Epigenome-wide association study of incident Type 2 diabetes in a British population: EPIC-Norfolk study

Alexia Cardona[1,2]*, Felix R. Day[1], John R.B. Perry[1], Marie Loh[3,4,5], Audrey Y. Chu[6,7], Benjamin Lehne[3], Dirk S. Paul[8], Luca A. Lotta[1], Isobel D. Stewart[1], Nicola D. Kerrison[1], Robert A. Scott[1], Kay-Tee Khaw[9], Nita G. Forouhi[1], Claudia Langenberg[1], Chunyu Liu[6,7,10], Michael M. Mendelson[6,7,10,11], Daniel Levy[6,7], Stephan Beck[12], R. David Leslie[13], Josée Dupuis[6,10], James B. Meigs[14.15.16], Jaspal S Kooner[17, 18,19,20], Jussi Pihlajamäki[21,22], Allan Vaag[23], Alexander Perfilyev[24], Charlotte Ling[24], Marie-France Hivert[25,26] , John C Chambers[3, 17,18, 19, 27], Nicholas J. Wareham[1], Ken K. Ong[1]*

1. MRC Epidemiology Unit, University of Cambridge, School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0QQ, United Kingdom.

2. Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, United Kingdom.

3. Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, United Kingdom.

4. Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research, Singapore 138648, Singapore.

5. Department of Biochemistry, National University of Singapore, Singapore 117596, Singapore.

6. National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA, United States of America.

7. The Population Sciences Branch, Division of Intramural Research, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD, United States of America.

8. MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, CB1 8RN, United Kingdom.

9. Department of Public Health and Primary Care, University of Cambridge, United Kingdom.

10. Department of Biostatistics, Boston University School of Public Health, Boston, MA, United States of America.

11. Department of Cardiology, Boston Children's Hospital, Boston, MA, United States of America.

12. Medical Genomics, UCL Cancer Institute, University College London, London WC1E 6BT, United Kingdom.

13. The Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, United Kingdom.

14. Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, United States of America.

15. Harvard Medical School, Boston, MA, United States of America.

16. Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA, United States of America.

17. Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, United Kingdom.

18. Imperial College Healthcare NHS Trust, London W12 0HS, United Kingdom.

19. MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, United Kingdom.

20. National Heart and Lung Institute, Imperial College London, London W12 0NN, United Kingdom.

21. Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Joensuu, Finland.

22. Clinical Nutrition and Obesity Center, Kuopio University Hospital, Kuopio, Finland.

23. Cardiovascular and Metabolic Disease (CVMD) Translational Medicine Unit, Early Clinical Development, IMED Biotech Unit, AstraZeneca, Gothenburg, Sweden.

24. Epigenetics and Diabetes Unit, Department of Clinical Sciences, Lund University Diabetes Centre, Scania University Hospital, Malmö, Sweden.

25. Department of Population Medicine, Harvard Medical School, Harvard Pilgrim Health Care Institute, Boston, MA, United States of America.

26. Massachusetts General Hospital, Boston, MA, United States of America.

27. Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore.

*Correspondence to Alexia Cardona (alexia.cardona@mrc-epid.cam.ac.uk, +44 (0)1223 330315) and Ken Ong (ken.ong@mrc-epid.cam.ac.uk, +44 (0) 1223 330315).

Abstract

Epigenetic changes may contribute substantially to risks of diseases of ageing. Previous studies reported seven methylation variable positions (MVPs) robustly associated with incident type 2 diabetes mellitus (T2DM). However, their causal roles in T2DM are unclear. In an incident T2DM case-cohort study nested within the population-based EPIC-Norfolk cohort, we used whole blood DNA collected at baseline, up to 11 years before T2DM onset to investigate the role of methylation in the aetiology of T2DM. We identified 15 novel MVPs with robust associations with incident T2DM, and robustly confirmed three MVPs identified previously (near to *TXNIP*, *ABCG1* and *SREBF1*). All 18 MVPs showed directionally consistent associations with incident and prevalent T2DM in independent studies. Further conditional analyses suggested that the identified epigenetic signals appear related to T2DM via glucose and obesity-related pathways acting before the collection of baseline samples. We integrated genome-wide genetic data to identify methylation-associated quantitative trait loci robustly associated with 16 of the 18 MVPs, and found one MVP, cg00574958 at *CPT1A*, with a possible direct causal role on T2DM. None of the implicated genes was previously highlighted by genetic association studies, suggesting that DNA methylation studies may reveal novel biological mechanisms involved in tissue responses to glycemia.

Introduction

Type 2 diabetes mellitus (T2DM) is a major and increasing public health problem. Genome-wide studies have identified more than 240 genetic variants[1] that are robustly associated with T2DM. However, these only explain a minor portion of T2DM susceptibility variance[2,3]. Environmental factors, including diet and physical activity, and also early life factors during fetal and early postnatal development are reported to contribute to the aetiology of T2DM. Epigenetic variation can occur as a result of genetic and/or environmental factors[4]. DNA methylation (DNAm) at cytosine-guanine dinucleotides (CpG sites) is the most commonly studied epigenetic mechanism to date, due to its role in expression regulation and available assays to quantify DNAm intensity at multiple sites across the epigenome that are applicable to large scale studies. Unlike genotypic variation, DNAm intensity patterns are liable to change over time, with age or following disease or other exposure, and therefore disease-associated changes may be either causal or consequential[5].

Previous epigenome-wide association studies (EWAS) have identified seven methylation variable positions (MVPs) that are significantly associated ($P<1{\cdot}0\text{x}10^{-7}$) with incident T2DM[6,7]. However, the causal role of those markers in T2DM is unclear. Here, we aimed to elucidate DNAm determinants of T2DM by performing an EWAS for incident T2DM in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Norfolk study[8]. By further integrating genome-wide genetic array data, we aimed to identify methylation quantitative trait loci (methQTLs) for any T2DM-associated MVPs, in order to assess the likely causal role of DNAm markers on T2DM through Mendelian randomization analyses[9].

Methods

Cohort descriptions

The discovery phase EWAS was undertaken in an incident T2DM case-cohort study nested within the EPIC-Norfolk study[8], a prospective cohort study that recruited 25,639 individuals aged between 40-79 years at baseline in 1993-1997. The cohort was representative of the general population of England and Wales for age, sex, anthropometric measures, blood pressure and serum lipids, but differed in that 99.7% of the cohort were of European descent. We defined a random sub-cohort of the whole EPIC-Norfolk study population excluding known prevalent cases of diabetes at baseline using the same definitions as used in the InterAct project[10] who had available genotype data. Incident T2DM cases were ascertained from multiple sources: two follow-up health and lifestyle questionnaires providing self-reported information on doctor-diagnosed diabetes or medications; medications brought to the second clinical exam; and medical record linkage. Record linkage to external sources included the listing of any EPIC-Norfolk participant in the general practice diabetes register, local hospital diabetes register, hospital admissions data with screening for diabetes-related admissions, and Office of National Statistics mortality data with coding for diabetes. Participants who self-reported a history of diabetes which could not be confirmed against any other sources were not considered as confirmed cases. Follow-up was censored at date of diagnosis of T2DM, 31 July 2006, or date of death, whichever came first. By definition in case-cohort design, there are cases within- and outside- the random sub-cohort but for the purposes of this analysis, we considered them in the incident case set only, with non-cases forming the comparison group. BMI and HbA1c levels were measured for each participant at baseline (**Table 1**). All participants in the EPIC-Norfolk study gave signed informed consent and the study was approved by the Local Research Ethics Committee.

Confirmation of top signals from the discovery EWAS was sought in two further studies. The London Life Sciences Prospective Population (LOLIPOP) study is a prospective population study of Indian Asian (*N*=17,606) and European (*N*=7,766) individuals, recruited at age 35–75 years from the lists of 58 family doctors in west London, UK, between May 1, 2002, and September 12, 2008. Indian Asians

had all four grandparents born on the Indian subcontinent (India, Pakistan, Sri Lanka, or Bangladesh). The LOLIPOP study is approved by the National Research Ethics Service (07/H0712/150) and all participants gave written informed consent at enrolment. The LOLIPOP nested case-control study of incident T2DM has been previously described[6]. Briefly, at follow-up, on Dec 31, 2013, individuals with T2DM were identified by primary care electronic health records and structured queries. Participants with incident T2DM were defined as those who did not have T2DM at baseline, but who developed the disease during follow-up. Controls were identified from a random subset of 7640 participants who attended a clinical assessment of fasting blood glucose concentration and HbA_{1c} and questionnaire assessment between Jan 11, 2010, and Dec 31, 2013.

The Framingham Heart Study (FHS) is a community-based longitudinal study of participants living in and near Framingham, MA, at the start of the study in 1948[11]. The Offspring cohort comprised the children and spouses of the original FHS participants, as described previously[12]. Briefly, enrolment for the Offspring cohort began in 1971 ($N = 5{,}124$), and in-person evaluations occurred approximately every 4 to 8 years thereafter. The current analysis was limited to participants from the Offspring cohort who survived until the eighth examination cycle (2005 to 2008) and consented to genetics research. DNAm data of peripheral blood samples collected at the eighth examination cycle were available in 2,741 participants. Prevalent T2DM was defined as having fasting glucose ≥ 7 mmol/L or as reporting taking T2DM medication at any examination cycle, up to the eighth examination. All participants provided written informed consent at the time of each examination visit. The study protocol was approved by the Institutional Review Board at Boston University Medical Center (Boston, MA, USA).

Methylation array profiling

In all studies, DNAm intensity was measured using the Illumina HumanMethylation450 array (12-sample array for FHS, 96-sample array for EPIC-Norfolk and LOLIPOP). Bisulfite conversion of DNA was performed using the EZ DNA methylation kit (Zymo Research, Orange, CA, USA).

For 1,378 EPIC-Norfolk participants, DNAm was measured in DNA extracted from whole blood samples collected at baseline. Converted DNA was assayed by PCR (Polymerase Chain Reaction) and gel electrophoresis. Each 96 well DNA sample plate contained two duplicate samples. The average correlation between the duplicate samples was 98%.

In LOLIPOP, DNAm was measured among the first 1,074 Indian Asian participants with incident T2DM and 1,590 matched Indian Asian controls. Controls were matched to cases by age (5-year groups) and sex. DNAm was quantified in the baseline DNA samples collected at study enrolment. Samples were analysed in random order, masked to case-control status.

In FHS, peripheral blood samples were collected at the eighth examination (2005 to 2008). Genomic DNA was extracted from buffy coat using the Gentra Puregene DNA extraction kit (Qiagen). Bisulphite converted DNA samples were hybridised to the 12 sample Illumina HumanMethylation450 array using the Infinium HD Methylation protocol and Tecan robotics (Illumina, San Diego, CA, USA). DNAm quantification was conducted in two laboratory batches.

EWAS quality control and normalisation

In EPIC-Norfolk, epigenome-wide DNAm data were analysed in R (version 3.2.2). Initial quality control was performed as recommended by the array manufacturer; methylation intensity values were corrected using the Illumina Background Correction algorithm as implemented in *minfi*[13], methylation intensities with a detection *P*-value ≥0·01 were set to 'missing' and methylation intensity beta values were calculated for each methylation marker per sample. For duplicate samples, the sample with the lower CpG detection percentage was excluded.

Sample call rates were calculated as the proportion of missing data in each sample, by autosomal, X and Y chromosomes. For the autosomal data, 77 samples with a call rate ≤0·99 were excluded. All samples passed the call rate threshold on the X chromosome. For the Y chromosome, seven male samples that did not pass the call-rate and two further female samples were excluded. Distributions of methylation intensities were also inspected by autosomal and sex chromosomes, and separately in

females and males leading to the exclusion of two additional samples that had an unusual distribution of methylation intensities. After those quality control procedures, data on 1,290 samples remained. All further downstream analyses were restricted to autosomal methylation markers.

Marker call rates were calculated as the proportion of missing data at each CpG site. 8,775 CpGs with a call rate ≤0·95 were excluded. The R package *ENmix*[14] was used to identify CpG sites with multimodal distributions of methylation intensity, which typically arise from technical artefacts; 3,295 such CpG sites were excluded. A further 18,874 CpG sites with probes previously identified as mapping to more than one genomic location were also excluded[15].

To ensure reliability of the data, filtering on sample and marker call rates were repeated until all samples and all markers passed their respective call rate thresholds. After excluding prevalent T2DM cases at baseline, the final dataset comprised 1,264 samples (563 incident T2D cases, including 22 cases from the subcohort, and 701 non-cases) with methylation intensities at 442,920 autosomal CpG sites. Quantile normalisation of methylation intensity beta values was applied separately to the different sub-groups of markers based on colour channel, probe type and methylated/unmethylated subtypes as proposed by Lehne *et al.*[16]

In LOLIPOP, DNAm data were analysed in R (version 2·15) using *minfi*[13] and other R scripts. Marker intensities were normalised by quantile normalisation as previously described[6].

In FHS, DNAm data were normalized using the DASEN methodology implemented in the wateRmelon package[17]. Sample exclusion criteria included poor SNP matching of control positions, missing rate >1%, outliers from multi-dimensional scaling, and sex mismatch. Probes were excluded if missing rate >20%. Data from laboratory batches were pooled leaving up to 2,635 samples and 443,304 CpG probes for analysis. Additional information on DNAm, normalization and quality control is available in Asbeykian *et al*.[18]. Differences in DNAm data generation, quality control and statistical models are summarised in **Table S1**.

EWAS statistical analyses

In EPIC-Norfolk, to identify MVPs associated with incident T2DM, we performed a logistic regression model for each methylation marker with incident T2DM status, adjusted for age, sex, estimated cell counts, and sample plate using the EWAC pipeline. A conservative multiple test corrected *P*-value threshold was applied ($P<1x10^{-7}$). Different methylation profiles have been observed between the different cell types in whole blood[19] and blood-based profile of DNAm was shown to predict the underlying distribution of cell types[20]. To correct for cell composition variability[21], first the proportions of different cell types (CD4+, CD8+ T cell subtypes, natural killer cells, monocytes, granulocytes and B cells) were estimated from DNAm data using the algorithm described by Houseman *et al.*[22] as implemented in the R package *minfi*[13]. These cell count estimates were then used as covariates in the epigenome-wide regression models for incident T2DM.

We used STRING[23] to perform gene-set enrichment on the significant genes associated with the 18 significant MVPs identified in the EWAS. We also performed a modified version of the MAGENTA[24] pipeline to identify the pathways associated with genes at the loci of the significant MVPs. Since MAGENTA uses SNP data to identify loci, we assigned to each CpG a "nearest SNP" based on HapMap3 data and using build 36 positions for both the CpG site and the SNPs (average distance to the nearest SNP=4,175 base pairs, IQR=1,375-4,859; 1,707 out of 466,039 CpGs were not assigned a SNP). In effect, rather than using a SNP *P*-value to rank genes to assess enrichment we use the *P*-value from the methylation site to run MAGENTA.

