logistic regression on the SNP genotype correcting for age, sex, and technical covariates (slide and position on slide).

### Power

Power calculations for the discovery (stage 1) analysis indicate that the study had 80% power to detect a true  $R^2 = 0.051$ , where  $n = 896$  and  $\alpha = 0.05/(60 \times 487,000)$ . No power calculation is provided for the replication of these results in the IoW and BSGS cohorts given the lack of independence of the two series of analyses. For stage 2, detecting a mQTL that correlated with a methylation-variable locus in the ARIES study sample, the study had 80% power to detect a true  $R^2 = 0.059$ , where  $n = 896$  and  $\alpha = 0.05/(118 \times 8,000,000)$ . In further analyses, we then assessed the relationship of these independent mQTLs with type 2 diabetes risk using available summary data from DIAGRAM. At this stage, the study had 80% power to detect a true  $R^2 = 0.00024$ , where  $n_{(\text{DIAGRAM})} = 87,167$  and  $\alpha = 0.05/226$ . In MAGIC, we estimate 80% power to detect a true  $R^2 = 0.00030$ , where  $n_{(\text{MAGIC})} = 46,186$  and  $\alpha = 0.05/(3 \times 4)$ .

## RESULTS

### Population Characteristics

Characteristics of the ALSPAC subjects selected for analysis are shown in Table 1. Measurements did not indicate presence of diabetes in any ALSPAC subjects included in the study.

### Stage 1: Identification of Associations Between Type 2 Diabetes Genetic Risk Variants and DNA Methylation

Thirty-three of 59 individual type 2 diabetes SNPs had one or more associations with 118 CpG sites at an epigenome-wide significance threshold  $P$  value of  $P \leq 1.17 \times 10^{-7}$  for each SNP. No single CpG site was associated with more than one SNP, but several SNPs were associated with methylation across clusters of CpGs; for example, rs10190052 on chromosome 2 was associated with three CpG sites spanning 7.2 kb at a distance of 17.5 kb from the SNP. Full results are shown in Supplementary Table 2. Figure 4 shows the genomic distribution of associations identified.  $R^2$  values showed that type 2 diabetes mQTLs explained 3–63% of the variation in methylation. On average, SNPs in close proximity to CpGs explained a greater proportion of variation in methylation than more distant SNPs (Fig. 5). Between SNPs and CpG sites positioned <50 kb apart, 74 associations were

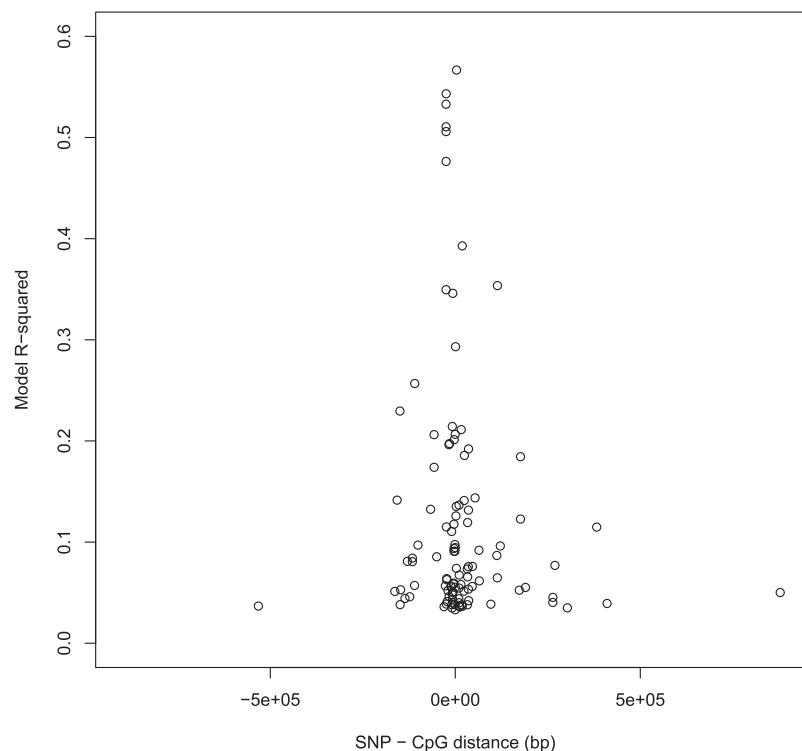
**Table 1—Population characteristics of ALSPAC subjects included in analysis**

