The role of DNA methylation in type 2 diabetes aetiology – using genotype as a causal anchor

Short running title: The role of DNA methylation in type 2 diabetes aetiology

Authors

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

Author Affiliations

1. MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK
2. Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK.
3. Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK.
4. Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia
5. The University of Queensland Diamantina Institute, Translational Research Institute, University of Queensland, Brisbane, QLD, Australia.

Corresponding author:

Hannah R. Elliott,
University of Bristol
Oakfield House
Oakfield Road
Bristol
BS8 2BN
United Kingdom
hannah.elliott@bristol.ac.uk
Tel. +44 (0)117 331 3344
Fax. +44 (0)117 331 4052

Word count: 4178

Number of tables: 2

Number of Figures: 6

Abstract

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 aetiology of type 2 diabetes.

Illumina HumanMethylation450 BeadChip data was generated on 1,018 young individuals from the 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 non-causal 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.

Keywords

Avon Longitudinal Study of Parents and Children, ALSPAC DNA methylation, Type 2 diabetes, causality, epigenetic epidemiology, Mendelian randomization

Introduction

Type 2 diabetes is a major global health problem, affecting around 660 million people in Europe alone (1). Several large-scale genome wide association studies 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 cases compared to controls in various tissues (11–13). *FTO* has haplotype-specific methylation patterns, again observed when comparing type 2 diabetes cases to controls (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 cases and controls, raising 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. (Figure 1, Model A). However, genetic risk variants may be associated with DNA methylation through their influence on disease itself (Figure 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 (Figure 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 which 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 body mass index (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 following 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 SNPs) and disease outcome is already well-established.

In the first stage of this study (Figure 2, Figure 3), we investigate whether any known type 2 diabetes SNPs are associated with DNA methylation (i.e. identify type 2 diabetes SNPs that can be categorised as methylation QTLs) in young individuals from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort (18–20). Because these subjects are young and non-diabetic, such an association is indicative of a causal role of DNA methylation in mediating disease pathogenesis (Figure 1, Model A), although Figure 1, Model C cannot be discounted.

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 i) were associated with CpGs identified in stage 1 (i.e. were mQTLs) but ii) were not in linkage disequilibrium (LD) with type 2 diabetes SNPs (i.e. were independent of known type 2 diabetes risk SNPs) (Figure 2, Stage2, Figure 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 summarised in Figure 3.

Research Design and Methods

Samples

ALSPAC is a large prospective cohort study based in the South West of the UK. ALSPAC recruited 14,541 pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st 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 is 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 has been generated in 1,018 mother-offspring pairs from the ALSPAC cohort (20). The 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 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 hours.

Height was measured using a Harpenden stadiometer while weight and bioelectrical impedance were measured using a Tanita Body Fat Analyser. Body mass index (kg/m^2) was then calculated.

Epigenetic data

Epigenetic data were generated using the Illumina HumanMethylation450 BeadChip (Illumina, San Diego, CA, USA). Detailed methods and normalisation procedures have been described previously (20).

Genetic Data

Genome-Wide Association Study (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. Briefly, individuals were removed if there was evidence of gender 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 ($p < 5x10^{-7}$) 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 cases and 83,964 controls (2). Data are freely available from the consortium website www.diagram-consortium.org.

Data utilised on glycaemic traits have been contributed by MAGIC investigators and have been downloaded from www.magicinvestigators.org (23,24).

Gene expression data were derived from the GTEx Portal (release v6) www.gtexportal.org (25).

Statistical analysis

For identification of epigenome-wide associations between SNPs and DNA methylation, the Matrix 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 ten 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 utilised: base, stats, MatrixEQTL, plyr, snpStats, xlxs, pwr, RCircos, Biobase and GEOquery.

Replication studies

Isle of Wight Birth Cohort (IoW)

In 1989, a whole population birth cohort was recruited on the Isle of Wight to assess the impact of heredity and environment on the development of allergic disorders and allergen sensitisation. 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 pre-processed 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 modelled using generalised 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 (BSGS)

A subset of 469 individuals from (BSGS) (33,34) with ages less than 20 years was utilised. This consisted of MZ and DZ 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 1000Genomes 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*487000)). 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*8000000). 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. 0.00024, where $n_{(DIAGRAM)}$=87167 and alpha=0.05/226. In MAGIC, we estimate 80% power to detect a true R^2= 0.00030 where $n_{(MAGIC)}$=46186 and alpha=0.05/(3*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

33 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 3 CpG sites spanning 7.2kb at a distance of 17.5kb from the SNP. Full results are shown in Supplementary Table 2. Figure 4 shows the genomic distribution of associations identified. R^2 values showed type 2 diabetes mQTLs explained 3% to 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 (Figure 5). 74 associations were observed between SNPs and CpG sites positioned less than 50kb apart. Seven associations were observed between SNPs and CpG sites on different chromosomes and the remainder (n=37) are on the same chromosome but with >50kb in distance between the SNP and CpG site.

Stage 2: Identification of independent mQTLs

For each CpG site associated with a type 2 diabetes SNP, we attempted to identify a further independent set of mQTLs ($p<1x10^{-07}$) using ALSPAC ARIES data where the Linkage Disequilibrium (LD) r^2 between the index diabetes SNP and additional mQTL was <0.05. To distinguish them from the type 2 diabetes mQTLs identified initially, these mQTLs are referred to as stage 2 mQTLs. A table documenting the stage 2 mQTLs for each CpG is shown in Supplementary Table 3. Of the 118 type 2 diabetes-CpG associations identified in stage 1, a further 226 independent mQTLs were identified in stage 2 for 81 of these 118 CpG sites. For each CpG, resultant mQTLs were independent of each other and the type 2 diabetes SNP. No stage 2 mQTLs were found for CpGs associated with rs17106184, rs2028299, rs2075423 or rs4273712.