For LOLIPOP, an epigenome-wide association of DNAm was performed in Indian Asians with incident T2DM who were identified from the 8-year follow-up of the study. Differential white blood cell (lymphocyte, monocyte, and granulocyte) count was available for all participants, and epigenome-wide methylation scores were used to impute a further four lymphocyte subsets (CD4, CD8, natural killer, and B cells). Principal components analysis was performed to quantify latent structure in the data, including batch effects. Associations between incident T2DM and the 18 significant MVPs identified in EPIC-Norfolk were tested using logistic regression including intensity

values from Infinium 450K assay control probes, bisulfite conversion batch, measured white cells and imputed white cell subsets, and the first five principal components as covariates. Association results were corrected for the genomic control inflation factor. For testing the predictive ability of the 18 markers for incident T2DM, univariate logistic regressions were run for each of the 18 markers to obtain individual effect sizes (betas) for incident T2DM. A weighted methylation risk score (MRS) was subsequently calculated from these betas, and receiver operating curve (ROC) analyses were performed to provide estimates for area under the curve (AUC).

In FHS, association between each identified MVP (associated with incident T2DM in EPIC-Norfolk) was tested for association with prevalent diabetes and glycemic traits (fasting glucose, fasting insulin, HbA1c). The analysis of glycemic traits included only non-diabetic individuals. Fasting insulin was natural log-transformed. Random effects statistical models were used to analyze the data to account for sibling correlation and included adjustments for age, sex, white blood cell counts, technical covariates, batch effects and BMI, with DNAm as the dependent variable.

We also examined each T2DM-associated MVP for additional cross-sectional association with Type 1 diabetes (T1DM) in an earlier EWAS of 52 monozygous twin pairs discordant for T1DM, in cell-sorted peripheral blood mononuclear cells (monocytes, B cells or T cells)[25] . As T2DM and T1DM have largely differing aetiologies, MVPs that are consistently associated with both outcomes may indicate metabolic effects of diabetes on DNAm.

Other tissues

The relevance of changes in DNAm intensity in whole blood to other tissues was tested by analyzing genome-wide DNAm data, generated using the Illumina HumanMethylation450 array, from human liver, adipose tissue, and skeletal muscle, as previously published[26]. Human liver DNAm data were from participants of the Kuopio Obesity Surgery Study (KOBS); 35 with T2D and 60 without[27]. Data on adipose tissue (14 pairs), skeletal muscle (17 pairs) and blood (19 pairs) were from monozygotic twins discordant for T2DM[26,28,29]. Adipose tissue and skeletal muscle from the same individual were available for most of these twin pairs (16 pairs in blood/muscle 14 pairs in blood/fat); concordance in

DNAm intensity across these tissues was tested for each highlighted MVP by Spearman correlation tests. We further tested cross-tissue correlations in DNAm at T2DM-associated MVPs between blood and other tissues of relevance to T2DM aetiology, liver and pancreas, in publicly available 450K methylation array data from 6 cadavers sampled within 12 hours post-mortem (mean age 65.5 years, SD = 7.2)[30] .

Mendelian randomization analyses

We performed bi-directional Mendelian randomization analyses to test whether any T2DM-associated MVP had a causal effect on T2DM or are a consequence of metabolic differences that had originated before the baseline measurement in this study. To predict the causal effect of each of T2DM-associated MVP on T2DM, methQTLs associated with each MVP (FDR <0·05) in whole blood in 3,841 adults of European descent were identified using the BIOS QTL browser[31]. To run Mendelian randomization analyses, the Z-score for each methQTL was converted to *beta* and standard error using the formulas[32]:

$$beta = \frac{Z}{\sqrt{N \times 2 \times MAF \times (1 - MAF)}}$$

$$SE = \frac{1}{\sqrt{N \times 2 \times MAF \times (1 - MAF)}}$$

where *N* is the sample size and *MAF* is the minor allele frequency. We then tested these methQTLs in Mendelian randomization analyses[9] for T2DM. Genetic associations with T2DM were estimated in 69,677 cases and 551,081 controls from the UK Biobank study[33], the EPIC-InterAct study[10] and the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) consortium[2]. A summary statistics method (inverse variance weighted, IVW) that combines all the SNPs for each MVP as a genetic instrument was used to predict the effect of that MVP on T2DM[34]. To ensure that the instruments are independent, clumping was performed. MR-Egger regression was also used to assess the sensitivity of the results to violations of Mendelian randomization assumptions. Mendelian randomization analyses were run using the R package *TwoSampleMR*[35].

For the reverse direction causal assessment, we tested SNPs with previously reported associations with T2DM[2] or related metabolic phenotypes (BMI[36], fasting glucose[37], 2-hour glucose[38], fasting insulin[39], fasting insulin adjusted for BMI[37], insulin resistance[40], insulin secretion[41] and waist-hip-ratio adjusted for BMI[42]) to test whether these traits have causal effects on methylation intensity at any T2DM-associated MVP. We used summary statistics methods (IVW and Egger's tests) that combine all the SNPs for each trait as a genetic instrument to predict the effect of that trait on each T2DM-associated MVP[34] in the cohort control samples of the EPIC-Norfolk (*N*=613) in whom genotype data were generated using the Affymetrix Axiom UK Biobank chip. All genotypes passed standard QC criteria as specified by the Affymetrix Best Practices pipeline and SNPs with MAF<5% in this sample were excluded.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. AC, FD, JRBP, NJW and KKO had full access to all of the data in the study and AC and KKO had final responsibility for the decision to submit for publication.

RESULTS

MVPs associated with incident T2DM

In the EPIC-Norfolk study, we identified 18 MVPs that are associated with incident T2DM at $P<1x10^{-7}$, including 15 novel signals (**Table 2**). None of these was reported to have a SNP on the target CpG[15]. The two strongest associations were the previously reported signals at *TXNIP* (cg19693031; $P=2.7x10^{-21}$), *ABCG1* (cg06500161; $P=6.4\times10^{-14}$)[6,7]. We confirmed a third previously reported signal at *SREBF1* (cg11024682; $P=6.0x10^{-10}$), and provide supportive evidence for an additional signal at *PROC* (cg09152259; $P=4.2x10^{-4}$) that had previously not been considered to be true due to lack of replication in Europeans **(Table S2).**

We sought confirmation of the top 18 MVPs in data on 1,074 incident T2DM cases and 1,590 control samples from the LOLIPOP study and in cross-sectional data from FHS (403 with prevalent T2DM

and 2,204 controls) **(Table 3)**. All 18 MVPs showed directionally consistent associations with incident T2DM (14 at $P<0{\cdot}05$) and prevalent T2DM (16 at $P<0{\cdot}05$).

Novel MVPs associated with incident T2DM include cg14476101 ($P=2{\cdot}8\text{x}10^{-10}$), located in the gene body of *PHGDH* which encodes phosphoglycerate dehydrogenase, an enzyme involved the synthesis of L-serine and other amino acids, and cg00574958 ($P=5.2\text{x}10^{-9}$) in the 5'UTR of *CPT1A* which encodes an enzyme that initiates mitochondrial oxidation of long-chain fatty acids (**Table S11).** Four of the 18 MVPs were located within solute carrier family genes (*SLC1A5*, *SLC43A1*, *SLC9A1* and *SLC9A3R1*), which encode plasma membrane proteins that regulate cell transport of amino acids and other metabolites.

To systematically explore the biological pathways implicated by T2DM-associated methylation signals, we first tested the 18 MVPs for gene set enrichment using STRING and identified significant enrichment for three pathways: "*positive regulation of cholesterol biosynthetic process*" (indicated by MVPs at *ABCG1*, *SREBF1* and *POR*), "*carnitine metabolic process*" (indicated by *CPT1A* and *POR*) and "*AMPK signalling*" (indicated by *PFKFB3*, *CPT1A* and *SREBF1*). We then tested the full EWAS dataset in a modified MAGENTA pipeline and identified significant enrichment for T2DM-associated methylation signals in 10 pathways (**Table S4**), including "*insulin receptor signalling*", "*IGF-1 signalling*", "*Erythropoietin signalling*", "*JAK signalling*" and "*Integrin signalling*".

MVPs associated with glycemic traits

In non-diabetic control FHS samples, all 18 T2DM-associated MVPs showed directionally concordant associations with fasting glucose, fasting insulin levels and BMI, and 16 of the 18 MVPs showed directionally concordant associations with HbA1c (**Table S5**). In additional conditional models in the EPIC-Norfolk discovery sample, the associations between all individual 18 MVPs with incident T2DM were markedly attenuated when models were further adjusted for baseline BMI and HbA1c (median attenuation 49%, **Table S3)**, indicating that these DNAm intensity changes largely reflect baseline differences between future incident T2DM cases and other cohort participants.

Furthermore, among 52 monozygous twin pairs discordant for Type 1 diabetes (T1DM), 7 of the 18 T2DM-associated MVPs showed cross-sectional differences in DNAm intensity in peripheral white blood cells (monocytes, B cells or T cells) between the T1DM-affected and unaffected twin, consistent with an effect of glycemia on DNAm intensity (at *TXNIP*, *SLC9A3R1*, *SREBF1*, *CPT1A*, *C7orf50*, *PFKFB3* and cg08309687) (**Table S6**).

Relevance of whole blood MVPs to other tissues

To explore the possible relevance of changes in DNAm intensity in whole blood to other tissues, relevant to T2DM pathogenesis, we examined these 18 MVPs in liver, adipose tissue, and skeletal muscle from individuals with and without T2DM. Nominal associations (P<0.05) were found with only our 2 strongest whole blood MVP signals: cg06500161 at *ABCG1* in adipose tissue (as previously published[26] and cg19693031 at *TXNIP* in skeletal muscle (Table 4). Furthermore, at 12 of the 18 MVPs there was evidence for a positive correlation in DNAm intensity between whole blood and liver, pancreas, adipose tissue or muscle (**Table S7**).

Causal effects on T2DM

To investigate the potential causal effects of the 18 T2DM-associated MVPs, we used the BIOS QTL browser[31] to identify methQTLs (genetic sequence variants) that are robustly associated (at $P<5x10^{-8}$) with DNAm intensity at any of the 18 MVPs. We found 54 methQTLs (33 *cis*, 21 *trans*) each associated with one of 16 MVPs (**Table S8**). We then used these methQTLs as instrumental variables in Mendelian randomization analyses, based on aggregated publicly-available GWAS data in 69,677 T2DM cases and 551,081 controls (DIAGRAM[2], UK Biobank[33] and EPIC-InterAct[10]). Only one of the 16 T2DM-associated MVPs with an identified methQTL showed nominal evidence for a direct causal association with T2DM, cg00574958 at *CPT1A* (P=0.01), however, for other MVPs the genetic-predicted effects overlapped with the observed effects in the LOLIPOP study **(Figure 1, Table S9)**.

We performed reverse direction causal analyses, to identify causal effects of BMI and glycemic traits on methylation intensity at the 18 MVPs. Among non-T2DM participants in EPIC-Norfolk (*N*=613), none of the genetic instruments for the tested glycemic or metabolic traits (T2DM, BMI, fasting glucose, 2-hour glucose, fasting insulin, fasting insulin adjusted for BMI, insulin resistance, insulin secretion and waist-hip-ratio adjusted for BMI) showed a consistent association with any of the 18 T2DM-associated MVPs (**Table S10**).

Prediction of T2DM

In the LOLIPOP study sample, which was independent of the discovery EWAS, the top 18 T2DM-associated MVPs in aggregate showed no predictive ability for incident T2DM (AUC=0·53). Furthermore, the addition of these 18 MVPs did not improve on a prediction model based on other baseline phenotypes (BMI, HbA1c, age, sex: AUC=0·761; BMI, HbA1c, age, sex, plus 18 MVPs: AUC=0·762).

Discussion

In this prospective study, we substantially increased the number of MVPs in whole blood that are robustly associated with incident T2DM. Associations for 17 of the 18 MVPs were confirmed with either incident or prevalent T2DM in two independent studies, which indicates the consistency of T2DM-associated whole blood DNAm intensity changes across different settings and ethnicities. Genetic causal modelling identified evidence to support a causal effect of DNAm on T2DM at one of these MVPs, cg00574958 at *CPT1A*.

The prospective designs of the EPIC-Norfolk and LOLIPOP studies aimed to identify MVPs that precede the development of T2DM. However, the identified T2DM-associated DNAm intensity changes were largely attenuated by adjustment for differences in BMI and glycemia that had developed prior to the baseline measurement in the prospective studies. Our Mendelian randomization analyses failed to find evidence for direct causal effects for the majority of T2DM-associated, as indicated by no detectable genetic-predicted effect of DNAm intensity on T2DM, and a wide discordance between the observed and genetic-predicted effects. Conversely, overlap between EWAS signals for T2DM and T1DM are consistent with effects of glycemia on DNAm intensity for at least 7 of the 18 T2DM-associated MVPs.

Whether or not they show directly causal associations, these novel and consistent T2DM-associated MVPs are highly informative with regard to implicated genes and biological pathways. Notably, none of the genes implicated by this EWAS was previously identified by genetic variant association studies. This stark difference may suggest that T2DM-associated DNAm intensity changes may reveal novel biological mechanisms involved in tissue responses to glycemia rather than in the pathogenesis of insulin resistance or insulin secretion, which are implicated by those genetic studies. The highest signal, cg19693031 which lies on *TXNIP*, is also the most significant observation in other T2DM EWAS studies[6,7]. Phosphoglycerate dehydrogenase (PHGDH) catalyzes the first and rate-limiting step in glucose-derived serine synthesis and may indicate consequent purine and deoxythymidine nucleotide synthesis in response to hyperglycemia and potential tissue proliferative responses[43]. Functional variation in carnitine palmitoyltransferase 1 (CPT1A) regulates the composition of

circulating polyunsaturated n-3 fatty acids and docosahexaoenic acid[44], is reported to activate lipolysis and mitochondrial activity in brown fat[45,46], and to maintain pancreatic islet secretion of the principal hyperglycemic hormone, glucagon[47]. Solute carrier family members are sodium-dependent membrane transporters that regulate intracellular cell pH, cell volume, and other cellular events such as adhesion, migration, and proliferation, and also contribute to systemic homeostasis of fluid volume, acid-base balance and electrolytes. Specifically, SLC9A3R1 (NHERF1) binds to PTEN to activate the PI3 kinase signaling cascade involved in cell survival, growth, proliferation[48], and is a key component of insulin and IGF-1 signalling pathways that we found enriched for T2DM EWAS associations. These highlighted pathways could potentially contribute to the pathogenesis of micro- and macro-vascular complications of hyperglycemia. PFBK3, a regulator of glycolysis and insulin signalling in mice, was recently highlighted by a SNP association with late onset autoimmune diabetes, and we here provide independent evidence to support its role in human glucose regulation[49].