<i>N</i>	896
Age (years)	17.1 (1.0)
BMI (kg/m <sup>2</sup> )	22.2 (3.8)
Body fat %	23.36 (10.4)
Waist circumference (cm)	76.7 (8.6)
Fasting glucose (mmol/L)	5.1 (0.4)
Fasting insulin (pmol/L)	48.2 (28.5)
Sex (% male), <i>n</i> (%)	434 (48.4)

Data are *N* or mean (SD), unless otherwise stated.







**Figure 5**—Plot showing the relationship between  $R^2$  and the distance in base pairs between *cis* mQTLs.

from stage 2. Of the potential mQTLs with sufficient data to allow validation, 238 (82%) were nominally associated ( $P < 0.05$ ) and 135 (46%) were associated at  $P < 1.17 \times 10^{-7}$ . The average age of BSGS participants at methylation analysis was 13.9 years (SD 2.2). Participants had a mean BMI of 20.3 kg/m<sup>2</sup> (SD 3.5) and 52% were male. Results of the mQTL analysis in the BSGS cohort can be found in Supplementary Table 5.

#### mQTL Associations With Type 2 Diabetes in DIAGRAM

For each stage 2 and type 2 diabetes mQTL, the association between the mQTL and diabetes was extracted from DIAGRAM consortium data (2). A summary of SNPs available from DIAGRAM data are shown in Supplementary Table 6.

One methylation site associated with a type 2 diabetes risk variant in *KCNQ1* also showed association between an independent mQTL and diabetes in DIAGRAM. One methylation site associated with a risk variant in *IGF2BP2* showed a nominal association notwithstanding adjustment for multiple testing in DIAGRAM. This suggests that for at least one of these two loci, there is evidence that methylation is implicated in the causal pathway between the common genetic variant and type 2 diabetes (Fig. 1, model A). These findings are summarized in Table 2 (below) with full details for all SNPs shown in Supplementary Table 6. However, the majority of independent mQTLs did not show any associations between the SNP and diabetes in DIAGRAM, giving no further supporting evidence to suggest that methylation may be on a causal pathway from these type 2 diabetes SNPs to disease.

#### Cross-Tissue DNA Methylation Patterns

DNA methylation patterns may vary across tissue type, defining tissue-specific transcriptional regulation. We therefore sought to evaluate methylation at the 118 CpG sites most strongly associated with type 2 diabetes SNPs to identify if they have tissue-specific methylation profiles. A subset of data from the Gene Expression Omnibus data entry GSE48472 was used, which included data from blood and a range of type 2 diabetes-relevant tissues, including pancreas, fat, and muscle (35). Although sample numbers were small, mean methylation in blood versus other tissues showed high levels of correlation (Pearson correlation coefficients 0.66–0.91), suggesting measurement in blood was a good proxy for methylation levels in other tissues at the sites under investigation (Fig. 6). This was also true of the two CpG sites for which we have any evidence of mediation (Fig. 1, model A). These CpG sites are indicated in red (cg23956648) and blue (cg14637411) in Fig. 6.

#### Associations Between mQTLs and Type 2 Diabetes-Related Traits in MAGIC

To evaluate whether the SNPs in *KCNQ1* and *IGF2BP2* that may increase risk of type 2 diabetes via methylation are associated with glycemic traits, summary data from the MAGIC consortium was used (23,24). There was no strong evidence to suggest that the SNPs tested are associated with fasting glucose, fasting insulin, or HbA<sub>1c</sub> (Supplementary Table 7), although effect sizes were of the same magnitude and direction in each locus.