Replication of mQTLs

From the IoW birth cohort, data were available on 35/118 mQTL associations from stage 1 and 14/226 mQTL associations from stage 2. Of the 49 potential mQTLs with sufficient data to allow validation, 37 (76%) were nominally associated ($p<0.05$) and 12 (24.5%) were associated at $p<1.17x\ 10^{-7}$). The average age of IoW participants at methylation analysis was 17.7 years (SD 0.48 years). Participants had a mean BMI of 23.7 kg/m^2 (SD 4) and 41.4% were male. Results of the mQTL analysis in the IoW cohort can be found in Supplementary Table 4.

From the BSGS, data were available on 109/118 mQTL associations from stage 1 and 183/226 mQTL associations 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 $p<1.17x10^{-7}$. 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^2 (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 is 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 not withstanding 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 (Figure 1, Model A). These findings are summarised 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 (Figure 6). This was also true of the two CpG sites for which we have any evidence of mediation (Figure 1, Model A). These CpG sites are indicated in red (cg23956648) and blue (cg14637411) in Figure 6.

Associations between mQTLs and type 2 diabetes related traits in the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC)

To evaluate whether the SNPs in *KCNQ1* and *IGF2BP2* that may increase risk of type 2 diabetes via methylation are associated with glycaemic 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 HbA1c (Supplementary Table 7) although effect sizes were of the same magnitude and direction in each locus.

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 +/- 1 Mb *cis* window around the transcription 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-normalised gene expression (95% CI: 0.18, 0.40, $p=6.15x10^{-7}$). 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, non-diabetic, cross-sectional cohort. 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 (Figure 1, model A) or could be a non-causal biomarker (Figure 1, Models B & C) (36).

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 non-causal biomarkers of later disease (Figure 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 and colleagues demonstrate that BMI-associated methylation variation is a very effective predictor of subsequent type 2 diabetes (17).

To support our data, we sought replication in similarly aged samples from the Isle of Wight cohort and Brisbane Systems Genomics Study. The replication samples were of much smaller size, so 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 (EWAS) design (37–41). These studies have reported methylation variable loci, including *KCNQ1*, but have largely utilised 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 generalise 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 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 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 (Figure 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 aetiology of type 2 diabetes. For *KNCQ1,* there is evidence that variation in methylation potentially plays a role in type 2 diabetes including differential methylation between type 2 diabetes cases and controls 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 foetal 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 increasing 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 body mass index 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 non-causal 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 (Figure 1, model B) and whether such changes might be implicated in downstream co-morbidities of this disease.

Author contributions

HRE researched data and wrote the manuscript. HAS, GAL and AFM researched data and reviewed the manuscript. JWH, GDS, SMR, TRG and CLR contributed to discussion and reviewed the manuscript.

Acknowledgements

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 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 ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This work was carried out in the MRC Integrative Epidemiology Unit (MC_UU_12013/2, MC_UU_12013/8). Methylation data in the ALSPAC cohort was generated as part of the UK BBSRC funded (BB/I025751/1) Accessible Resource for Integrated Epigenomic Studies (ARIES, http://www.ariesepigenomics.org.uk). GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corportation of America) using support from 23andMe.

HRE is supported by an Oak Foundation post-doctoral research fellowship award.

The BSGS data was supported by NHMRC grants 1010374, 496667 and 1046880. AFM is supported by the NHMRC Fellowship Scheme (1083656). Peter Visscher, Grant Montgomery and Nicholas Martin (all University of Queensland, Australia) are acknowledged for their role in generating the BSGS datset.

HRE 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.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Tables

Table 1 population characteristics of ALSPAC subjects included in analysis

Measure	**Mean(SD) or count**
Number	896
Age (years)	17.1 (1.0)
BMI (kg/m^2)	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)	434 (48%)

Table 2 SNP associations in DIAGRAM

SNP	Associated CpG	type 2 diabetes gene name	Variance in methylation explained by SNP (r^2)	Linkage disequilibrium between SNPs (r^2)	DIAGRAM associations between SNP and type 2 diabetes		
					OR	95% confidence interval	p-value
rs4402960 *	cg23956648	*IGF2BP2*	7.3%	0.02	1.12	1.09, 1.14	$9.4x10^{-18}$
rs9850770 †			4.3%		1.04	1.01, 1.07	0.01
rs163184*	cg14637411	*KCNQ1*	5.5%	0.03	1.11	1.08, 1.14	$1.7x10^{-14}$
rs2237896 †			5.9%		1.24	1.18, 1.3	$2.7x10^{-19}$

*type 2 diabetes mQTL; † stage 2 mQTL. Complete details including genomic locations for SNPs and CpG sites are included in Supplementary Table 6.

Figure Legends

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.

Figure 2 The primary analyses conducted in ALSPAC/ARIES. In stage one, 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 was 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.

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

Figure 4 Circos Plot showing distribution of SNP-methylation associations (mQTLs) throughout the genome. Wide numbered grey bands represent chromosomes. Each SNP is labelled with its approximate genomic location. *Cis* associations are linked with red lines. Blue lines connect associated CpGs and SNPs that are positioned on different chromosomes.