We recognize a number of limitations of our study. Both of the prospective study samples displayed large differences in baseline glycemia and BMI between incident T2DM cases and non-cases. This nested prospective study design aimed to identify interactions between genetic factors and baseline lifestyle factors measured prior to the development of clinically-diagnosed T2DM[10]. Since it is impossible to develop T2DM except by passing through a phase of non-diabetic hyperglycemia, it is inevitable that people who go on to get incident diabetes in a cohort study will have raised glucose levels at baseline if follow up is of short or medium duration. Future studies that have samples stored many years prior to disease onset would be required to identify when in the development of diabetes the T2DM-MVP associations become apparent. Secondly, our assessments of other, non-blood, tissues were limited in the range of tissues and numbers of samples available. Despite concordant changes in DNAm intensity between whole blood and various tissues relevant to T2DM pathogenesis at 12 of the 18 T2DM-associated MVPs, nominal differences in DNAm were found only for our strongest two MVPs, which suggest that larger study samples are needed. We recognize that whole blood is not a tissue of interest to the pathogenesis of T2DM, however current, and most likely future, large-scale EWAS are confined to such samples, and functional insights will depend on follow-up of

whole blood signals in other tissues[50,51]. The same issue of appropriate tissue of interest limits our genetic modelling approach, which identified genetic markers of DNAm intensity in peripheral blood. Furthermore, the sample size for this approach (*N*=3,841 in BIOS QTL[31] and *N*=613 in the EPIC-Norfolk cohort control group) is relatively small compared to data on QTLs for gene expression in peripheral blood (*N*=8,086 in Westra *et al.*[52]). Hence, we found only nominal evidence for a causal effect of DNAm at only one of the 18 T2DM-associated MVPs, at *CPT1A*, and for several MVPs the genetic-predicted effects were overlapping with the observed effects. Similarly, a recent large EWAS for BMI found a causal role of methylation at only one MVP (cg26663590 at *NFATC2IP*)[53]. There are various possible conceptualisations of the functional interplay between SNP, MVP and T2D, which provide alternative explanations other than SNP-to-DNAm-to-T2D[54], but they do not limit the statistical detection of apparent causal signals. Future, larger reference data on QTLs for DNAm intensity in whole blood are being generated (GoDMC), which will allow more powerful tests for causality, although their relevance to DNAm in tissues of interest remains an important question. Finally, the determinants of the identified T2DM-associated MVPs remain unknown. Again, larger reference panels of GWAS and DNAm array data, as well as new methods to integrate findings across multiple methQTLs for each MVP, will inform future causal analyses. Future studies are needed to identify the potential lifestyle and developmental determinants of these T2DM-associated MVPs.

In conclusion, we identified several robust and consistent DNAm markers for incident T2DM. These appear to be related to T2DM via glucose and obesity-related pathways that had their effects before the collection of baseline samples in these cohort studies, which commenced in midlife. These associations indicate several plausible biological mechanisms involved in tissue responses and co-morbidities of hyperglycemia.

Acknowledgements

We are grateful to all of the participants and staff of the EPIC Norfolk, LOLIPOP and Framingham Heart Study cohorts. We thank Dr Stephen Burgess for his advice on methQTLs and Dr Jan Bert van Klinken for his advice on the BIOS-QTL data, Stephen Sharp and Dr Jian'an Luan for their advice on

statistical analyses and Ylva Wessman, Per-Anders Jansson and Emma Nilsson for their help with the Twin study. AC and KKO are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Author Contributors

K-TK, NGF, CL, JD, JBM, JSK, M-FH, JCC, NJW and KKO contributed to the study design.

LAL, NDK, RAS, K-TK, NGF, SB, RDL, CL, MMM, DL, JD, JBM, JSK, M-FH, JCC, NJW and KKO contributed to the data collection.

AC, FD, JRBP, DSP, ML, AYC, ChL, BL and IDS performed data analyses.

AC, KKO, and FD drafted the manuscript. AC constructed the figure.

All authors contributed to data interpretation and revisions of the manuscript.

Declaration of interests

AYC is currently employed by Merck Research Laboratories. The other authors declare no competing interests.

Disclaimer

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

Funding

EPIC-Norfolk is supported by programme grants from the Medical Research Council (MRC) [G9502233; G0401527; G100143] and Cancer Research UK [C864/A8257]. The generation and management of the Illumina 450K methylation array data in this cohort is supported through the MRC Cambridge initiative in metabolomic science [MR/L00002/1]. The genome-wide genotyping data in

EPIC-Norfolk was funded by an MRC award MC_PC_13048. This work is also supported by MRC programme grants [MC_UU_12015/1, MC_UU_12015/2 and MC_UU_12015/5].

The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z, 090532 & 098381) the NIHR (RP-PG-0407-10371), the NIHR Official Development Assistance (ODA, award 16/136/68), the European Union FP7 (EpiMigrant, 279143) and H2020 programmes (iHealth-T2D, 643774). We acknowledge support of the MRC-PHE Centre for Environment and Health, and the NIHR Health Protection Research Unit on Health Impact of Environmental Hazards. The work was carried out in part at the NIHR/Wellcome Trust Imperial Clinical Research Facility. JC is supported by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research Investigator (STaR) Award (NMRC/STaR/0028/2017).

The Framingham Heart Study is supported by grants: N01-HC-25195 and HHSN268201500001I. The laboratory work for this investigation was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, and the NIH Director's Challenge Award (PI: D. Levy). The analytical component of this project was funded by the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the Center for Information Technology, National Institutes of Health, Bethesda, MD. JBM is supported by grants: NIDDK U01 DK078616 and K24 DK080140.

Data on Type 1 diabetes discordant twin pairs arose from studies funded by the EU-FP7 project BLUEPRINT (282510). The Cardiovascular Epidemiology Unit at the University of Cambridge is supported by the UK MRC (MR/L003120/1), BHF (RG/13/13/30194) and National Institute for Health Research (Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Data from human tissues are from studies supported by grants from the Novonordisk foundation, Swedish Research Council, Region Skåne (ALF), EFSD, Exodiab, Swedish Foundation for Strategic Research for IRC15-0067, Swedish Diabetes Foundation and Albert Påhlsson Foundation.

URL

Full summary data from the discovery EWAS for incident T2DM in the EPIC-Norfolk Study are available at: https://www.repository.cam.ac.uk/

BIOS-QTL Browser: http://atlas.bbmrirp3-lumc.surf-hosted.nl/

GoDMC: http://www.godmc.org.uk/

References

1 Mahajan A, Taliun D, Thurner M, *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018; **50**: 1505–13.

2 Morris A, Voight B, Teslovich T. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 2012; **44**: 981–90.

3 Mahajan A, Go MJ, Zhang W, *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014; **46**: 234–44.

4 Bernstein BE, Meissner A, Lander ES. The Mammalian Epigenome. Cell. 2007; **128**: 669–81.

5 Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 2012; **12**: 529–41.

6 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**: 526–34.

7 Soriano-Tárraga C, Jiménez-Conde J, Giralt-Steinhauer E, *et al.* Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. *Hum Mol Genet* 2015; : 1–11.

8 Day N, Oakes S, Luben RN, *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* 1999; **80 Suppl 1**: 95–103.

9 Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. *Int J Epidemiol* 2013; **42**: 1134–44.

10 Langenberg C, Sharp SJ, Forouhi NG, *et al.* Design and cohort description of the InterAct Project: an examination of the interaction of genetic and lifestyle factors on the incidence of type 2 diabetes in the EPIC Study. *Diabetologia* 2011; **54**: 2272–82.

11 Dawber TR, Meadors GF, Moore FE. Epidemiological Approaches to Heart Disease: The Framingham Study . *Am J Public Heal Nations Heal* 1951; **41**: 279–86.

12 Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* 1979; **110**: 281–90.

13 Aryee MJ, Jaffe AE, Corrada-Bravo H, *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 2014; **30**: 1363–9.

14 Xu Z, Niu L, Li L, Taylor JA. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res* 2015; : gkv907-.

15 Naeem H, Wong N, Chatterton Z, *et al.* Reducing the risk of false discovery enabling identification of biologically significant genome-wide methylation status using the HumanMethylation450 array. *BMC Genomics* 2014; **15**: 51.

16 Lehne B, Drong AW, Loh M, *et al.* A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol* 2015; **16**: 37.

17 Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* 2013; **14**: 293.

18 Aslibekyan S, Demerath EW, Mendelson M, *et al.* Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity* 2015; **23**: 1493–501.

19 Baron U, T??rbachova I, Hellwag A, *et al.* DNA methylation analysis as a tool for cell typing. *Epigenetics* 2006; **1**: 55–60.

20 Koestler DC, Christensen BC, Karagas MR, *et al.* Blood-based profiles of DNA methylation predict the underlying distribution of cell types: A validation analysis. *Epigenetics* 2013; **8**: 816–26.

21 Jaffe AE, Irizarry R a. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol* 2014; **15**: R31.

22 Houseman EA, Accomando WP, Koestler DC, *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012; **13**: 86.

23 Szklarczyk D, Franceschini A, Wyder S, *et al.* STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015; **43**: D447–52.

24 Ayellet VS, Groop L, Mootha VK, Daly MJ, Altshuler D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* 2010; **6**. DOI:10.1371/journal.pgen.1001058.

25 Paul DS, Teschendorff AE, Dang MAN, *et al.* Increased DNA methylation variability in type 1 diabetes across three immune effector cell types. *Nat Commun* 2016; **7**:13555. DOI:10.1038/ncomms13555.

26 Dayeh T, Tuomi T, Almgren P, *et al.* DNA methylation of loci within ABCG1 and PHOSPHO1 in blood DNA is associated with future type 2 diabetes risk. *Epigenetics* 2016; **11**: 482–8.

27 Nilsson E, Matte A, Perfilyev A, *et al.* Epigenetic alterations in human liver from subjects with type 2 diabetes in parallel with reduced folate levels. *J Clin Endocrinol Metab* 2015; **100**: jc20153204.

28 Nitert MD, Dayeh T, Volkov P, *et al.* Impact of an exercise intervention on DNA methylation in skeletal muscle from first-degree relatives of patients with type 2 diabetes. *Diabetes* 2012; **61**: 3322–32.

29 Nilsson E, Jansson PA, Perfilyev 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**: 2962–76.

30 Slieker RC, Bos SD, Goeman JJ, *et al.* Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics and Chromatin* 2013; **6**. DOI:10.1186/1756-8935-6-26.

31 Bonder MJ, Luijk R, Zhernakova DV, *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 2017; **49**:131-138.

32 Rietveld CA, Medland SE, Derringer J, *et al.* GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science* 2013; **340**: 1467–71.

33 Allen NE, Sudlow C, Downey P, *et al.* UK Biobank: Current status and what it means for epidemiology. *Heal Policy Technol* 2012; **1**: 123–6.

34 Burgess S, Scott RA, Timpson NJ, Smith GD, Thompson SG. Using published data in Mendelian randomization: A blueprint for efficient identification of causal risk factors. *Eur J Epidemiol* 2015; **30**: 543–52.

35 Hemani G, Zheng J, Wade KH, *et al.* MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. *bioRxiv* 2016; : 78972.

36 Locke AE, Kahali B, Berndt SI, *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**: 197–206.

37 Manning AK, Hivert M-F, Scott RA, *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet* 2012; **44**: 659–69.

38 Saxena R, Hivert M-F, Langenberg C, *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nat Genet* 2010; **42**: 142–8.

39 Scott RA, Lagou V, Welch RP, *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 2012; **44**: 991–1005.

40 Lotta LA, Gulati P, Day FR, *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat Genet* 2016; published online Nov. DOI:10.1038/ng.3714.

41 Prokopenko I, Poon W, Mägi R, *et al.* A Central Role for GRB10 in Regulation of Islet Function in Man. *PLoS Genet* 2014; **10**: e1004235.

42 Shungin D, Winkler TW, Croteau-Chonka DC, *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* 2015; **518**: 187–96.

43 Pacold ME, Brimacombe KR, Chan SH, *et al.* A PHGDH inhibitor reveals coordination of serine synthesis and one-carbon unit fate. *Nat Chem Biol* 2016; **12**: 452–8.

44 Skotte L, Koch A, Yakimov V, *et al. CPT1A* Missense Mutation Associated With Fatty Acid Metabolism and Reduced Height in Greenlanders. *Circ Cardiovasc Genet* 2017; **10**: e001618.

45 Clemente FJ, Cardona A, Inchley CE, *et al.* A Selective Sweep on a Deleterious Mutation in CPT1A in Arctic Populations. *Am J Hum Genet* 2014; **95**: 584–9.

46 Calderon-Dominguez M, Sebastián D, Fucho R, *et al.* Carnitine palmitoyltransferase 1 increases lipolysis, UCP1 protein expression and mitochondrial activity in brown adipocytes. *PLoS One* 2016; **11**. DOI:10.1371/journal.pone.0159399.

47 Linford Briant AJ, Dodd MS, Chibalina M V, *et al.* CPT1a-Dependent Long-Chain Fatty Acid Oxidation Contributes to Maintaining Glucagon Secretion from Pancreatic Islets. *CellReports* 2018; **23**: 3300–11.

48 Takahashi Y, Morales FC, Kreimann EL, Georgescu MM. PTEN tumor suppressor associates with NHERF proteins to attenuate PDGF receptor signaling. *EMBO J* 2006; **25**: 910–20.

49 Cousminer DL, Ahlqvist E, Mishra R, *et al.* First genome-wide association study of latent autoimmune diabetes in adults reveals novel insights linking immune and metabolic diabetes. In: Diabetes Care. 2018: 2396–403.

50 Davegårdh C, García-Calzón S, Bacos K, Ling C. DNA methylation in the pathogenesis of type 2 diabetes in humans. Mol. Metab. 2018; **14**: 12–25.

51 Bacos K, Gillberg L, Volkov P, *et al.* Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat Commun* 2016; **7**. DOI:10.1038/ncomms11089.

52 Westra H-J, Peters MJ, Esko T, *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013; **45**: 1238–43.

53 Wahl S, Drong A, Lehne B, *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* 2016. DOI:10.1038/nature20784.

54 VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. *Epidemiology* 2014; **25**: 427–35.

Table 1: Baseline characteristics of participants the EPIC-Norfolk, LOLIPOP and Framingham Heart study samples

	EPIC-Norfolk		LOLIPOP		FHS	
	Discovery phase		Confirmation phase		Confirmation phase	
	Incident T2DM	Non-cases	Incident T2DM	Non-cases	Prevalent T2DM	Non-cases
N	563	701	1,074	1,590	403	2,204
Sex (F)	474 (84%)	407 (58%)	352 (36.3%)	507 (31.8%)	173 (43.0%)	1245 (56.5%)
Age (years)	61.6 (8.1)	59.1 (9.2)	52.5 (10.2)	49.9 (9.8)	69.3 (8.4)	65.8 (8.9)
Ethnicity	European	European	Indian Asian	Indian Asian	European	European
HbA1c (%)	6.5 (1.3)	5.5 (0.33)	5.77 (0.49)	5.37 (0.48)	6.67 (1.15) %	5.55 (0.27) %
HbA1c (mmol/mol)	47.4 (14.2)	36.2 (3.6)	40 (5.4)	35 (5.2)	49 (12.6)	37 (3)
BMI (kg/m2)	29.2 (4.5)	25.6 (3.6)	28.9 (4.6)	26.7 (3.9)	31.6 (6.2)	27.7 (5.0)

Means (standard deviations) or number (%) are shown.