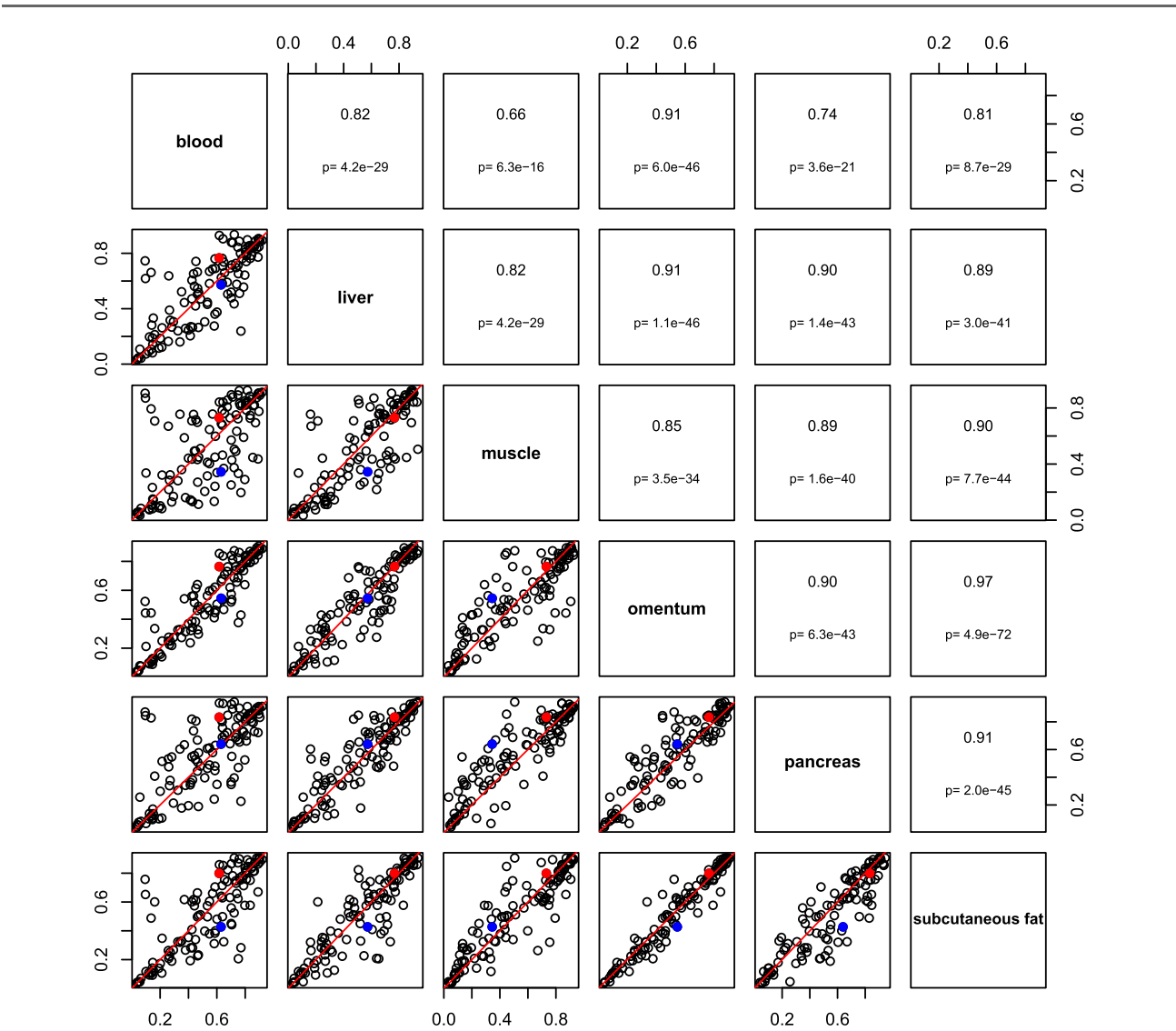
Table 2—SNP associations in DIAGRAM						DIAGRAM associations between SNP and type 2 diabetes		
SNP	Associated CpG	Type 2 diabetes gene name	Variance in methylation explained by SNP ( $R^2$ )	LD between SNPs ( $R^2$ )	OR	95% CI	$P$ value	
rs4402960*	cg23956648	<i>IGF2BP2</i>	7.3%	0.02	1.12	1.09, 1.14	$9.4 \times 10^{-18}$	
rs9850770†			4.3%		1.04	1.01, 1.07	0.01	
rs163184*	cg14637411	<i>KCNQ1</i>	5.5%	0.03	1.11	1.08, 1.14	$1.7 \times 10^{-14}$	
rs2237896†			5.9%		1.24	1.18, 1.3	$2.7 \times 10^{-19}$	

Complete details including genomic locations for SNPs and CpG sites are included in Supplementary Table 6. \*Type 2 diabetes mQTL.  
†Stage 2 mQTL.