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

Figure 6 Pairwise comparisons across tissues, of 118 CpG sites most strongly associated with type 2 diabetes SNPs. Seven tissue types are shown (blood: n=11; muscle, omentum & subcutaneous fat: n=6; liver: n=5; pancreas: n=4; spleen: n=3). The upper panel shows the Pearson correlation coefficient and p values; the lower panel shows the pairwise scatterplot (trendline 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).

References

1. International Diabetes Federation Diabetes Atlas [Internet]. Seventh. 2015. Available from: http://www.diabetesatlas.org

2. Mahajan A, Go MJ, Zhang W, Below JE, Gaulton KJ, Ferreira T, et al. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. Nat Genet [Internet]. 2014;46(3):234–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24509480

3. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, et al. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet [Internet]. 2010;42(7):579–89. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20581827

4. Hara K, Fujita H, Johnson TA, Yamauchi T, Yasuda K, Horikoshi M, et al. Genome-wide association study identifies three novel loci for type 2 diabetes. Hum Mol Genet [Internet]. 2014;23(1):239–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23945395

5. Williams AL, Jacobs SB, Moreno-Macias H, Huerta-Chagoya A, Churchhouse C, Marquez-Luna C, et al. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. Nature [Internet]. 2014;506(7486):97–101. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24390345

6. Kooner JS, Saleheen D, Sim X, Sehmi J, Zhang W, Frossard P, et al. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. Nat Genet [Internet]. 2011;43(10):984–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21874001

7. Saxena R, Saleheen D, Been LF, Garavito ML, Braun T, Bjonnes A, et al. Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. Diabetes [Internet]. 2013;62(5):1746–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23300278

8. Lockett GA, Patil VK, Soto-Ramirez N, Ziyab AH, Holloway JW, Karmaus W. Epigenomics and allergic disease. Epigenomics [Internet]. 2013;5(6):685–99. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24283882

9. Richmond RC, Hemani G, Tilling K, Davey Smith G, Relton CL. Challenges and novel approaches for investigating molecular mediation. Hum Mol Genet. 2016;

10. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic identification of genetic influences on methylation across the human life course. Genome Biol [Internet]. 2016;17(1):61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27036880

11. Nilsson E, Jansson PA, Perfilyev A, Volkov P, Pedersen M, Svensson MK, et al. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. Diabetes [Internet]. 2014;63(9):2962–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24812430

12. Dayeh T, Volkov P, Salo S, Hall E, Nilsson E, Olsson AH, 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. 2014/03/08. 2014;10(3):e1004160.

13. Ribel-Madsen R, Fraga MF, Jacobsen S, Bork-Jensen J, Lara E, Calvanese V, et al. Genome-wide analysis of DNA methylation differences in muscle and fat from monozygotic twins discordant for type 2 diabetes. PLoS One. 2012/12/20. 2012;7(12):e51302.

14. Bell CG, Finer S, Lindgren CM, Wilson GA, Rakyan VK, Teschendorff AE, et al. Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. PLoS One. 2010/12/03. 2010;5(11):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 [Internet]. 2012;41(1):161–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22422451

16. Richmond RC, Sharp GC, Ward ME, Fraser A, Lyttleton O, McArdle WL, et al. DNA methylation and BMI: Investigating identified methylation sites at HIF3A in a causal framework. Diabetes. 2016;65(5):1231–44.

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

18. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: the "children of the 90s"--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol [Internet]. 2013;42(1):111–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22507743

19. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol [Internet]. 2013;42(1):97–110. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22507742

20. Relton CL, Gaunt T, McArdle W, Ho K, Duggirala A, Shihab H, et al. Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES). Int J Epidemiol [Internet]. 2015;44(4):1181–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25991711

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 [Internet]. 2009;5(6):e1000529. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19543373

22. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 [Internet]. 2011;1(6):457–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22384356

23. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet [Internet]. 2010;42(2):105–16. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/20081858

24. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, et al. Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways. Diabetes [Internet]. 2010;59(12):3229–39. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20858683

25. GTEx Consortium TGte. The Genotype-Tissue Expression (GTEx) project. Nat Genet [Internet]. 2013;45(6):580–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23715323%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4010069

26. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics [Internet]. 2012;28(10):1353–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22492648

27. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics [Internet]. 2010;11:587. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21118553

28. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics [Internet]. 2012;13:86. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22568884

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. ClinExpAllergy. 1993;23(6):504–11.

30. Wang D, Yan L, Hu Q, Sucheston LE, Higgins MJ, Ambrosone CB, et al. IMA: An R package for high-throughput analysis of Illumina's 450K Infinium methylation data. Bioinformatics. 2012;28(5):729–30.

31. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8(1):118–27.

32. Lockett GA, Soto-Ramírez N, Ray MA, Everson TM, Xu CJ, Patil VK, et al. Association of season of birth with DNA methylation and allergic disease. Allergy Eur J Allergy Clin Immunol. 2016;71(9):1314–24.

33. Powell JE, Henders AK, McRae AF, Caracella A, Smith S, Wright MJ, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. PLoS One [Internet]. 2012;7(4):e35430. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22563384

34. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biol [Internet]. 2014;15(5):R73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24887635

35. Slieker RC, Bos SD, Goeman JJ, Bovee J V, Talens RP, van der Breggen R, et al. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics Chromatin [Internet]. 2013;6(1):26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23919675

36. Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. PLoS Med [Internet]. 2010;7(10):e1000356. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21048988

37. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol. 2015/06/23. 2015;3(7):526–34.

38. Kulkarni H, Kos MZ, Neary J, Dyer TD, Kent JW, G??ring HHH, et al. Novel epigenetic determinants of type 2 diabetes in Mexican-American families. Hum Mol Genet. 2015;24(18):5330–44.