Table 2: Methylation variable positions associated with incident type 2 diabetes at *P*<1.0E-07 in the EPIC-Norfolk study (*N*=1,264)

CpG ID	Chr	Position	OR	CI 95%	P-value	FDR	Gene name	Gene position
cg19693031	1	144152909	0.52	[0.46-0.6]	2.7E-21	1.3E-15	*TXNIP*	3'UTR
cg06500161	21	42529656	1.65	[1.45-1.89]	6.4E-14	1.5E-08	*ABCG1*	Body
cg14476101	1	120057515	0.67	[0.59-0.76]	2.8E-10	3.9E-05	*PHGDH*	Body
cg14020176	17	70276580	1.63	[1.4-1.9]	3.3E-10	3.9E-05	*SLC9A3R1*	3'UTR
cg11024682	17	17670819	1.56	[1.35-1.79]	6.0E-10	5.7E-05	*SREBF1*	Body
cg06397161	22	38090005	1.51	[1.32-1.73]	4.5E-09	3.3E-04	*SYNGR1*	Body;TSS200
cg00574958	11	68364198	0.69	[0.61-0.78]	5.2E-09	3.3E-04	*CPT1A*	5'UTR
cg06235429	11	67129690	1.49	[1.3-1.7]	5.5E-09	3.3E-04	*NDUFV1*	TSS1500
cg05778424	17	52524507	1.69	[1.42-2.02]	7.4E-09	3.9E-04	*AKAP1*	5'UTR
cg11376147	11	57017774	0.68	[0.59-0.77]	1.3E-08	6.0E-04	*SLC43A1*	Body
cg04816311	7	1033176	1.51	[1.31-1.75]	1.7E-08	7.2E-04	*C7orf50*	Body
cg02711608	19	51979804	0.69	[0.6-0.79]	4.5E-08	1.5E-03	*SLC1A5*	1stExon;5'UTR
cg08309687	21	34242466	0.68	[0.6-0.78]	4.5E-08	1.5E-03		
cg13514042	7	1158728	1.42	[1.25-1.61]	4.5E-08	1.5E-03		
cg08994060	10	6254032	0.65	[0.55-0.76]	5.2E-08	1.6E-03	*PFKFB3*	Body
cg01676795	7	75424284	1.56	[1.33-1.84]	6.5E-08	1.8E-03	*POR*	Body
cg25130381	1	27313308	1.49	[1.29-1.73]	6.7E-08	1.8E-03	*SLC9A1*	Body
cg11183227	15	89256411	1.49	[1.29-1.72]	7.0E-08	1.8E-03	*MAN2A2*	Body

Position: by HapMap Build37. OR: odds ratio per +1 standard deviation in methylation intensity. Genes: Gene names in which the CpG falls between 1500bp upstream of the transcriptional start site to the end of the 3' UTR as in Illumina's HM450 manifest file.

Table 3: Confirmation of the top 18 T2DM-associated MVPs in the LOLIPOP and Framingham Heart studies

			Discovery Incident T2DM		LOLIPOP Incident T2DM			Framingham Heart Study* Prevalent T2DM		
CpG ID	**Chr**	**Gene**	**OR**	**CI 95%**	**OR**	**CI 95%**	**P**	**beta**	**se**	**P**
cg19693031	1	*TXNIP*	0.52	[0.46-0.6]	**0.68**	**[0.62-0.75]**	**1.2E-14**	**-2.6E-02**	**2.7E-03**	**1.6E-21**
cg06500161	21	*ABCG1*	1.65	[1.45-1.89]	**1.44**	**[1.31-1.58]**	**2.6E-14**	**1.5E-02**	**1.8E-03**	**7.1E-17**
cg14476101	1	*PHGDH*	0.67	[0.59-0.76]	**0.81**	**[0.75-0.89]**	**3.0E-06**	**-1.6E-02**	**3.6E-03**	**1.5E-05**
cg14020176	17	*SLC9A3R1*	1.63	[1.4-1.9]	**1.14**	**[1-1.29]**	**4.3E-02**	**5.4E-03**	**1.5E-03**	**3.9E-04**
cg11024682	17	*SREBF1*	1.56	[1.35-1.79]	**1.40**	**[1.26-1.57]**	**2.2E-09**	**8.6E-03**	**1.6E-03**	**5.4E-08**
cg06397161	22	*SYNGR1*	1.51	[1.32-1.73]	**1.17**	**[1.06-1.28]**	**1.1E-03**	**9.6E-03**	**2.2E-03**	**1.6E-05**
cg00574958	11	*CPT1A*	0.69	[0.61-0.78]	**0.80**	**[0.74-0.88]**	**1.1E-06**	**-6.7E-03**	**7.9E-04**	**4.8E-17**
cg06235429	11	*NDUFV1*	1.49	[1.3-1.7]	1.11	[1-1.24]	5.8E-02	2.4E-03	1.3E-03	6.5E-02
cg05778424	17	*AKAP1*	1.69	[1.42-2.02]	**1.44**	**[1.21-1.71]**	**3.5E-05**	**4.9E-03**	**1.6E-03**	**2.5E-03**
cg11376147	11	*SLC43A1*	0.68	[0.59-0.77]	**0.85**	**[0.74-0.97]**	**1.5E-02**	**-3.2E-03**	**1.2E-03**	**8.4E-03**
cg04816311	7	*C7orf50*	1.51	[1.31-1.75]	**1.13**	**[1-1.27]**	**4.4E-02**	**2.0E-02**	**3.2E-03**	**8.4E-10**
cg02711608	19	*SLC1A5*	0.69	[0.6-0.79]	**0.84**	**[0.76-0.93]**	**9.7E-04**	**-7.9E-03**	**1.7E-03**	**2.0E-06**
cg08309687	21	-	0.68	[0.6-0.78]	**0.82**	**[0.74-0.91]**	**1.9E-04**	**-7.8E-03**	**3.0E-03**	**1.0E-02**
cg13514042	7	-	1.42	[1.25-1.61]	1.04	[0.94-1.15]	4.4E-01	1.8E-04	1.4E-03	9.0E-01
cg08994060	10	*PFKFB3*	0.65	[0.55-0.76]	**0.81**	**[0.72-0.92]**	**6.6E-04**	**-1.6E-02**	**2.5E-03**	**8.5E-10**
cg01676795	7	*POR*	1.56	[1.33-1.84]	1.09	[0.95-1.26]	2.2E-01	**9.2E-03**	**2.4E-03**	**1.2E-04**
cg25130381	1	*SLC9A1*	1.49	[1.29-1.73]	**1.23**	**[1.09-1.39]**	**1.2E-03**	**6.5E-03**	**1.7E-03**	**1.7E-04**
cg11183227	15	*MAN2A2*	1.49	[1.29-1.72]	1.08	[0.97-1.2]	1.9E-01	**4.6E-03**	**2.0E-03**	**2.2E-02**

MVPs and individual cells with confirmed association $P<0.05$ are highlighted in bold. FHS: T2DM (403 cases, 2,204 controls). LOLIPOP: (1,074 cases, 1,590 controls). OR: odds ratio for T2DM per +1 standard deviation in methylation intensity.
*In FHS, beta indicates difference in percentage DNA methylation intensity between cases and controls, adjusted for age, sex, PC1-3 (calculated from methylation data), batch and family structure.

Table 4: Analysis of the top 18 T2DM-associated MVPs in non-blood tissues

BLOOD (T2DM EPIC-Norfolk)					LIVER			FAT					MUSCLE				
		T2D-CO	T2D-CON		T2D-CON	T2D-CON	T2D-CON	T2D	CON	T2D-CON	T2D-CON	T2D-CON	T2D	CON	T2D-CON	T2D-CON	T2D-CON
CpG ID	**Chr**	**OR**	***P-value***	**Gene**	**lm_coeff**	***P-value***	**Consistent**	**(mean beta)**	**(mean beta)**	**(mean beta)**	***P-value***	**Consistent**	**(mean beta)**	**(mean beta)**	**(mean beta)**	***P-value***	**Consistent**
cg19693031	1	0.52	*2.7E-21*	TXNIP	-0.041	*0.65*	**TRUE**	0.501	0.499	0.002	*0.71*	FALSE	**0.594**	**0.634**	**-0.040**	***0.02***	**TRUE**
cg06500161	21	1.65	*6.4E-14*	ABCG1	0.035	*0.64*	**TRUE**	**0.477**	**0.439**	**0.038**	***0.02***	**TRUE**	0.360	0.349	0.011	*0.75*	**TRUE**
cg14476101	1	0.67	*2.8E-10*	PHGDH	0.080	*0.14*	FALSE	0.565	0.561	0.004	*0.86*	FALSE	0.483	0.484	-0.001	*0.85*	**TRUE**
cg14020176	17	1.63	*3.3E-10*	SLC9A3R1	0.065	*0.26*	**TRUE**	0.678	0.686	-0.008	*0.36*	FALSE	0.741	0.748	-0.007	*0.52*	FALSE
cg11024682	17	1.56	*6.0E-10*	SREBF1	0.081	*0.24*	**TRUE**	0.646	0.640	0.007	*0.58*	**TRUE**	0.406	0.391	0.016	*0.96*	**TRUE**
cg06397161	22	1.51	*4.5E-09*	SYNGR1	-0.145	*0.07*	FALSE	0.501	0.516	-0.015	*0.09*	FALSE	0.562	0.566	-0.005	*0.78*	FALSE
cg00574958	11	0.69	*5.2E-09*	CPT1A	0.059	*0.42*	FALSE	0.091	0.096	-0.005	*0.22*	**TRUE**	0.097	0.103	-0.006	*0.33*	**TRUE**
cg06235429	11	1.49	*5.5E-09*	NDUFV1	-0.037	*0.60*	FALSE	0.793	0.793	0.001	*0.43*	**TRUE**	0.764	0.765	-0.001	*0.82*	FALSE
cg05778424	17	1.69	*7.4E-09*	AKAP1	0.019	*0.75*	**TRUE**	0.583	0.594	-0.011	*0.36*	FALSE	0.644	0.648	-0.005	*0.64*	FALSE
cg11376147	11	0.68	*1.3E-08*	SLC43A1	0.076	*0.15*	FALSE	0.301	0.305	-0.005	*0.54*	**TRUE**	0.282	0.285	-0.003	*0.64*	**TRUE**
cg04816311	7	1.51	*1.7E-08*	C7orf50	0.026	*0.63*	**TRUE**	0.853	0.861	-0.008	*0.81*	FALSE	0.879	0.885	-0.006	*0.35*	FALSE
cg02711608	19	0.69	*4.5E-08*	SLC1A5	0.018	*0.72*	FALSE	0.245	0.256	-0.011	*0.06*	**TRUE**	0.356	0.376	-0.020	*0.06*	**TRUE**
cg08309687	21	0.68	*4.5E-08*		0.017	*0.85*	FALSE	0.634	0.653	-0.019	*0.12*	**TRUE**	0.446	0.445	0.001	*0.85*	FALSE
cg13514042	7	1.42	*4.5E-08*		0.130	*0.09*	**TRUE**	0.710	0.706	0.003	*0.71*	**TRUE**	0.732	0.724	0.007	*0.55*	**TRUE**
cg08994060	10	0.65	*5.2E-08*	PFKFB3	-0.095	*0.17*	**TRUE**	0.170	0.175	-0.005	*0.86*	**TRUE**	0.154	0.143	0.011	*0.55*	FALSE
cg01676795	7	1.56	*6.5E-08*	POR	-0.090	*0.47*	FALSE	0.851	0.852	-0.001	*0.86*	FALSE	0.890	0.892	-0.002	*0.85*	FALSE
cg25130381	1	1.49	*6.7E-08*	SLC9A1	-0.069	*0.14*	FALSE	0.621	0.610	0.011	*0.19*	**TRUE**	0.764	0.776	-0.011	*0.40*	FALSE
cg11183227	15	1.49	*7.0E-08*	MAN2A2	-0.105	*0.19*	FALSE	0.945	0.946	-0.001	*0.63*	FALSE	0.906	0.910	-0.004	*0.40*	FALSE

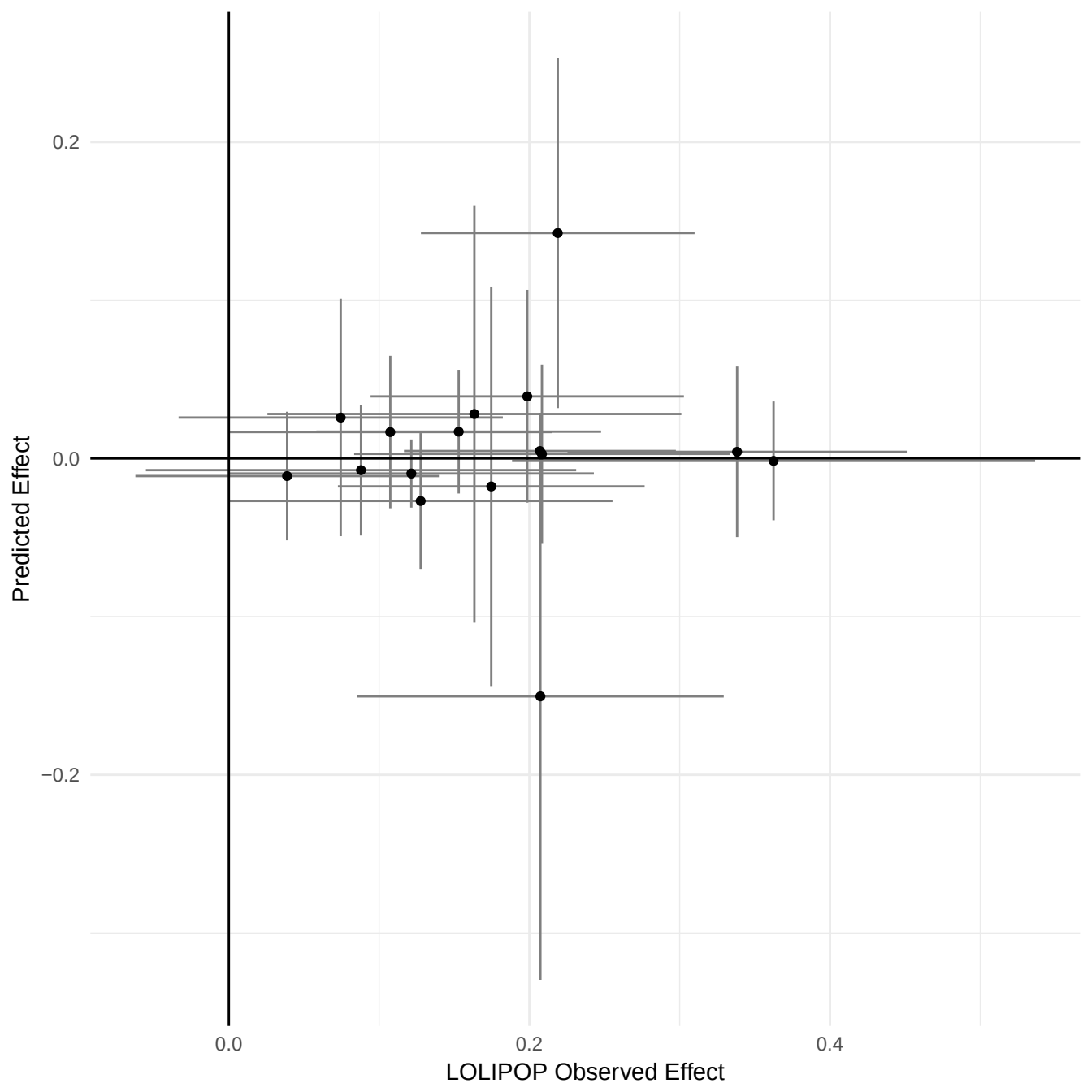
T2D: Type 2 diabetes cases

CON: Controls

Figure Legend

Figure 1: Predicted causal effects of DNA methylation on type 2 diabetes

The scatterplot shows the genetic-predicted effects of DNA methylation intensity on risk for Type 2 diabetes (Y-axis) plotted against observed effect estimates (from the LOLIPOP confirmation phase; X-axis) at each of 16 top-hit methylation variable positions (see **Table S7**). Effect sizes are log-odds ratios per 1 unit change in normalised methylation intensity aligned to higher observed odds of Type 2 diabetes.