Associations Between mQTLs and Gene Expression

To investigate whether the SNPs that may increase risk of type 2 diabetes via methylation showed evidence of association with gene expression, we obtained eQTL data for

single tissues from the GTEx Portal (release v6) for SNPs rs4402960, rs9850770, rs163184, and rs2237896 (25). eQTLs were included for tissues with data from >70 samples using a  $\pm 1$  Mb *cis* window around the transcription



**Figure 6**—Pairwise comparisons across tissues of 118 CpG sites most strongly associated with type 2 diabetes SNPs. Six tissue types are shown (blood:  $n = 11$ ; muscle, omentum, and subcutaneous fat:  $n = 6$ ; liver:  $n = 5$ ; pancreas:  $n = 4$ ). The upper panel shows the Pearson correlation coefficient and  $P$  values; the lower panel shows the pairwise scatterplot (trend line shown in red). Data points for cg23956648 are red and for cg14637411 are blue. These data are a subset of Gene Expression Omnibus data entry GSE48472 (35).

start site. Only one SNP, rs4402960, was identified as an eQTL; this was for *IGF2BP2* in thyroid tissue. For each copy of the minor (type 2 diabetes risk) allele, there was a 0.29-unit increase in rank-normalized gene expression (95% CI 0.18, 0.40;  $P = 6.15 \times 10^{-7}$ ). This eQTL is within intron 2 of the *IGF2BP2* gene.

## DISCUSSION

This analysis examined whether genetic variants predisposing to type 2 diabetes exert their influence on disease via changes in DNA methylation in a young cross-sectional cohort without diabetes. Using genetic variants as causal anchors, we identified that around half of known type 2 diabetes SNPs are associated with variation in DNA methylation and postulated that methylation could either be on a causal pathway to future disease (Fig. 1, model A) or could be a noncausal biomarker (36) (Fig. 1, models B and C).

We then further identified a set of independent mQTLs and assessed their associations with type 2 diabetes in later life using DIAGRAM data. For almost all of these associations, we were unable to provide additional strong evidence that methylation is a key pathway through which SNPs are having an effect. For these SNPs, methylation at the associated CpGs could simply be noncausal biomarkers of later disease (Fig. 1, model C), with potential utility in disease prediction. Whether such information on methylation levels adds anything further to genotype information with respect to risk prediction warrants a more detailed statistical appraisal. Recent work in this area by Wahl et al. (17) demonstrates that BMI-associated methylation variation is a very effective predictor of subsequent type 2 diabetes.

To support our data, we sought replication in similarly aged samples from the IoW cohort and BSGS. The replication samples were a smaller size, so they are likely to be underpowered to detect some of the associations captured in the mQTL analysis of the discovery cohort (ARIES). However, the majority of associations were replicated and showed similar effect sizes. Secondary analysis of large-scale GWAS consortia data (DIAGRAM and MAGIC) provided a suitably powered analysis of the potential consequences of variation in DNA methylation on type 2 diabetes risk and related traits. However, power could be further improved by increasing sample size as and when data become available. Analysis of methylation and gene expression reference data highlighted the broader application of our findings in other tissues, despite the primary analyses being conducted on DNA methylation measurements undertaken in peripheral blood. However, in these analyses, there were insufficient data to draw conclusions about mechanisms by which mQTLs are exerting biological effects.

Several recent studies have sought to identify methylation variation associated with type 2 diabetes using an epigenome-wide association study design (37–41). These studies have reported methylation-variable loci, including *KCNQ1*, but have largely used a case-control design and have not focused on delineating the direction of causation from disease to methylation or vice versa. A particular

strength of this study is the use of young subjects who are not only disease free but are unlikely to be in pre-clinical stages of disease. This enabled exploration of SNP-methylation relationships without measuring methylation differences that result from reverse causation.