39. Florath I, Butterbach K, Heiss J, Bewerunge-Hudler M, Zhang Y, Schöttker B, et al. Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. Diabetologia. 2016;59(1):130–8.

40. Kriebel J, Herder C, Rathmann W, Wahl S, Kunze S, Molnos S, et al. Association between DNA Methylation in whole blood and measures of glucose metabolism: Kora F4 study. PLoS One. 2016;11(3).

41. Soriano-Tárraga C, Jiménez-Conde J, Giralt-Steinhauer E, Mola-Caminal M, Vivanco-Hidalgo RM, Ois A, et al. Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. Hum Mol Genet. 2016;25(3):609–19.

42. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet [Internet]. 2014;23(R1):R89-98. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25064373

43. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet [Internet]. 2016;48(5):481–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27019110

44. Rakitsch B, Stegle O. Modelling local gene networks increases power to detect trans-acting genetic effects on gene expression. Genome Biol [Internet]. 2016;17:33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26911988

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. 2013/03/07. 2013;56(5):1036–46.

46. Travers ME, Mackay DJ, Dekker Nitert M, Morris AP, Lindgren CM, Berry A, 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 [Internet]. 2013;62(3):987–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23139357

47. Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, et al. Parental origin of sequence variants associated with complex diseases. Nature. 2009/12/18. 2009;462(7275):868–74.

48. Christiansen J, Kolte AM, Hansen T, Nielsen FC. IGF2 mRNA-binding protein 2: biological function and putative role in type 2 diabetes. J Mol Endocrinol [Internet].

2009;43(5):187–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19429674

49. Smith AC, Choufani S, Ferreira JC, Weksberg R. Growth regulation, imprinted genes, and chromosome 11p15.5. Pediatr Res. 2007/04/07. 2007;61(5 Pt 2):43r–47r.

50. Delaval K, Wagschal A, Feil R. Epigenetic deregulation of imprinting in congenital diseases of aberrant growth. Bioessays [Internet]. 2006;28(5):453–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16615080

51. Su R, Wang C, Feng H, Lin L, Liu X, Wei Y, 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 [Internet]. 2016;11(2):e0148399. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26840070

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 Heal. 2015/02/14. 2015;69(7):639–47.

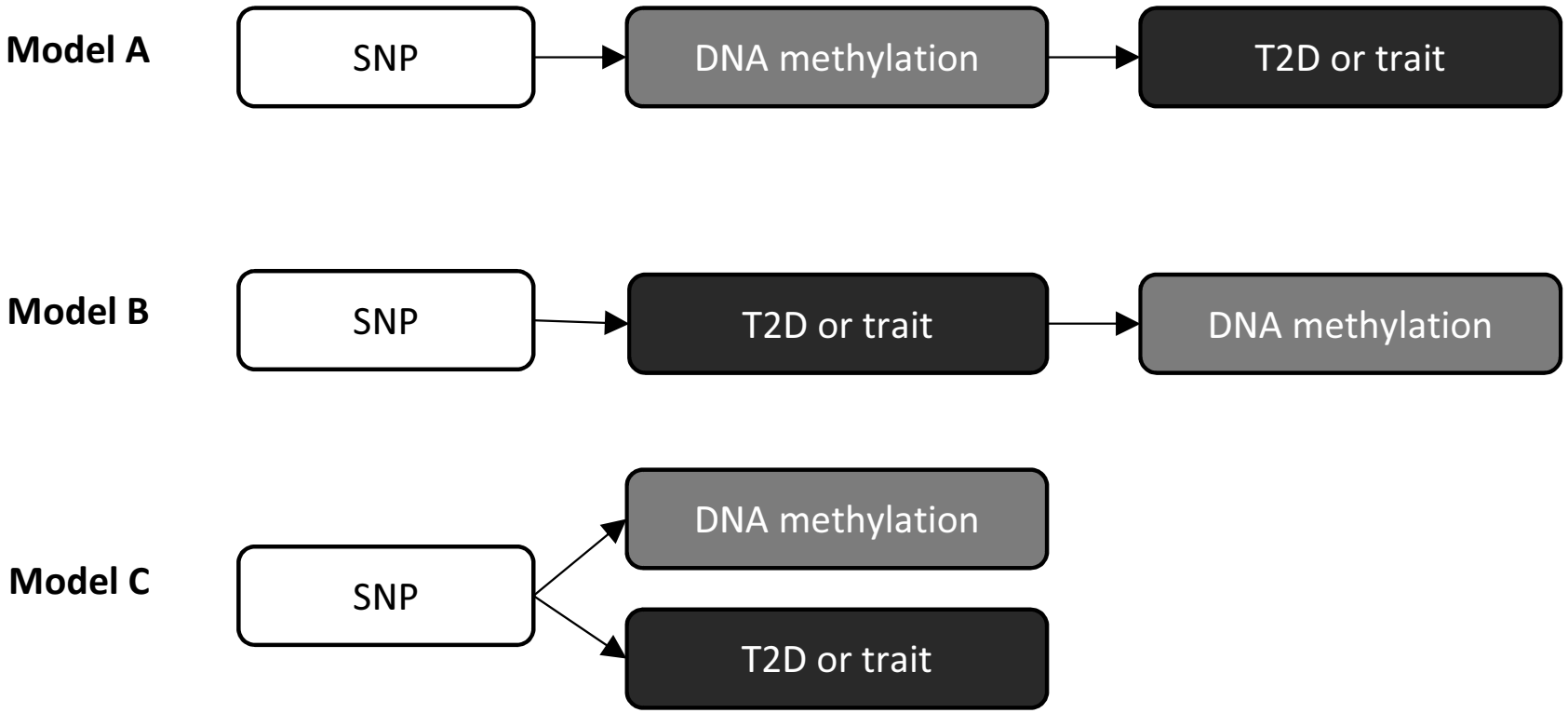
53. Relton CL, Davey Smith G. Mendelian randomization: applications and limitations in epigenetic studies. Epigenomics [Internet]. 2015;7(8):1239–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26639554