0.2
0.0
−0.2
Predicted Effect
0.0
0.2
0.4
LOLIPOP Observed Effect

Supplementary Tables 1-11

Epigenome-wide association study of incident Type 2 diabetes in a British population: EPIC-Norfolk study

Table S1. Summary of DNA methylation array data generation and processing across the cohorts

Cohort	EPIC-Norfolk	LOLIPOP	FHS
Sample size	1264	1,074 (incident T2DM)/ 1,590 (controls)	2741
DNA source	whole blood	whole blood	whole blood
Methylation array	Illumina HM450k	Illumina HM450k	Illumina HM450k
normalisation	quantile normalisation	quantile normalisation	DASEN
QC method	Ref. 17	Ref. 17	Ref. 19
Statistical model	T2DM status ~ methylation + age + sex + estimated white blood cell counts + sample plate	T2DM status ~ methylation + control probes + bisulfite conversion batch + measured white blood cells + imputed white blood cell subsets + 5 PCA (corrected for genomic control inflation factor).	Methylation ~ T2DM status + age + sex + white blood cell counts + technical covariates + batch effects +BMI

Table S2: Confirmation in EPIC-Norfolk of the 5 MVPs with consistent associations with incident T2DM reported by Chambers *et al.* (2015)

*These additional two MVPs were reportedly not significantly associated (P>0.05) with T2DM in their European replication samples.
MVPs with confirmed association *P*<7.1E-03 (=0.05/7) are highlighted.

				Chambers et al. 2015				EPIC-Norfolk Study		
CpG	**Chr**	**Position**	**Gene(s)**	**RR (Indian Asians)**	**P (Indian Asians)**	**RR (Europeans)**	**P (Europeans)**	**OR**	**CI 95%**	**P**
cg19693031	1	145441552	*TXNIP*	0·92 (0·91–0·94)	1.0E-13	0·96 (0·94–0·98)	2.50E-05	0.524	[0.46-0.6]	2.7E-21
cg06500161	21	43656587	*ABCG1*	1·08 (1·06–1·10)	2.2E-13	1·04 (1·02–1·06)	1.2E-04	1.655	[1.45-1.89]	6.4E-14
cg11024682	17	17730094	*SREBF1*	1·06 (1·04–1·08)	8.4E-09	1·03 (0·95–0·99)	5.4E-03	1.556	[1.35-1.79]	6.0E-10
cg02650017	17	47301614	*PHOSPHO1*	0·94 (0·92–0·96)	2.1E-09	0·97 (0·95–0·99)	1.2E-03	0.863	[0.76-0.97]	1.8E-02
cg18181703	17	76354621	*SOCS3*	0·95 (0·93–0·97)	2.1E-07	0·97 (0·95–0·99)	1.6E-03	0.909	[0.81-1.02]	1.2E-01
Reported inconsistent associations										
* cg04999691	7	150027050	*C7orf29*	0·95 (0·93–0·96)	1.4E-08	1·00 (0·98–1·02)	7.1E-01	0.999	[0.88-1.14]	9.9E-01
* cg09152259	2	128156114	*PROC*	0·95 (0·93–0·97)	9.3E-08	0·99 (0·97–1·01)	3.2E-01	0.799	[0.71-0.91]	4.2E-04

Table S3: Conditional models for the top 18 MVPs associated with incident T2DM - adjusted for baseline BMI (B), HbA1c (C), BMI & HbA1c (D), in the EPIC-Norfolk study

Model A: adjusted for age, sex, estimated cell counts and technical covariates (Discovery).
Model B: model A + adjustment for baseline BMI.
Model C: model A + adjustment for baseline HbA1c.
Model D: model A + adjustement for baseline HbA1c and BMI.
OR: Odds Ratio per +1 SD in normalised methylation intensity; CI 95%: 95% Confidence Interval; Attentn: % Attenuation in the odds ratio from Model A.

				Model A			Model B (n=1260)				Model C (n=1262)				Model D (n=1256)			
CpG ID	**Chr**	**Position**	**Gene**	**OR**	**CI 95%**	**P-value**	**OR**	**CI 95%**	**P-value**	**Attentn.**	**OR**	**CI 95%**	**P-value**	**Attentn.**	**OR**	**CI 95%**	**P-value**	**Attentn.**
cg19693031	1	144152909	*TXNIP*	0.52	[0.46-0.6]	2.7E-21	0.55	[0.47-0.63]	2.5E-16	5%	0.68	[0.57-0.81]	8.5E-06	33%	0.68	[0.57-0.81]	1.9E-05	33%
cg06500161	21	42529656	*ABCG1*	1.65	[1.45-1.89]	6.4E-14	1.47	[1.27-1.69]	1.2E-07	29%	1.48	[1.25-1.75]	4.5E-06	26%	1.36	[1.13-1.62]	8.0E-04	46%
cg14476101	1	120057515	*PHGDH*	0.67	[0.59-0.76]	2.8E-10	0.78	[0.68-0.9]	4.4E-04	35%	0.68	[0.58-0.8]	4.1E-06	5%	0.76	[0.64-0.91]	1.9E-03	29%
cg14020176	17	70276580	*SLC9A3R1*	1.63	[1.4-1.9]	3.3E-10	1.48	[1.26-1.75]	3.6E-06	24%	1.36	[1.11-1.65]	2.4E-03	44%	1.26	[1.02-1.55]	3.1E-02	59%
cg11024682	17	17670819	*SREBF1*	1.56	[1.35-1.79]	6.0E-10	1.43	[1.23-1.67]	3.4E-06	22%	1.31	[1.08-1.57]	4.7E-03	45%	1.25	[1.03-1.52]	2.4E-02	55%
cg06397161	22	38090005	*SYNGR1*	1.51	[1.32-1.73]	4.5E-09	1.44	[1.23-1.68]	3.3E-06	14%	1.30	[1.09-1.56]	4.4E-03	41%	1.26	[1.04-1.53]	1.6E-02	49%
cg00574958	11	68364198	*CPT1A*	0.69	[0.61-0.78]	5.2E-09	0.79	[0.69-0.9]	5.5E-04	31%	0.73	[0.63-0.86]	1.0E-04	14%	0.81	[0.69-0.96]	1.3E-02	39%
cg06235429	11	67129690	*NDUFV1*	1.49	[1.3-1.7]	5.5E-09	1.38	[1.2-1.6]	1.2E-05	22%	1.32	[1.11-1.57]	1.9E-03	35%	1.25	[1.04-1.5]	1.6E-02	48%
cg05778424	17	52524507	*AKAP1*	1.69	[1.42-2.02]	7.4E-09	1.57	[1.3-1.91]	5.1E-06	17%	1.41	[1.12-1.78]	3.6E-03	41%	1.35	[1.06-1.72]	1.6E-02	50%
cg11376147	11	57017774	*SLC43A1*	0.68	[0.59-0.77]	1.3E-08	0.72	[0.62-0.84]	1.4E-05	14%	0.66	[0.55-0.79]	5.2E-06	-4%	0.68	[0.56-0.82]	6.8E-05	2%
cg04816311	7	1033176	*C7orf50*	1.51	[1.31-1.75]	1.7E-08	1.47	[1.26-1.72]	1.4E-06	8%	1.18	[0.98-1.43]	7.4E-02	64%	1.19	[0.98-1.44]	8.2E-02	63%
cg02711608	19	51979804	*SLC1A5*	0.69	[0.6-0.79]	4.5E-08	0.77	[0.67-0.89]	5.5E-04	27%	0.79	[0.66-0.93]	5.0E-03	31%	0.84	[0.7-1.01]	6.3E-02	50%
cg08309687	21	34242466	-	0.68	[0.6-0.78]	4.5E-08	0.75	[0.64-0.87]	1.4E-04	20%	0.79	[0.67-0.95]	1.0E-02	35%	0.86	[0.71-1.03]	1.1E-01	56%
cg13514042	7	1158728	-	1.42	[1.25-1.61]	4.5E-08	1.33	[1.16-1.53]	4.0E-05	20%	1.36	[1.16-1.59]	1.9E-04	15%	1.30	[1.1-1.54]	2.3E-03	28%
cg08994060	10	6254032	*PFKFB3*	0.65	[0.55-0.76]	5.2E-08	0.67	[0.56-0.79]	4.5E-06	6%	0.73	[0.59-0.89]	2.1E-03	23%	0.74	[0.59-0.91]	5.1E-03	25%
cg01676795	7	75424284	*POR*	1.56	[1.33-1.84]	6.5E-08	1.48	[1.24-1.77]	1.3E-05	14%	1.34	[1.08-1.67]	7.6E-03	39%	1.28	[1.02-1.61]	3.3E-02	50%
cg25130381	1	27313308	*SLC9A1*	1.49	[1.29-1.73]	6.7E-08	1.42	[1.22-1.67]	1.0E-05	14%	1.23	[1.02-1.48]	3.2E-02	54%	1.21	[1-1.48]	5.3E-02	57%
cg11183227	15	89256411	*MAN2A2*	1.49	[1.29-1.72]	7.0E-08	1.42	[1.21-1.66]	1.7E-05	14%	1.18	[0.99-1.42]	6.8E-02	62%	1.16	[0.96-1.41]	1.3E-01	67%

Table S4: Significant gene-set enrichment analysis results for the EWAS result with MAGENTA.

Significance is based on the FDR 75% cut-off. For further details see the original paper (Segrè *et al.* 2010).

Database	Gene Set	Nominal GSEA P-val 75% cut-off	FDR 75% cut-off	Exp. Genes above 75% cutoff	Obs. Genes above 75% cutoff
Ingenuity	JAK.Stat.Signaling	3.00E-04	3.90E-03	3	8
Ingenuity	Erythropoietin.Signaling	4.00E-04	4.50E-03	4	10
Ingenuity	Insulin.Receptor.Signaling	1.17E-04	5.60E-03	9	19
Ingenuity	Fc.Epsilon.RI.Signaling	3.00E-04	6.20E-03	4	10
Ingenuity	PI3K.AKT.Signaling	1.20E-03	1.03E-02	7	15
Ingenuity	T.Cell.Receptor.Signaling	2.10E-03	1.26E-02	8	16
Ingenuity	Integrin.Signaling	1.00E-03	1.29E-02	9	18
Ingenuity	B.Cell.Receptor.Signaling	2.20E-03	1.32E-02	8	16
Ingenuity	Neuregulin.Signaling	2.10E-03	1.56E-02	6	13
Ingenuity	IGF-1.Signaling	6.10E-03	2.10E-02	5	10

Table S5: Evaluation of the top 18 T2DM-associated MVPs with glycemic traits in the Framingham Heart Studies and with BMI in Wahl *et al.*, 2017

Direction of methylation: Associations are aligned to T2DM-increasing methylation changes. MVPs with associations P<0.05 are highlighted in bold.
*Adjusted for age, sex, PC1-3 (calculated from methylation data), batch and family structure.
** Additionally adjusted for BMI.

				Fasting glucose (n=2117)*			Fasting insulin (n=2151)**			HbA1c (n=2156)*			BMI Wahl (n=10,261)		
CpG ID	**Chr**	**Gene**	**Direction of methylation**	**beta**	**se**	**p**	**beta**	**se**	**p**	**beta**	**se**	**p**	**beta**	**se**	**p**
cg19693031	1	*TXNIP*	Decreasing	**7.4E-04**	**1.2E-04**	**4.0E-10**	**3.4E-03**	**1.5E-03**	**2.1E-02**	**1.1E-02**	**3.9E-03**	**6.0E-03**	**-9.8**	**2.0**	**1.5E-06**
cg06500161	21	*ABCG1*	Increasing	**5.8E-04**	**7.9E-05**	**4.6E-13**	**8.3E-03**	**9.5E-04**	**3.9E-18**	**1.4E-02**	**2.6E-03**	**2.0E-07**	**34.8**	**2.3**	**1.7E-53**
cg14476101	1	*PHGDH*	Decreasing	**8.5E-04**	**1.6E-04**	**9.5E-08**	**5.5E-03**	**1.9E-03**	**4.4E-03**	-1.4E-03	5.3E-03	7.9E-01	**-14.5**	**1.6**	**4.1E-20**
cg14020176	17	*SLC9A3R1*	Increasing	6.9E-05	6.7E-05	3.1E-01	**2.2E-03**	**8.4E-04**	**7.9E-03**	**7.0E-03**	**2.2E-03**	**1.9E-03**	**17.5**	**3.0**	**3.3E-09**
cg11024682	17	*SREBF1*	Increasing	**3.5E-04**	**6.9E-05**	**6.9E-07**	**5.4E-03**	**8.4E-04**	**1.5E-10**	**9.6E-03**	**2.3E-03**	**3.0E-05**	**32.5**	**3.0**	**3.6E-27**
cg06397161	22	*SYNGR1*	Increasing	8.6E-05	9.7E-05	3.8E-01	**3.7E-03**	**1.2E-03**	**2.2E-03**	6.1E-03	3.2E-03	6.1E-02	**13.6**	**2.2**	**5.7E-10**
cg00574958	11	*CPT1A*	Decreasing	**2.6E-04**	**3.6E-05**	**4.4E-13**	**3.2E-03**	**4.3E-04**	**4.4E-13**	**3.6E-03**	**1.2E-03**	**2.4E-03**	**-40.2**	**3.1**	**1.4E-38**
cg06235429	11	*NDUFV1*	Increasing	7.5E-05	5.8E-05	1.9E-01	3.2E-04	7.1E-04	6.5E-01	1.5E-03	1.9E-03	4.3E-01	**15.5**	**3.9**	**8.3E-05**
cg05778424	17	*AKAP1*	Increasing	**2.1E-04**	**7.1E-05**	**3.5E-03**	**3.3E-03**	**8.8E-04**	**2.1E-04**	**7.5E-03**	**2.4E-03**	**1.6E-03**	**21.6**	**2.8**	**1.2E-14**
cg11376147	11	*SLC43A1*	Decreasing	**1.5E-04**	**5.5E-05**	**5.8E-03**	1.1E-03	6.8E-04	1.2E-01	2.5E-03	1.8E-03	1.7E-01	**-25.3**	**3.0**	**2.9E-17**
cg04816311	7	*C7orf50*	Increasing	**4.8E-04**	**1.4E-04**	**6.7E-04**	**6.2E-03**	**1.7E-03**	**3.6E-04**	**1.8E-02**	**4.7E-03**	**1.7E-04**	**9.3**	**1.8**	**1.9E-07**
cg02711608	19	*SLC1A5*	Decreasing	**1.7E-04**	**7.3E-05**	**2.2E-02**	**2.8E-03**	**9.1E-04**	**2.4E-03**	-8.2E-04	2.4E-03	7.3E-01	**-17.2**	**2.4**	**1.3E-12**
cg08309687	21	-	Decreasing	**6.6E-04**	**1.3E-04**	**7.9E-07**	**6.9E-03**	**1.7E-03**	**3.0E-05**	**1.7E-02**	**4.4E-03**	**9.6E-05**	**-15.2**	**1.9**	**9.4E-16**
cg13514042	7	-	Increasing	1.2E-04	6.3E-05	5.9E-02	8.8E-04	7.7E-04	2.6E-01	2.8E-03	2.1E-03	1.8E-01	**14.7**	**4.4**	**9.1E-04**
cg08994060	10	*PFKFB3*	Decreasing	**3.4E-04**	**1.1E-04**	**2.2E-03**	**4.2E-03**	**1.4E-03**	**3.0E-03**	**9.7E-03**	**3.8E-03**	**9.7E-03**	**-7.2**	**1.7**	**1.8E-05**
cg01676795	7	*POR*	Increasing	1.6E-04	1.0E-04	1.2E-01	**3.1E-03**	**1.3E-03**	**1.5E-02**	**1.3E-02**	**3.5E-03**	**1.3E-04**	**10.7**	**2.4**	**7.9E-06**
cg25130381	1	*SLC9A1*	Increasing	**1.6E-04**	**7.6E-05**	**3.2E-02**	**3.5E-03**	**9.4E-04**	**2.0E-04**	**6.1E-03**	**2.6E-03**	**1.7E-02**	**14.1**	**2.5**	**3.1E-08**
cg11183227	15	*MAN2A2*	Increasing	**2.6E-04**	**9.0E-05**	**4.4E-03**	**3.1E-03**	**1.1E-03**	**4.9E-03**	5.5E-03	3.0E-03	6.4E-02	**18.9**	**2.8**	**3.0E-11**

Table S6: Evaluation of the top 18 T2DM-associated MVPs with cross-sectional differences in DNA methylation intensity between monozygous twin pairs discordant for Type 1 diabetes (T1DM).