One potential drawback of this study is that the type 2 diabetes-associated SNPs used in the initial analysis were drawn from only one study; however, this is one of the largest *trans*-ethnic GWAS available. Data analysis in ARIES, BSGS, and IoW was restricted to samples of predominantly white European ancestry. It is therefore not possible to generalize these findings to other ethnicities. The three study cohorts were ethnically homogeneous; however, other factors such as lifestyle, demographic, or socioeconomic factors may have affected the consistency of observations between the cohorts. It is also possible that methylation may mediate the risk SNP–disease relationship in an age-dependent manner, and this was not addressed in this study.

In stage 2 of the analysis, we used genetic variants tagging CpG sites as causal anchors to attempt to build on evidence that methylation is a possible pathway through which SNPs are influencing later disease. This adopts the principle of Mendelian randomization but without formal instrumental variables analysis (15,42). However, it should be noted although the LD between stage 2 and type 2 diabetes mQTLs was low ( $<0.05$ ), most SNP pairs identified in this study were still in *cis*. As discussed in previous gene expression studies (43), it is still possible that the stage 2 and type 2 diabetes SNPs could each have direct effects on type 2 diabetes and methylation (Fig. 1, model C). This issue can only be fully resolved by identification of *trans* variants from larger methylation GWAS, when power will be large enough to make stronger claims (44).

Further analysis, particularly of CpG sites in the imprinted gene *KCNQ1*, will deepen our understanding of the etiology of type 2 diabetes. For *KCNQ1*, there is evidence that variation in methylation potentially plays a role in type 2 diabetes, including differential methylation between type 2 diabetes case and control subjects in both adipose and pancreatic islets (11,12,45). Interestingly, *KCNQ1* risk alleles also show parent of origin-specific effects, influencing disease susceptibility when maternally inherited; these risk alleles also appear to impact on local DNA methylation (46,47). To our knowledge, there is no prior evidence that methylation variation at the *IGF2BP2* locus has been associated with type 2 diabetes. However, *IGF2BP2* acts as a key regulator of *IGF2* translation (48), and *IGF2* is an imprinted locus whose methylation affects fetal growth (49–52). Genetic variance in methylation at these CpG sites explains a relatively small proportion of the total variation in methylation observed; however, in the context of this study, this genetic variance is used as an instrumental variable from which we can draw causal inference (53). Methylation may be responsive to environmental stimuli as well as to genetic variation, which may increase an individual's disease risk further (11,54,55).

A further potential extension of this work is that a methylation score predicting future type 2 diabetes risk



could potentially be generated from the 118 type 2 diabetes SNP-associated CpG sites identified in this study; a similar approach has been used previously to predict exposure to cigarette smoke from DNA methylation data (56), or could be used in combination with a genetic risk score, as has been applied in the context of trait prediction for BMI and height (57). This would require more extensive statistical appraisal involving the training and testing of such a methylation score in independent data sets.

The study design applied here provides a framework for the exploration of DNA methylation as a causal mechanism linking established common genetic variants with disease outcomes and is relevant to a wide range of common complex diseases. This study design focused on the identification of methylation variation that may be implicated in the pathogenesis of type 2 diabetes. Given that only one stage 2 mQTL was identified (in *KCNQ1*), it is highly unlikely that methylation mediates the genetic effects on type 2 diabetes identified to date. A set of probable noncausal biomarkers of later disease were identified. Further work is required to identify any potential predictive utility of these methylation sites. For one locus (*KCNQ1*), we were able to provide further evidence that methylation is likely to be on the causal pathway to disease in later life. Further confirmation of this finding could be achieved with further research including laboratory analyses. Further work is also required to establish whether DNA methylation changes might be induced as a consequence of type 2 diabetes (Fig. 1, model B) and whether such changes might be implicated in downstream comorbidities of this disease.

**Acknowledgments.** The authors are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. Peter Visscher, Grant Montgomery, and Nicholas Martin (The University of Queensland) are acknowledged for their role in generating the BSGS data set.