54. Ligthart S, Steenaard R V, Peters MJ, van Meurs JB, Sijbrands EJ, Uitterlinden AG, et al. Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. Diabetologia [Internet]. 2016;59(5):998–1006. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26825526

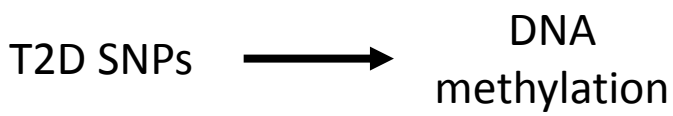
55. Gomez-Uriz AM, Milagro FI, Mansego ML, Cordero P, Abete I, De Arce A, et al. Obesity and ischemic stroke modulate the methylation levels of KCNQ1 in white blood cells. Hum Mol Genet [Internet]. 2015;24(5):1432–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25429063

56. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clin Epigenetics [Internet]. 2014;6(1):4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24485148

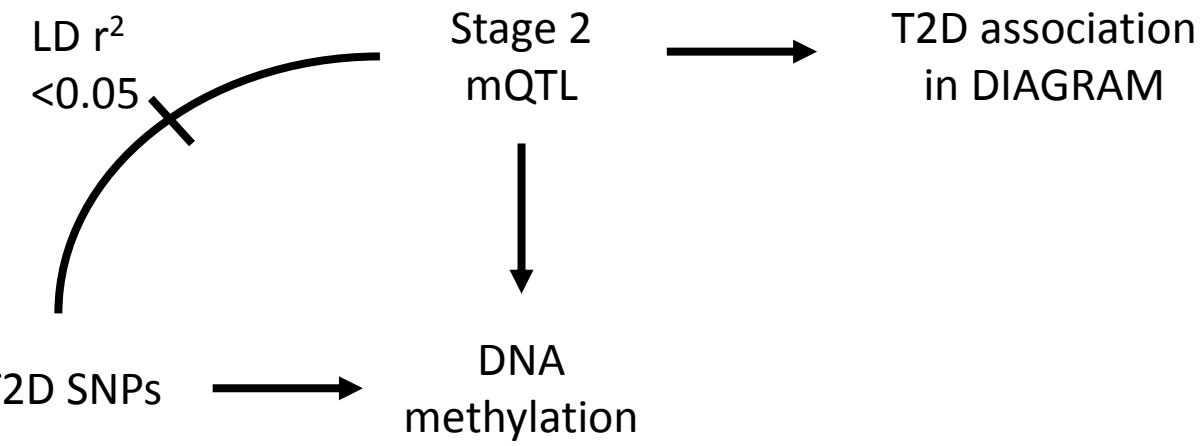
57. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, et al. Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. Am J Hum Genet [Internet]. 2015;97(1):75–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26119815