*Odds ratio for incident T2DM in the discovery EPIC Norfolk study
FDR: false-discovery rate, based on paired T-tests. Delta Beta: Controls minus T1DM
No T1DM data were available for cg06397161 (SYNGR1) or cg02711608 (SLC1A5)

			Monocyte				B cell				T cell			
CpG ID	**Gene**	**OR for T2DM***	**Beta Controls**	**Beta T1DM**	**Delta Beta**	**FDR**	**Beta Controls**	**Beta T1DM**	**Delta Beta**	**FDR**	**Beta Controls**	**Beta T1DM**	**Delta Beta**	**FDR**
cg19693031	***TXNIP***	**0.52**	**0.680**	**0.744**	**-0.064**	**2.1E-05**	**0.886**	**0.919**	**-0.033**	**4.3E-04**	0.806	0.803	0.003	7.9E-01
cg06500161	*ABCG1*	1.65	0.741	0.740	0.001	8.7E-01	0.491	0.481	0.010	3.2E-01	0.705	0.713	-0.009	3.5E-01
cg14476101	*PHGDH*	0.67	0.658	0.649	0.009	5.3E-01	0.806	0.804	0.002	8.7E-01	0.883	0.883	0.001	9.0E-01
cg14020176	***SLC9A3R1***	**1.63**	**0.639**	**0.624**	**0.015**	**4.3E-04**	0.862	0.854	0.009	8.6E-01	0.905	0.903	0.002	7.9E-01
cg11024682	***SREBF1***	**1.56**	**0.414**	**0.405**	**0.010**	**2.9E-02**	0.504	0.500	0.004	8.6E-01	0.525	0.516	0.009	3.5E-01
cg00574958	***CPT1A***	**0.69**	0.032	0.033	-0.001	8.5E-01	0.034	0.032	0.001	8.6E-01	**0.136**	**0.157**	**-0.021**	**8.3E-03**
cg06235429	*NDUFV1*	1.49	0.891	0.889	0.002	8.2E-01	0.936	0.934	0.002	8.6E-01	0.942	0.946	-0.004	3.5E-01
cg05778424	*AKAP1*	1.69	0.087	0.084	0.004	8.2E-01	0.764	0.769	-0.005	8.6E-01	0.929	0.926	0.003	7.9E-01
cg11376147	*SLC43A1*	0.68	0.304	0.301	0.003	8.2E-01	0.246	0.245	0.001	9.3E-01	0.164	0.165	0.000	9.0E-01
cg04816311	***C7orf50***	**1.51**	**0.360**	**0.333**	**0.026**	**2.6E-02**	0.328	0.333	-0.005	8.6E-01	0.940	0.941	-0.001	7.9E-01
cg08309687	***NA***	**0.68**	**0.657**	**0.689**	**-0.032**	**4.0E-03**	0.634	0.657	-0.022	1.1E-01	0.069	0.067	0.002	7.9E-01
cg13514042	*NA*	1.42	0.814	0.813	0.001	8.7E-01	0.854	0.850	0.004	8.6E-01	0.912	0.910	0.002	7.9E-01
cg08994060	***PFKFB3***	**0.65**	**0.090**	**0.105**	**-0.015**	**2.0E-02**	**0.334**	**0.377**	**-0.043**	**1.2E-05**	**0.576**	**0.594**	**-0.018**	**4.7E-02**
cg01676795	*POR*	1.56	0.901	0.900	0.001	8.7E-01	0.960	0.963	-0.002	8.6E-01	0.982	0.980	0.002	3.5E-01
cg25130381	*SLC9A1*	1.49	0.676	0.665	0.011	3.0E-01	0.819	0.817	0.002	8.7E-01	0.436	0.428	0.008	5.1E-01
cg11183227	*MAN2A2*	1.49	0.981	0.981	-0.001	8.2E-01	0.976	0.975	0.001	8.6E-01	0.983	0.982	0.002	3.5E-01

Table S7: Spearman correlations in DNA methylation intensity between whole blood and liver, pancreas, adipose tissue or muscle at the top 18 T2DM-associated MVPs.

Liver and Pancreas samples are from 6 cadaveric samples; Spearman Rho values >=0.5 are highlighted in bold
Fat and muscle samples are from 14 and 16 twin pairs, respectively; correlations at P<0.05 are highlighted in bold

BLOOD (T2DM EPIC-Norfolk)			**LIVER**		**PANCREAS**		**FAT**		**MUSCLE**	
CpG ID	**Chr**	**Gene**	**Rho**	***P-value***	**Rho**	***P-value***	**Rho**	***P-value***	**Rho**	***P-value***
cg19693031	1	***TXNIP***	**0.7**	0.23	**0.8**	0.33	-0.2	*0.24*	0.2	*0.25*
cg06500161	21	*ABCG1*	-0.7	0.23	0.2	0.92	0.1	*0.68*	0.0	*0.87*
cg14476101	1	***PHGDH***	**0.7**	0.23	0.2	0.92	0.2	*0.23*	-0.1	*0.78*
cg14020176	17	***SLC9A3R1***	**0.7**	0.23	**0.8**	0.33	**0.4**	***0.05***	0.0	*0.81*
cg11024682	17	***SREBF1***	0.4	0.52	0.0	1.00	**0.5**	***0.01***	0.0	*0.79*
cg06397161	22	***SYNGR1***	0.3	0.68	0.0	1.00	**0.5**	***0.01***	0.2	*0.18*
cg00574958	11	***CPT1A***	**0.5**	0.45	-0.4	0.75	**0.5**	***0.01***	-0.2	*0.21*
cg06235429	11	*NDUFV1*	-0.8	0.13	0.4	0.75	0.3	*0.08*	0.0	*0.85*
cg05778424	17	*AKAP1*	-0.5	0.45	-0.8	0.33	0.2	*0.20*	0.0	*0.95*
cg11376147	11	***SLC43A1***	**0.7**	0.23	0.4	0.75	0.2	*0.40*	-0.2	*0.20*
cg04816311	7	***C7orf50***	**0.6**	0.35	0.2	0.92	-0.1	*0.64*	0.1	*0.63*
cg02711608	19	***SLC1A5***	**1.0**	**0.02**	0.0	1.00	0.2	*0.37*	-0.1	*0.47*
cg08309687	21	***NA***	0.1	0.95	0.2	0.92	**0.7**	***0.0001***	0.0	*0.95*
cg13514042	7	*NA*	-0.5	0.45	0.2	0.92	-0.1	*0.76*	0.0	*0.95*
cg08994060	10	***PFKFB3***	0.0	1.00	-0.4	0.75	0.2	*0.24*	**0.4**	***0.03***
cg01676795	7	*POR*	-0.9	0.08	1.0	0.08	0.2	*0.43*	-0.2	*0.37*
cg25130381	1	*SLC9A1*	-0.3	0.68	0.0	1.00	0.2	*0.32*	-0.1	*0.58*
cg11183227	15	***MAN2A2***	**0.7**	0.23	**0.6**	0.42	**0.4**	***0.02***	0.1	*0.63*

Table S8: Individual methQTLs identified in the BIOS-QTL browser for 16 of the 18 T2DM-associated MVPs.
Methylation quantitative trait loci (methQTLs) are SNPs with a genome-wide significant association (FDR< 5%) with any of the 18 T2DM-associated MVPs.
* aligned to increasing T2DM.

MVP and its observed association with T2DM						SNP to Methylation							
MVP	**chr**	**position**	**Gene**	**OR*$_{MVP-T2DM}$**	**CI 95%**	**SNP**	**chr**	**position**	**Type**	**A1**	**A2**	**Z-score $_{A1-to-MVP}$**	**P-value**
cg05778424	17	52524507	*AKAP1*	1.69	[1.42-2.02]	rs1047891	2	211540507	trans	A	C	-8.23	1.93E-16
cg05778424	17	52524507	*AKAP1*	1.69	[1.42-2.02]	rs715	2	211543055	trans	C	T	-8	1.20E-15
cg14020176	17	70276580	*SLC9A3R1*	1.63	[1.4-1.9]	rs7529925	1	199007208	trans	C	T	6.64	3.12E-11
cg11183227	15	89256411	*MAN2A2*	1.49	[1.29-1.72]	rs9790517	4	106084778	trans	T	C	-6.48	9.44E-11
cg05778424	17	52524507	*AKAP1*	1.69	[1.42-2.02]	rs2216405	2	211616894	trans	G	A	-6.44	1.20E-10
cg05778424	17	52524507	*AKAP1*	1.69	[1.42-2.02]	rs1544196	1	224632782	trans	A	G	-6.43	1.26E-10
cg08309687	21	34242466		1.46	[1.28-1.68]	rs6763931	3	141102833	trans	A	G	5.7	1.23E-08
cg08309687	21	34242466		1.46	[1.28-1.68]	rs1991431	3	141133450	trans	A	G	5.69	1.25E-08
cg08309687	21	34242466		1.46	[1.28-1.68]	rs724016	3	141105570	trans	G	A	5.69	1.25E-08
cg08309687	21	34242466		1.46	[1.28-1.68]	rs6440003	3	141094209	trans	A	G	5.63	1.84E-08
cg00574958	11	68364198	*CPT1A*	1.45	[1.28-1.64]	rs964184	11	116648917	trans	G	C	-5.54	2.99E-08
cg08994060	10	6254032	*PFKFB3*	1.55	[1.32-1.81]	rs6763931	3	141102833	trans	A	G	5.48	4.30E-08
cg00574958	11	68364198	*CPT1A*	1.45	[1.28-1.64]	rs3741298	11	116657561	trans	C	T	-5.48	4.34E-08
cg08309687	21	34242466		1.46	[1.28-1.68]	rs9310736	3	24350811	trans	A	G	-5.48	4.36E-08
cg08994060	10	6254032	*PFKFB3*	1.55	[1.32-1.81]	rs6440003	3	141094209	trans	A	G	5.46	4.70E-08
cg08994060	10	6254032	*PFKFB3*	1.55	[1.32-1.81]	rs724016	3	141105570	trans	G	A	5.43	5.72E-08
cg02711608	19	51979804	*SLC1A5*	1.45	[1.27-1.66]	rs4948102	7	56097265	trans	C	G	-5.41	6.25E-08
cg08994060	10	6254032	*PFKFB3*	1.55	[1.32-1.81]	rs1991431	3	141133450	trans	A	G	5.29	1.23E-07
cg11024682	17	17670819	*SREBF1*	1.56	[1.35-1.79]	rs7701414	5	131585958	trans	G	A	5.21	1.94E-07
cg11024682	17	17670819	*SREBF1*	1.56	[1.35-1.79]	rs7529925	1	199007208	trans	C	T	-5.15	2.62E-07
cg11183227	15	89256411	*MAN2A2*	1.49	[1.29-1.72]	rs1544196	1	224632782	trans	A	G	-5.15	2.65E-07
cg08994060	10	6254032	*PFKFB3*	1.55	[1.32-1.81]	rs592866	10	6205922	cis	A	T	-41.2	3.27E-310
cg04816311	7	1033176	*C7orf50*	1.51	[1.31-1.75]	rs56048221	7	1092533	cis	G	A	32.6	4.80E-233
cg14476101	1	120057515	*PHGDH*	1.50	[1.32-1.7]	rs11583993	1	120255370	cis	A	G	-32.2	3.85E-228
cg01676795	7	75424284	*POR*	1.56	[1.33-1.84]	rs10954750	7	75664921	cis	C	G	-22.8	7.21E-115

cg06397161	22	38090005	*SYNGR1*	1.51	[1.32-1.73]	rs2069235	22	39747780	cis	A	G	-17.5	1.72E-68
cg08309687	21	34242466		1.46	[1.28-1.68]	rs8128167	21	35307200	cis	C	T	17.5	1.92E-68
cg13514042	7	1158728		1.42	[1.25-1.61]	rs2949204	7	1200845	cis	G	A	16.1	4.85E-58
cg06235429	11	67129690	*NDUFV1*	1.49	[1.3-1.7]	rs4244823	11	67361356	cis	A	G	-13.9	1.24E-43
cg14020176	17	70276580	*SLC9A3R1*	1.63	[1.4-1.9]	rs12601504	17	72754829	cis	C	T	13.5	2.02E-41
cg06397161	22	38090005	*SYNGR1*	1.51	[1.32-1.73]	rs1569499	22	39769818	cis	T	C	-13.4	3.50E-41
cg04816311	7	1033176	*C7orf50*	1.51	[1.31-1.75]	rs12701708	7	1090333	cis	T	C	-12.4	4.25E-35
cg05778424	17	52524507	*AKAP1*	1.69	[1.42-2.02]	rs2160115	17	55165612	cis	C	T	-12.1	7.15E-34
cg08309687	21	34242466		1.46	[1.28-1.68]	rs2051387	21	35364046	cis	A	T	-10.9	1.22E-27
cg13514042	7	1158728		1.42	[1.25-1.61]	rs13226093	7	1168754	cis	C	T	-10.1	8.45E-24
cg11024682	17	17670819	*SREBF1*	1.56	[1.35-1.79]	rs8070432	17	17480474	cis	C	T	8.92	4.83E-19
cg05778424	17	52524507	*AKAP1*	1.69	[1.42-2.02]	rs148596028	17	55210022	cis	T	C	-8.45	2.87E-17
cg14476101	1	120057515	*PHGDH*	1.50	[1.32-1.7]	rs41276626	1	120262112	cis	A	G	7.59	3.14E-14
cg01676795	7	75424284	*POR*	1.56	[1.33-1.84]	rs239950	7	75581714	cis	G	A	-6.85	7.24E-12
cg08309687	21	34242466		1.46	[1.28-1.68]	rs75586969	21	35265459	cis	T	C	-6.64	3.03E-11
cg04816311	7	1033176	*C7orf50*	1.51	[1.31-1.75]	rs7784559	7	1002973	cis	A	G	-6	1.99E-09
cg05778424	17	52524507	*AKAP1*	1.69	[1.42-2.02]	rs3094438	17	55188373	cis	C	G	5.92	3.13E-09
cg14476101	1	120057515	*PHGDH*	1.50	[1.32-1.7]	rs34291690	1	120099137	cis	A	G	5.34	9.22E-08
cg04816311	7	1033176	*C7orf50*	1.51	[1.31-1.75]	rs13226093	7	1168754	cis	C	T	-5.24	1.64E-07
cg11376147	11	57017774	*SLC43A1*	1.48	[1.29-1.69]	rs2511984	11	57278733	cis	T	C	-5.22	1.76E-07
cg01676795	7	75424284	*POR*	1.56	[1.33-1.84]	rs10954672	7	75481823	cis	A	G	-5.07	3.94E-07
cg01676795	7	75424284	*POR*	1.56	[1.33-1.84]	rs59882870	7	75638421	cis	A	G	-4.74	2.11E-06
cg01676795	7	75424284	*POR*	1.56	[1.33-1.84]	rs7777399	7	75618797	cis	T	C	-4.28	1.88E-05
cg05778424	17	52524507	*AKAP1*	1.69	[1.42-2.02]	rs72843415	17	55163250	cis	T	C	-4.14	3.51E-05
cg06235429	11	67129690	*NDUFV1*	1.49	[1.3-1.7]	rs12799746	11	67246630	cis	T	C	-4.1	4.06E-05
cg08994060	10	6254032	*PFKFB3*	1.55	[1.32-1.81]	rs584797	10	6220496	cis	A	G	4.06	4.98E-05
cg13514042	7	1158728		1.42	[1.25-1.61]	rs55633977	7	1122645	cis	T	C	-4.05	5.14E-05
cg01676795	7	75424284	POR	1.56	[1.33-1.84]	rs73145061	7	75353969	cis	C	G	-4.02	5.81E-05
cg25130381	1	27313308	SLC9A1	1.49	[1.29-1.73]	rs34079867	1	27407850	cis	T	C	-3.88	0.000105

Table S9: Genetically predicted effects of methylation intensity at 16 MVPs on T2DM.