**Funding.** The Isle of Wight birth cohort assessments have been supported by grants from the National Institutes of Health (R01 HL082925, R01 AI091905, R01 AI121226) and Asthma UK (364). The U.K. Medical Research Council and the Wellcome Trust (grant 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This work was performed in the Medical Research Council Integrative Epidemiology Unit (MC\_UU\_12013/2 and MC\_UU\_12013/8). Methylation data in the ALSPAC cohort was generated as part of the U.K. Biotechnology and Biological Sciences Research Council-funded (BB/I025751/1) ARIES (<http://www.ariesepigenomics.org.uk>). GWAS data were generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp using support from 23andMe. H.R.E. is supported by an Oak Foundation post-doctoral research fellowship award. The BSGS data were supported by National Health and Medical Research Council (NHMRC) grants 1010374, 496667, and 1046880. A.F.M. is supported by the NHMRC Fellowship Scheme (1083656).

**Duality of Interest.** No potential conflicts of interest relevant to this article were reported.

**Author Contributions.** H.R.E. researched data and wrote the manuscript. H.A.S., G.A.L., and A.F.M. researched data and reviewed the manuscript. J.W.H., G.D.S., S.M.R., T.R.G., and C.L.R. contributed to discussion and reviewed the manuscript. H.R.E. is the guarantor of this work and, as such, had full access

to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## References

1. International Diabetes Federation. *IDF Diabetes Atlas, 7th edition* [Internet]. 2015. Brussels, Belgium, International Diabetes Federation. Available from <http://www.diabetesatlas.org>. Accessed 1 July 2016
2. Mahajan A, Go MJ, Zhang W, et al.; DIABetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Asian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) Consortium; South Asian Type 2 Diabetes (SAT2D) Consortium; Mexican American Type 2 Diabetes (MAT2D) Consortium; Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) Consortium. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014;46:234–244
3. Voight BF, Scott LJ, Steinthorsdottir V, et al.; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010;42:579–589
4. Hara K, Fujita H, Johnson TA, et al.; DIAGRAM consortium. Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum Mol Genet* 2014;23:239–246
5. Williams AL, Jacobs SB, Moreno-Macías H, et al.; SIGMA Type 2 Diabetes Consortium. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* 2014;506:97–101
6. Kooner JS, Saleheen D, Sim X, et al.; DIAGRAM; MuTHER. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 2011;43:984–989
7. Saxena R, Saleheen D, Been LF, et al.; DIAGRAM; MuTHER; AGEN. Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes* 2013;62:1746–1755
8. Lockett GA, Patil VK, Soto-Ramírez N, Ziyab AH, Holloway JW, Karmaus W. Epigenomics and allergic disease. *Epigenomics* 2013;5:685–699
9. Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. *Hum Mol Genet* 2016;25(R2):R149–R156
10. Gaunt TR, Shihab HA, Hemani G, et al. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol* 2016;17:61
11. Nilsson E, Jansson PA, Perflyev A, et al. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes* 2014;63:2962–2976
12. Dayeh T, Volkov P, Salö S, et al. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet* 2014;10:e1004160
13. Ribel-Madsen R, Fraga MF, Jacobsen S, et al. Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. *PLoS One* 2012;7:e51302
14. Bell CG, Finer S, Lindgren CM, et al. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS One* 2010;5:e14040
15. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int J Epidemiol* 2012;41:161–176
16. Richmond RC, Sharp GC, Ward ME, et al. DNA methylation and BMI: Investigating identified methylation sites at HIF3A in a causal framework. *Diabetes* 2016;65:1231–1244
17. Wahl S, Drong A, Lehne B, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* 2017;541:81–86
18. Boyd A, Golding J, Macleod J, et al. Cohort profile: the ‘children of the 90s’—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* 2013;42:111–127
19. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* 2013;42:97–110