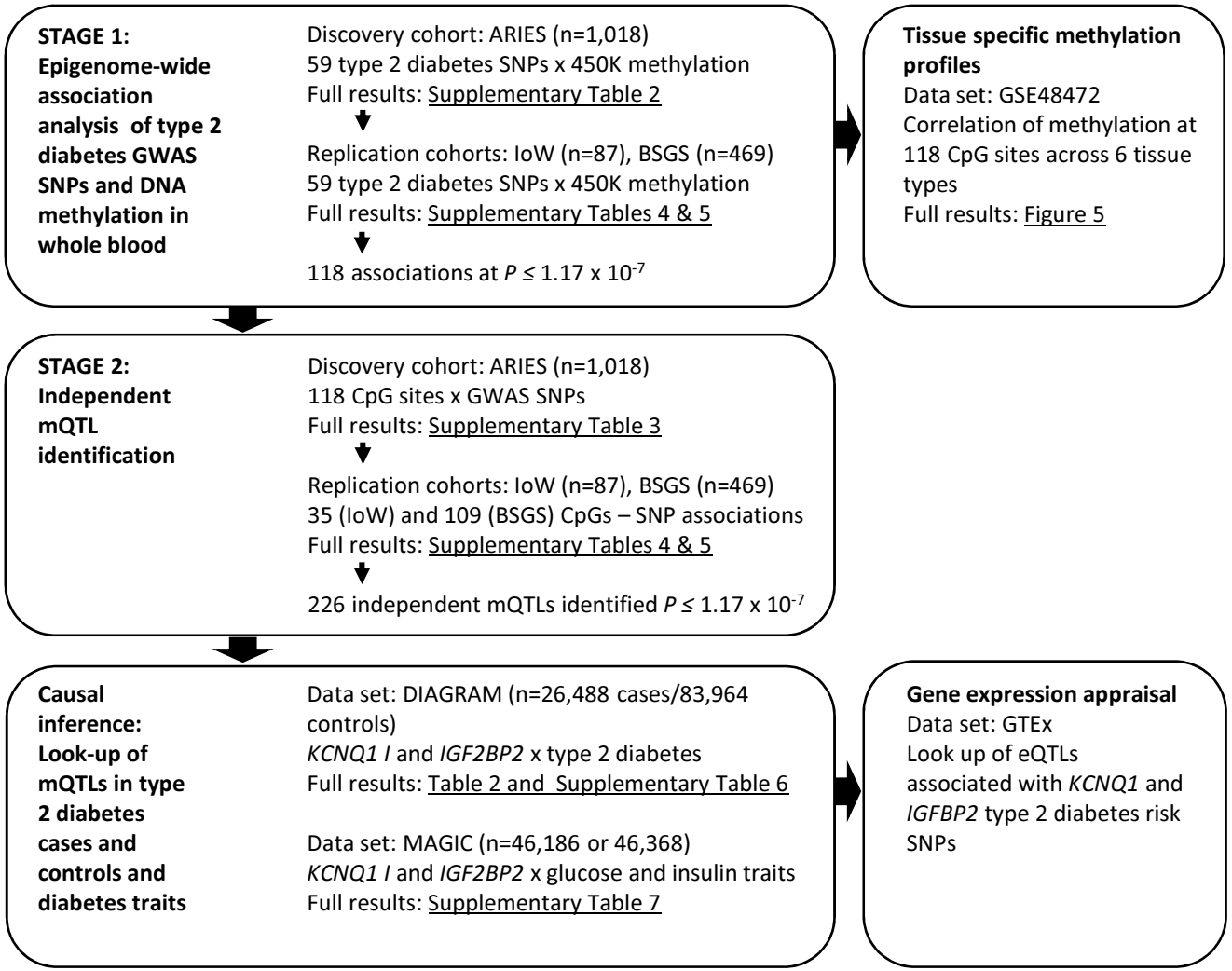

Model A
SNP
DNA methylation
T2D or trait
Model B
SNP
T2D or trait
DNA methylation
Model C
SNP
DNA methylation
T2D or trait