No of methQTLs: number of methQTLs used in Mendelian Randomization analysis after clumping the SNPs using the R package *TwoSampleMR*. The remaining columns are generated using the *TwoSampleMR* package in R. Beta's indicate the log-odds ratios per 1 unit increase normalised methylation intensity.

				Predicted effect			Heterogeneity			Directional pleiotropy		
MVP	Gene	No. of methQTLs	MR model	b	se	P	Q	df	P-val	Egger_intercept	se	P
cg08994060	*PFKFB3*	2	Inverse variance weighted	-0.003	0.029	0.92	0	1	1			
cg11183227	*MAN2A2*	2	Inverse variance weighted	0.026	0.038	0.50	0	1	1			
cg02711608	*SLC1A5*	1	Wald ratio	0.018	0.064	0.78						
cg08309687	-	3	MR Egger	0.001	0.085	0.99	6.62	1	0.01	-0.01	0.02	0.68
cg08309687	-	3	Inverse variance weighted	-0.039	0.034	0.25	4.32	2	0.12			
cg14020176	*SLC9A3R1*	2	Inverse variance weighted	-0.027	0.022	0.22	0	1	1			
cg06397161	*SYNGR1*	1	Wald ratio	0.017	0.020	0.40						
cg00574958	*CPT1A*	1	Wald ratio	-0.142	0.056	0.01						
cg11376147	*SLC43A1*	1	Wald ratio	-0.028	0.067	0.68						
cg25130381	*SLC9A1*	1	Wald ratio	-0.150	0.091	0.10						
cg11024682	*SREBF1*	3	MR Egger	0.058	0.054	0.48	0.20	1	0.65	-0.01	0.01	0.45
cg11024682	*SREBF1*	3	Inverse variance weighted	0.004	0.028	0.88	0.77	2	0.68			
cg05778424	*AKAP1*	3	MR Egger	0.022	0.108	0.87	0.04	1	0.85	-0.01	0.03	0.86
cg05778424	*AKAP1*	3	Inverse variance weighted	-0.002	0.019	0.94	0.04	2	0.98			
cg14476101	*PHGDH*	2	Inverse variance weighted	-0.005	0.010	0.65	0	1	1			
cg04816311	*C7orf50*	2	Inverse variance weighted	-0.010	0.011	0.39	0	1	1			
cg01676795	*POR*	4	MR Egger	-0.013	0.041	0.79	6.18	2	0.05	0.00	0.01	0.89
cg01676795	*POR*	4	Inverse variance weighted	-0.007	0.021	0.73	4.17	3	0.24			
cg06235429	*NDUFV1*	1	Wald ratio	0.017	0.025	0.50						
cg13514042	-	1	Wald ratio	-0.011	0.021	0.59						

Table S10: Mendelian Randomisation tests for determinants of the 18 T2DM-associated MVPs.

Results for the MR associations for the inverse-variance weighted and MR-Egger methods using 2 hour glucose (Saxena *et al*., 2010), BMI (Locke *et al.,* 2015), fasting glucose (Manning *et al.,* 2012), fasting insulin (Scott *et al.,* 2012), fasting insulin adjusted for BMI (Manning *et al*., 2012), insulin secretion (Prokopenko *et al*., 2014), insulin resistance (Lotta *et al.,* 2016), T2DM status (Morris *et al.,* 2012) and waist hip ratio (Shungin *et al.,* 2015) associated SNPs instrument variables.

Potential Determinant	MVP	IVW MR			Heterogenity		Egger MR					
		Beta	SE	p-value	Coch Q	Coch Q p-value	Beta	SE	p-value	intercept	SE intercept	p-value intercept
2 hour glucose	cg00574958	0.06	0.28	0.82	72.31	2.37E-05	-0.33	0.51	0.52	0.01	0.01	0.35
2 hour glucose	cg01676795	0.15	0.27	0.58	58.10	1.56E-03	-0.42	0.49	0.39	0.02	0.01	0.09
2 hour glucose	cg02711608	0.22	0.20	0.26	5.44	0.71	0.37	0.81	0.65	-0.01	0.06	0.85
2 hour glucose	cg04816311	-0.06	0.20	0.76	5.65	0.69	0.69	0.77	0.37	-0.06	0.06	0.31
2 hour glucose	cg05778424	0.26	0.18	0.15	4.55	0.80	-0.15	0.70	0.83	0.03	0.06	0.55
2 hour glucose	cg06235429	-0.10	0.18	0.57	3.08	0.93	-0.56	0.67	0.40	0.04	0.05	0.44
2 hour glucose	cg06397161	0.31	0.20	0.12	5.25	0.73	0.95	0.77	0.22	-0.05	0.06	0.39
2 hour glucose	cg06500161	0.02	0.20	0.90	5.39	0.72	0.68	0.78	0.39	-0.05	0.06	0.39
2 hour glucose	cg08309687	-0.35	0.19	0.06	4.45	0.81	-0.18	0.76	0.81	-0.01	0.06	0.82
2 hour glucose	cg08994060	0.14	0.18	0.42	3.48	0.90	0.69	0.67	0.30	-0.05	0.05	0.36
2 hour glucose	cg11024682	-0.17	0.18	0.34	2.88	0.94	-0.76	0.67	0.26	0.05	0.04	0.25
2 hour glucose	cg11183227	-0.04	0.18	0.80	3.02	0.93	0.48	0.67	0.48	-0.04	0.04	0.31
2 hour glucose	cg11376147	-0.20	0.28	0.48	10.85	0.21	-0.73	1.14	0.52	0.04	0.09	0.63
2 hour glucose	cg13514042	0.43	0.18	0.01	1.98	0.98	-0.03	0.67	0.96	0.04	0.04	0.30
2 hour glucose	cg14020176	0.36	0.23	0.12	8.81	0.36	0.38	0.96	0.69	0.00	0.08	0.98
2 hour glucose	cg14476101	-0.05	0.17	0.78	3.44	0.90	-0.51	0.67	0.44	0.04	0.05	0.46
2 hour glucose	cg19693031	-0.31	0.20	0.12	6.26	0.62	-1.37	0.71	0.05	0.09	0.06	0.12
2 hour glucose	cg25130381	0.17	0.18	0.33	2.45	0.96	0.69	0.68	0.31	-0.04	0.04	0.27
BMI	cg00574958	-0.14	0.17	0.44	176.24	5.1E-08	-0.80	0.42	0.06	0.02	0.01	0.08
BMI	cg01676795	-0.18	0.17	0.29	161.10	2.4E-06	-0.04	0.40	0.91	0.00	0.01	0.71

BMI	cg02711608	-0.12	0.17	0.46	161.70	2.1E-06	-0.25	0.41	0.53	0.00	0.01	0.73
BMI	cg04816311	-0.19	0.17	0.27	173.55	1.0E-07	-0.43	0.42	0.31	0.01	0.01	0.54
BMI	cg05778424	-0.19	0.18	0.29	184.66	5.3E-09	-0.29	0.43	0.50	0.00	0.01	0.79
BMI	cg06235429	-0.02	0.17	0.90	157.40	5.8E-06	-0.53	0.41	0.19	0.01	0.01	0.15
BMI	cg06397161	0.06	0.17	0.71	155.59	8.8E-06	-0.14	0.41	0.72	0.01	0.01	0.57
BMI	cg06500161	-0.15	0.18	0.42	200.01	6.9E-11	-0.05	0.44	0.91	0.00	0.01	0.81
BMI	cg08309687	0.03	0.17	0.87	180.81	1.5E-08	-0.37	0.42	0.38	0.01	0.01	0.30
BMI	cg08994060	0.12	0.19	0.52	200.93	5.3E-11	0.69	0.45	0.12	-0.02	0.01	0.16
BMI	cg11024682	-0.05	0.17	0.76	147.76	5.2E-05	0.06	0.41	0.88	0.00	0.01	0.75
BMI	cg11183227	-0.12	0.17	0.48	152.55	1.8E-05	-0.02	0.41	0.96	0.00	0.01	0.78
BMI	cg11376147	-0.03	0.17	0.86	175.42	6.4E-08	-0.39	0.42	0.34	0.01	0.01	0.34
BMI	cg13514042	-0.38	0.17	0.02	156.25	7.6E-06	-0.93	0.41	0.02	0.02	0.01	0.12
BMI	cg14020176	-0.30	0.18	0.08	178.84	2.6E-08	-0.83	0.42	0.05	0.02	0.01	0.17
BMI	cg14476101	-0.10	0.17	0.54	145.63	8.4E-05	-0.33	0.40	0.41	0.01	0.01	0.52
BMI	cg19693031	0.29	0.17	0.08	137.76	4.3E-04	0.19	0.40	0.64	0.00	0.01	0.77
BMI	cg25130381	0.13	0.17	0.44	165.30	8.5E-07	0.62	0.41	0.13	-0.01	0.01	0.19
Fasting glucose	cg00574958	0.06	0.28	0.82	72.31	2.4E-05	-0.33	0.51	0.52	0.01	0.01	0.35
Fasting glucose	cg01676795	0.15	0.27	0.58	58.10	1.6E-03	-0.42	0.49	0.39	0.02	0.01	0.09
Fasting glucose	cg02711608	0.13	0.27	0.64	59.01	1.2E-03	0.15	0.49	0.76	0.00	0.01	0.96
Fasting glucose	cg04816311	0.31	0.27	0.26	59.63	1.0E-03	0.31	0.50	0.54	0.00	0.01	0.99
Fasting glucose	cg05778424	0.19	0.34	0.57	89.54	7.7E-08	1.22	0.57	0.03	-0.03	0.02	0.03
Fasting glucose	cg06235429	-0.15	0.27	0.59	57.27	1.9E-03	-0.18	0.49	0.71	0.00	0.01	0.92
Fasting glucose	cg06397161	0.12	0.29	0.68	71.23	3.3E-05	0.26	0.54	0.62	0.00	0.01	0.75
Fasting glucose	cg06500161	0.17	0.27	0.52	60.28	8.5E-04	-0.38	0.49	0.44	0.02	0.01	0.14
Fasting glucose	cg08309687	-0.15	0.27	0.59	68.39	8.0E-05	-0.31	0.49	0.52	0.01	0.01	0.67
Fasting glucose	cg08994060	0.50	0.31	0.11	106.42	1.8E-10	0.29	0.57	0.61	0.01	0.02	0.65
Fasting glucose	cg11024682	0.05	0.27	0.86	59.01	1.2E-03	0.21	0.49	0.68	-0.01	0.01	0.69
Fasting glucose	cg11183227	-0.04	0.28	0.89	74.19	1.3E-05	-0.41	0.49	0.41	0.01	0.01	0.37
Fasting glucose	cg11376147	0.17	0.32	0.60	89.35	8.2E-08	0.17	0.58	0.77	0.00	0.02	1.00
Fasting glucose	cg13514042	0.08	0.31	0.79	78.59	3.1E-06	-0.09	0.56	0.87	0.01	0.02	0.70