20. Relton CL, Gaunt T, McArdle W, et al. Data resource profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). *Int J Epidemiol* 2015;44:1181–1190
21. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529
22. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)* 2011;1:457–470
23. Dupuis J, Langenberg C, Prokopenko I, et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010;42:105–116
24. Soranzo N, Sanna S, Wheeler E, et al.; WTCCC. Common variants at 10 genomic loci influence hemoglobin A<sub>1c</sub> levels via glycemic and nonglycemic pathways [published correction appears in *Diabetes* 2011;60:1050–1051]. *Diabetes* 2010;59:3229–3239
25. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580–585
26. Shabalín AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 2012;28:1353–1358
27. Du P, Zhang X, Huang CC, et al. Comparison of beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* 2010;11:587
28. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;13:86
29. Arshad SH, Stevens M, Hide DW. The effect of genetic and environmental factors on the prevalence of allergic disorders at the age of two years. *Clin Exp Allergy* 1993;23:504–511
30. Wang D, Yan L, Hu Q, et al. IMA: an R package for high-throughput analysis of Illumina's 450K Infinium methylation data. *Bioinformatics* 2012;28:729–730
31. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118–127
32. Lockett GA, Soto-Ramírez N, Ray MA, et al. Association of season of birth with DNA methylation and allergic disease. *Allergy* 2016;71:1314–1324
33. Powell JE, Henders AK, McRae AF, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One* 2012;7:e35430
34. McRae AF, Powell JE, Henders AK, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol* 2014;15:R73
35. Slieker RC, Bos SD, Goeman JJ, et al. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin* 2013;6:26
36. Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Med* 2010;7:e1000356
37. Chambers JC, Loh M, Lehne B, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol* 2015;3:526–534
38. Kulkarni H, Kos MZ, Neary J, et al. Novel epigenetic determinants of type 2 diabetes in Mexican-American families. *Hum Mol Genet* 2015;24:5330–5344
39. Florath I, Butterbach K, Heiss J, et al. Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. *Diabetologia* 2016;59:130–138
40. Kriebel J, Herder C, Rathmann W, et al. Association between DNA methylation in whole blood and measures of glucose metabolism: Kora F4 study. *PLoS One* 2016;11:e0152314
41. Soriano-Tárraga C, Jiménez-Conde J, Giral-Steinhauer E, et al.; GENESTROKE Consortium. Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum Mol Genet* 2016;25:609–619
42. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;23(R1):R89–R98
43. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016;48:481–487
44. Rakitsch B, Stegle O. Modelling local gene networks increases power to detect trans-acting genetic effects on gene expression. *Genome Biol* 2016;17:33
45. Dayeh TA, Olsson AH, Volkov P, Almgren P, Ronn T, Ling C. Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. *Diabetologia* 2013;56:1036–1046
46. Travers ME, Mackay DJ, Dekker Nitert M, et al. Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets. *Diabetes* 2013;62:987–992
47. Kong A, Steinthorsdottir V, Masson G, et al. Parental origin of sequence variants associated with complex diseases. *Nature* 2009;462:868–874
48. Christiansen J, Kolte AM, Hansen Tv, Nielsen FC. IGF2 mRNA-binding protein 2: biological function and putative role in type 2 diabetes. *J Mol Endocrinol* 2009;43:187–195
49. Smith AC, Choufani S, Ferreira JC, Weksberg R. Growth regulation, imprinted genes, and chromosome 11p15.5. *Pediatr Res* 2007;61:43R–47R
50. Delaval K, Wagschal A, Feil R. Epigenetic deregulation of imprinting in congenital diseases of aberrant growth. *BioEssays* 2006;28:453–459
51. Su R, Wang C, Feng H, et al. Alteration in expression and methylation of IGF2/H19 in placenta and umbilical cord blood are associated with macrosomia exposed to intrauterine hyperglycemia. *PLoS One* 2016;11:e0148399
52. King K, Murphy S, Hoyo C. Epigenetic regulation of newborns' imprinted genes related to gestational growth: patterning by parental race/ethnicity and maternal socioeconomic status. *J Epidemiol Community Health* 2015;69:639–647
53. Relton CL, Davey Smith G. Mendelian randomization: applications and limitations in epigenetic studies. *Epigenomics* 2015;7:1239–1243
54. Ligthart S, Steenaard RV, Peters MJ, et al.; BIOS consortium. Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. *Diabetologia* 2016;59:998–1006
55. Gómez-Úriz AM, Milagro FI, Mansego ML, et al. Obesity and ischemic stroke modulate the methylation levels of KCNQ1 in white blood cells. *Hum Mol Genet* 2015;24:1432–1440
56. Elliott HR, Tillin T, McArdle WL, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics* 2014;6:4
57. Shah S, Bonder MJ, Marioni RE, et al.; BIOS Consortium. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am J Hum Genet* 2015;97:75–85