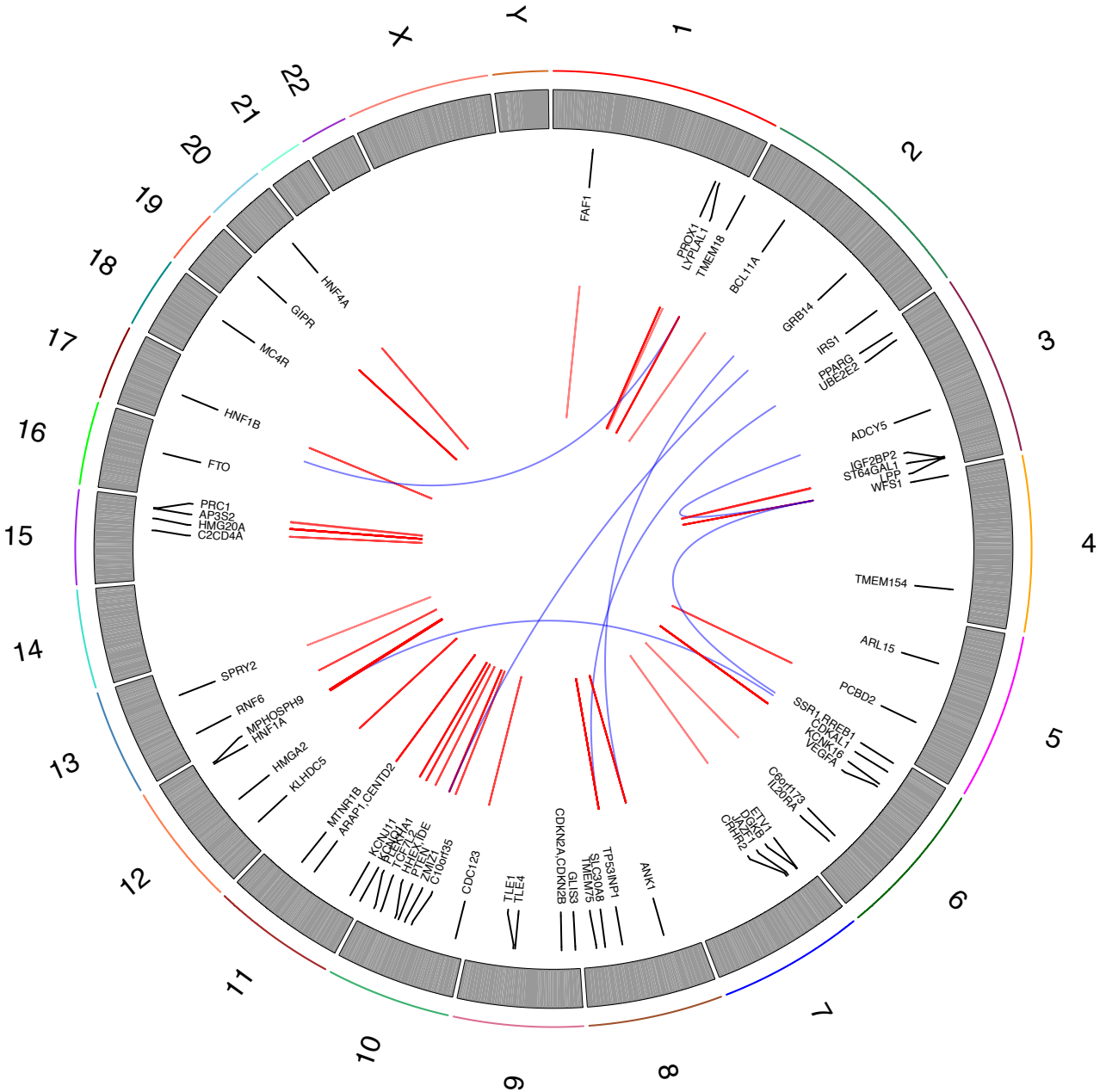
Stage 1

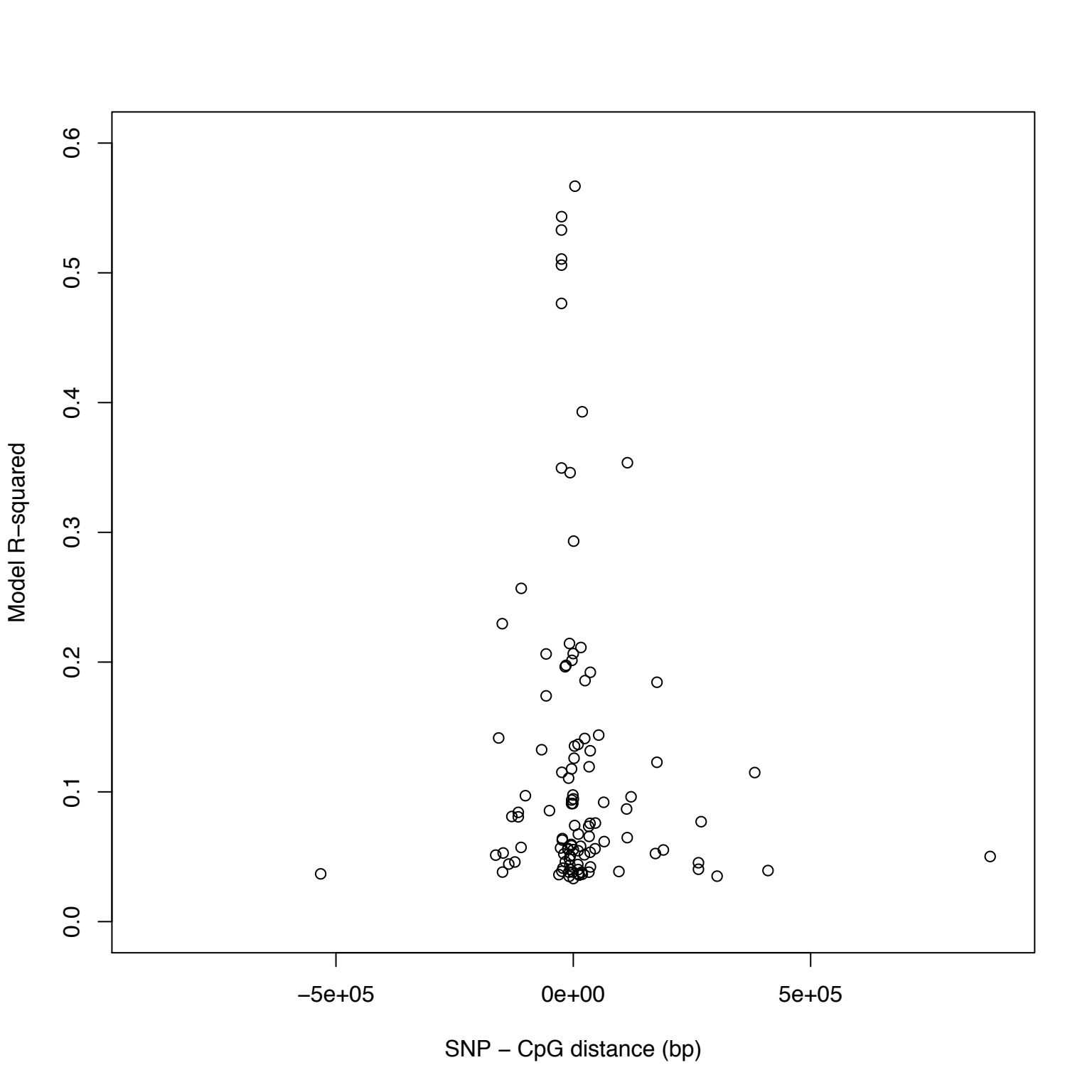


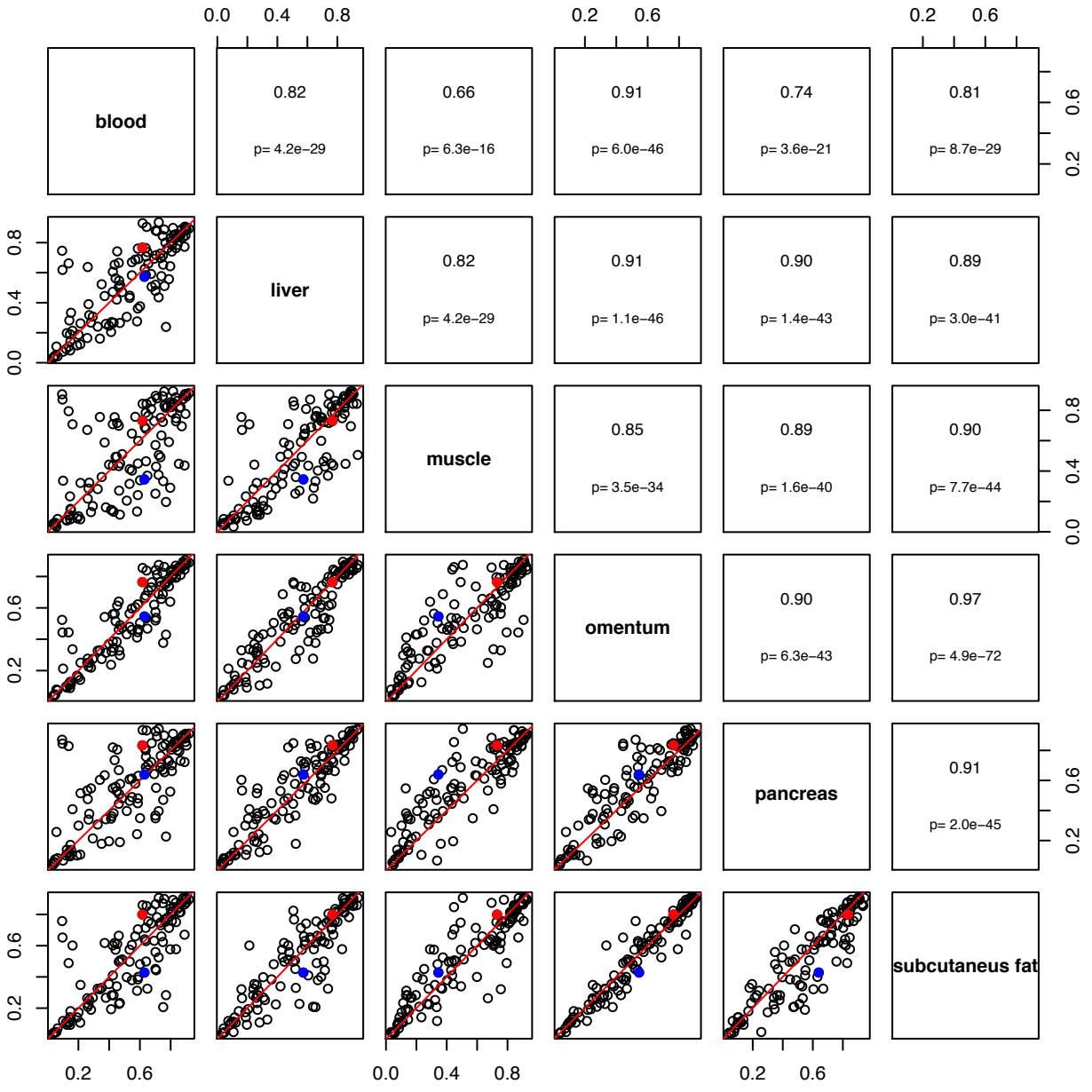
Stage 2




STAGE 1:
Epigenome-wide association analysis of type 2 diabetes GWAS SNPs and DNA methylation in whole blood
Discovery cohort: ARIES (n=1,018)
59 type 2 diabetes SNPs x 450K methylation
Full results: Supplementary Table 2
Replication cohorts: IoW (n=87), BSGS (n=469)
59 type 2 diabetes SNPs x 450K methylation
Full results: Supplementary Tables 4 & 5
118 associations at P ≤ 1.17 x 10-7
Tissue specific methylation profiles
Data set: GSE48472
Correlation of methylation at 118 CpG sites across 6 tissue types
Full results: Figure 5
STAGE 2:
Independent mQTL identification
Discovery cohort: ARIES (n=1,018)
118 CpG sites x GWAS SNPs
Full results: Supplementary Table 3
Replication cohorts: IoW (n=87), BSGS (n=469)
35 (IoW) and 109 (BSGS) CpGs – SNP associations
Full results: Supplementary Tables 4 & 5
226 independent mQTLs identified P ≤ 1.17 x 10-7
Causal inference:
Look-up of mQTLs in type 2 diabetes cases and controls and diabetes traits
Data set: DIAGRAM (n=26,488 cases/83,964 controls)
KCNQ1 I and IGF2BP2 x type 2 diabetes
Full results: Table 2 and Supplementary Table 6
Data set: MAGIC (n=46,186 or 46,368)
KCNQ1 I and IGF2BP2 x glucose and insulin traits
Full results: Supplementary Table 7
Gene expression appraisal
Data set: GTEx
Look up of eQTLs associated with KCNQ1 and IGFBP2 type 2 diabetes risk SNPs


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
FAF1
PROX1
LYPLAL1
TMEM18
BCL11A
GRB14
IRS1
PPARG
UBE2E2
ADCY5
IGF2BP2
ST6GAL1
LPP
WFS1
TMEM154
ARL15
PCBD2
SSR1,RREB1
CDKAL1
KCNK16
VEGFA
C6orf173
THBS2
ETV1
DGKB
CRHR1
ANK1
TP53INP1
SLC30A8
TMEM75
GLIS3
CDKN2A,CDKN2B
TLE1
TLE4
CDC123
KCNJ11
ATP5G1
TCF7L2
HHEX
PTPN11
ZMIZ1
C10orf35
MTNR1B
ARAP1,CENTD2
SPRY2
RNF6
MPHOSPH9
HNF1A
HMGA2
KLHDC5
PRC1
AP3S2
HMG20A
C2CD4A
FTO
HNF1B
MC4R
GIPR
HNF4A


Model R−squared
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−5e+05
0e+00
5e+05
SNP − CpG distance (bp)


blood
liver
muscle
omentum
pancreas
subcutaneus fat
0.82
p= 4.2e−29
0.66
p= 6.3e−16
0.91
p= 6.0e−46
0.74
p= 3.6e−21
0.81
p= 8.7e−29
0.82
p= 4.2e−29
0.91
p= 1.1e−46
0.90
p= 1.4e−43
0.89
p= 3.0e−41
0.85
p= 3.5e−34
0.89
p= 1.6e−40
0.90
p= 7.7e−44
0.90
p= 6.3e−43
0.97
p= 4.9e−72
0.91
p= 2.0e−45
0.0
0.2
0.4
0.6
0.8