Fasting glucose	cg14020176	0.07	0.27	0.81	72.23	2.4E-05	-0.15	0.50	0.77	0.01	0.01	0.61
Fasting glucose	cg14476101	0.06	0.28	0.82	72.52	2.2E-05	-0.02	0.51	0.97	0.00	0.01	0.85
Fasting glucose	cg19693031	0.09	0.27	0.73	62.25	4.9E-04	-0.93	0.49	0.06	0.03	0.01	3.9E-03
Fasting glucose	cg25130381	0.19	0.31	0.54	103.78	4.7E-10	-0.15	0.56	0.79	0.01	0.02	0.46
Fasting insulin (adj BMI)	cg00574958	0.25	0.68	0.71	32.92	4.8E-03	-3.13	2.94	0.29	0.05	0.04	0.17
Fasting insulin (adj BMI)	cg01676795	0.54	0.77	0.49	57.78	6.0E-07	8.83	2.94	0.00	-0.13	0.04	1.0E-03
Fasting insulin (adj BMI)	cg02711608	-0.31	0.79	0.69	59.60	3.0E-07	-1.68	3.52	0.63	0.02	0.05	0.69
Fasting insulin (adj BMI)	cg04816311	-0.56	0.81	0.49	69.28	6.0E-09	2.37	3.56	0.51	-0.04	0.05	0.40
Fasting insulin (adj BMI)	cg05778424	0.19	0.68	0.78	29.63	1.3E-02	1.38	2.95	0.64	-0.02	0.04	0.62
Fasting insulin (adj BMI)	cg06235429	-1.01	0.70	0.15	47.01	3.7E-05	0.06	3.13	0.98	-0.02	0.05	0.72
Fasting insulin (adj BMI)	cg06397161	0.68	0.68	0.32	44.07	1.1E-04	0.02	3.01	0.99	0.01	0.04	0.82
Fasting insulin (adj BMI)	cg06500161	-0.38	0.68	0.58	24.40	0.06	0.89	2.95	0.76	-0.02	0.03	0.56
Fasting insulin (adj BMI)	cg08309687	-0.61	0.83	0.46	69.62	5.2E-09	3.63	3.56	0.31	-0.06	0.05	0.22
Fasting insulin (adj BMI)	cg08994060	-0.54	0.68	0.42	44.25	1.0E-04	-2.13	2.94	0.47	0.02	0.04	0.58
Fasting insulin (adj BMI)	cg11024682	-0.02	0.68	0.97	12.04	0.68	-1.30	2.95	0.66	0.02	0.02	0.41
Fasting insulin (adj BMI)	cg11183227	0.06	0.72	0.93	48.31	2.3E-05	5.90	2.95	0.05	-0.09	0.04	0.03
Fasting insulin (adj BMI)	cg11376147	0.45	0.68	0.50	40.16	4.3E-04	-0.04	2.95	0.99	0.01	0.04	0.86
Fasting insulin (adj BMI)	cg13514042	0.04	0.92	0.96	78.15	1.5E-10	-9.01	3.27	0.01	0.14	0.05	4.5E-03
Fasting insulin (adj BMI)	cg14020176	-0.11	0.91	0.90	73.26	1.2E-09	2.14	4.04	0.60	-0.03	0.06	0.57
Fasting insulin (adj BMI)	cg14476101	0.18	0.68	0.80	31.53	0.01	-0.64	2.94	0.83	0.01	0.04	0.75
Fasting insulin (adj BMI)	cg19693031	-1.75	0.82	0.03	64.11	4.9E-08	-0.42	3.68	0.91	-0.02	0.05	0.71
Fasting insulin (adj BMI)	cg25130381	0.45	0.68	0.51	25.93	0.04	1.69	2.95	0.57	-0.02	0.03	0.58
Insulin secretion	cg00574958	-0.27	0.18	0.13	15.65	0.55	-0.60	0.47	0.20	0.02	0.03	0.45
Insulin secretion	cg01676795	0.10	0.16	0.52	7.56	0.98	0.00	0.42	0.99	0.01	0.02	0.70
Insulin secretion	cg02711608	-0.05	0.19	0.81	19.98	0.28	0.02	0.52	0.97	0.00	0.03	0.89
Insulin secretion	cg04816311	0.17	0.17	0.31	16.88	0.46	0.61	0.45	0.18	-0.03	0.03	0.30
Insulin secretion	cg05778424	-0.05	0.19	0.80	16.72	0.47	-0.03	0.51	0.95	0.00	0.03	0.98
Insulin secretion	cg06235429	0.22	0.16	0.17	18.89	0.33	1.14	0.42	0.01	-0.06	0.02	4.6E-04
Insulin secretion	cg06397161	-0.13	0.20	0.52	29.53	0.03	0.14	0.54	0.80	-0.02	0.03	0.59
Insulin secretion	cg06500161	0.05	0.16	0.78	14.19	0.65	-0.31	0.43	0.47	0.02	0.03	0.37

Insulin secretion	cg08309687	0.02	0.16	0.91	10.09	0.90	0.11	0.42	0.80	-0.01	0.02	0.79
Insulin secretion	cg08994060	-0.38	0.21	0.07	26.33	0.07	-0.29	0.55	0.59	-0.01	0.03	0.87
Insulin secretion	cg11024682	0.31	0.16	0.06	27.82	0.05	0.75	0.43	0.08	-0.03	0.03	0.26
Insulin secretion	cg11183227	0.14	0.16	0.38	12.52	0.77	-0.08	0.42	0.85	0.01	0.03	0.56
Insulin secretion	cg11376147	0.00	0.21	1.00	27.54	0.05	-1.02	0.50	0.04	0.07	0.03	0.03
Insulin secretion	cg13514042	0.09	0.20	0.64	21.36	0.21	-0.44	0.53	0.41	0.04	0.03	0.28
Insulin secretion	cg14020176	-0.13	0.17	0.45	20.05	0.27	0.60	0.42	0.15	-0.05	0.03	0.06
Insulin secretion	cg14476101	0.05	0.16	0.76	8.58	0.95	-0.32	0.42	0.45	0.02	0.02	0.15
Insulin secretion	cg19693031	-0.03	0.16	0.87	18.43	0.36	1.00	0.42	0.02	-0.07	0.02	9.6E-04
Insulin secretion	cg25130381	0.09	0.16	0.58	9.34	0.93	0.44	0.42	0.29	-0.02	0.02	0.27
Insulin resistance	cg00574958	-0.91	0.62	0.14	238.85	1.3E-28	0.98	1.91	0.61	-0.02	0.02	0.30
Insulin resistance	cg01676795	0.26	0.61	0.66	210.47	1.4E-23	1.37	1.89	0.47	-0.01	0.02	0.54
Insulin resistance	cg02711608	-0.58	0.69	0.40	272.93	8.6E-35	-1.14	2.16	0.60	0.01	0.02	0.78
Insulin resistance	cg04816311	0.39	0.61	0.52	214.72	2.6E-24	-0.05	1.92	0.98	0.00	0.02	0.81
Insulin resistance	cg05778424	0.52	0.59	0.38	159.10	6.1E-15	-0.15	1.83	0.93	0.01	0.02	0.66
Insulin resistance	cg06235429	-0.22	0.63	0.73	228.51	9.4E-27	-0.37	1.97	0.85	0.00	0.02	0.93
Insulin resistance	cg06397161	0.90	0.70	0.20	311.31	6.2E-42	2.57	2.18	0.24	-0.02	0.02	0.42
Insulin resistance	cg06500161	0.44	0.59	0.46	200.51	7.5E-22	1.14	1.83	0.53	-0.01	0.02	0.68
Insulin resistance	cg08309687	-0.30	0.61	0.63	210.36	1.5E-23	0.26	1.91	0.89	-0.01	0.02	0.76
Insulin resistance	cg08994060	-0.15	0.59	0.80	125.65	8.7E-10	-0.74	1.83	0.69	0.01	0.02	0.68
Insulin resistance	cg11024682	0.99	0.59	0.09	206.64	6.6E-23	1.37	1.83	0.45	0.00	0.02	0.82
Insulin resistance	cg11183227	-0.01	0.59	0.98	180.48	1.9E-18	-1.12	1.83	0.54	0.01	0.02	0.49
Insulin resistance	cg11376147	-0.28	0.59	0.64	205.12	1.2E-22	2.35	1.82	0.20	-0.03	0.02	0.13
Insulin resistance	cg13514042	0.52	0.65	0.43	236.40	3.7E-28	-1.72	2.01	0.39	0.02	0.02	0.24
Insulin resistance	cg14020176	0.15	0.71	0.83	294.11	1.0E-38	0.28	2.23	0.90	0.00	0.02	0.95
Insulin resistance	cg14476101	-0.24	0.59	0.69	178.26	4.5E-18	0.69	1.82	0.71	-0.01	0.02	0.57
Insulin resistance	cg19693031	-1.33	0.73	0.07	304.97	9.5E-41	-2.42	2.28	0.29	0.01	0.02	0.61
Insulin resistance	cg25130381	1.41	0.59	0.02	189.18	6.5E-20	1.25	1.83	0.50	0.00	0.02	0.92
Type 2 diabetes	cg00574958	-0.04	0.06	0.56	22.03	1.00	-0.07	0.15	0.66	0.00	0.01	0.83
Type 2 diabetes	cg01676795	0.10	0.06	0.11	22.89	0.99	0.01	0.15	0.94	0.01	0.02	0.50

Type 2 diabetes	cg02711608	0.07	0.07	0.35	25.51	0.98	0.01	0.17	0.95	0.01	0.02	0.70
Type 2 diabetes	cg04816311	0.03	0.06	0.59	21.60	1.00	0.01	0.15	0.95	0.00	0.01	0.84
Type 2 diabetes	cg05778424	0.08	0.06	0.17	17.59	1.00	0.05	0.14	0.71	0.00	0.01	0.77
Type 2 diabetes	cg06235429	0.01	0.06	0.83	22.21	0.99	0.05	0.14	0.73	0.00	0.01	0.77
Type 2 diabetes	cg06397161	-0.03	0.06	0.69	24.58	0.99	-0.06	0.15	0.67	0.00	0.01	0.78
Type 2 diabetes	cg06500161	0.06	0.06	0.34	20.50	1.00	0.19	0.14	0.17	-0.02	0.01	0.28
Type 2 diabetes	cg08309687	-0.08	0.06	0.18	18.68	1.00	-0.22	0.14	0.12	0.02	0.01	0.26
Type 2 diabetes	cg08994060	0.08	0.07	0.25	27.00	0.96	0.30	0.16	0.06	-0.02	0.02	0.13
Type 2 diabetes	cg11024682	0.00	0.07	0.95	27.30	0.96	0.05	0.15	0.72	-0.01	0.02	0.67
Type 2 diabetes	cg11183227	0.01	0.06	0.92	19.87	1.00	0.07	0.14	0.62	-0.01	0.01	0.60
Type 2 diabetes	cg11376147	-0.03	0.08	0.66	30.77	0.90	-0.03	0.18	0.86	0.00	0.02	0.98
Type 2 diabetes	cg13514042	-0.07	0.08	0.38	33.15	0.83	-0.01	0.18	0.95	-0.01	0.02	0.72
Type 2 diabetes	cg14020176	0.11	0.06	0.08	23.23	0.99	0.32	0.14	0.03	-0.02	0.01	0.11
Type 2 diabetes	cg14476101	-0.03	0.06	0.61	18.91	1.00	0.19	0.14	0.18	-0.02	0.01	0.06
Type 2 diabetes	cg19693031	-0.08	0.07	0.28	26.34	0.97	-0.02	0.16	0.91	-0.01	0.02	0.69
Type 2 diabetes	cg25130381	0.01	0.06	0.91	16.17	1.00	0.13	0.14	0.37	-0.01	0.01	0.27
WHR (adj. BMI)	cg00574958	-0.07	0.26	0.78	67.13	1.8E-03	-1.28	1.31	0.33	0.03	0.04	0.34
WHR (adj. BMI)	cg01676795	0.02	0.26	0.93	65.34	2.8E-03	1.14	1.31	0.38	-0.03	0.04	0.38
WHR (adj. BMI)	cg02711608	-0.01	0.28	0.98	70.11	8.2E-04	-1.28	1.39	0.35	0.04	0.04	0.35
WHR (adj. BMI)	cg04816311	-0.25	0.25	0.31	61.98	0.01	1.29	1.22	0.29	-0.04	0.03	0.20
WHR (adj. BMI)	cg05778424	0.00	0.24	0.99	49.18	0.09	-0.02	1.20	0.99	0.00	0.03	0.99
WHR (adj. BMI)	cg06235429	0.15	0.24	0.53	38.27	0.41	-1.38	1.21	0.25	0.04	0.03	0.10
WHR (adj. BMI)	cg06397161	0.09	0.24	0.70	59.73	0.01	-0.40	1.20	0.74	0.01	0.03	0.68
WHR (adj. BMI)	cg06500161	0.09	0.24	0.70	56.91	0.02	-1.06	1.20	0.38	0.03	0.03	0.31
WHR (adj. BMI)	cg08309687	-0.08	0.25	0.73	59.58	0.01	0.19	1.25	0.88	-0.01	0.03	0.82
WHR (adj. BMI)	cg08994060	0.15	0.25	0.55	61.49	0.01	-1.44	1.22	0.24	0.04	0.03	0.19
WHR (adj. BMI)	cg11024682	0.13	0.24	0.60	57.59	0.02	1.39	1.20	0.25	-0.04	0.03	0.28
WHR (adj. BMI)	cg11183227	-0.09	0.28	0.76	78.04	9.3E-05	-0.05	1.44	0.97	0.00	0.04	0.98
WHR (adj. BMI)	cg11376147	0.16	0.24	0.51	53.07	0.04	-1.70	1.20	0.16	0.05	0.03	0.09
WHR (adj. BMI)	cg13514042	-0.17	0.24	0.48	52.92	0.04	0.82	1.20	0.50	-0.03	0.03	0.40

WHR (adj. BMI)	cg14020176	-0.12	0.24	0.62	52.57	0.05	1.47	1.21	0.22	-0.04	0.03	0.16
WHR (adj. BMI)	cg14476101	0.10	0.24	0.68	37.83	0.43	0.29	1.20	0.81	-0.01	0.03	0.84
WHR (adj. BMI)	cg19693031	-0.28	0.24	0.24	57.56	0.02	-2.27	1.20	0.06	0.06	0.03	0.08
WHR (adj. BMI)	cg25130381	0.35	0.24	0.15	45.72	0.15	0.65	1.21	0.59	-0.01	0.03	0.77

Table S11: Gene annotations for the top 18 T2DM-associated MVPs.

MVP (CpG ID)	Chr	Position	P-value	Gene	Gene position	Full gene name	Gene function	Associated disorders
cg19693031	1	144152909	2.7E-21	*TXNIP*	3'UTR	thioredoxin interacting protein	The encoded protein inhibits the antioxidative function of thioredoxin resulting in the accumulation of reactive oxygen species and cellular stress. This protein also regulates cellular metabolism and endoplasmic reticulum (ER) stress, and may also function as a tumor suppressor.	
cg06500161	21	42529656	6.4E-14	*ABCG1*	Body	ATP-binding cassette, sub-family G (WHITE), member 1	The encoded protein is a member of the White subfamily of ATP-binding cassette (ABC) transporters, and is involved in macrophage cholesterol and phospholipids transport, and may regulate cellular lipid homeostasis in other cell types.	**Tangier disease**: Characterized by low levels of high-density lipoprotein cholesterol (HDL) in the blood and accordingly with moderately increased risk of cardiovascular disease. **Autosomal Recessive Nonsyndromic Deafness**
cg14476101	1	120057515	2.8E-10	*PHGDH*	Body	phosphoglycerate dehydrogenase	The encoded enzyme is involved in the early steps of L-serine synthesis in animal cells. L-serine is required for D-serine and other amino acid synthesis.	**Neu-Laxova Syndrome 1**: An autosomal recessive lethal multiple malformation syndrome. **Phosphoglycerate Dehydrogenase Deficiency**: characterized by microcephaly; impaired development of physical reactions, movements, and speech (psychomotor retardation); and recurrent seizures (epilepsy).
cg14020176	17	70276580	3.3E-10	*SLC9A3R1*	3'UTR	solute carrier family 9, subfamily A (NHE3, cation proton antiporter 3), member 3 regulator 1	Encodes a sodium/hydrogen exchanger regulatory cofactor, which regulates: the cystic fibrosis transmembrane conductance regulator, and G-protein coupled receptors such as the beta2-adrenergic receptor and the parathyroid hormone 1 receptor. Interacts with proteins that function as linkers between integral membrane and cytoskeletal proteins.	**Nephrolithiasis/osteoporosis, hypophosphatemic, 2**: Characterized by decreased renal phosphate absorption, renal phosphate wasting, hypophosphatemia, hyperphosphaturia, hypercalciuria, nephrolithiasis and osteoporosis.

cg11024682	17	17670819	6.0E-10	*SREBF1*	Body	sterol regulatory element binding transcription factor 1 [alias: sterol regulatory element binding protein 1; SREBP1	Encodes a transcription factor that binds to the sterol regulatory element-1 (SRE1), which is a decamer flanking the low density lipoprotein receptor gene and some genes involved in sterol biosynthesis. SREBP1 and SREBP2 (600481) are structurally related proteins that control cholesterol homeostasis by stimulating transcription of sterol-regulated genes.	This gene is located within the Smith-Magenis syndrome region (a developmental disorder characterised by intellectual disability, delayed speech and language skills, distinctive facial features, sleep disturbances, and behavioral problems).
cg06397161	22	38090005	4.5E-09	*SYNGR1*	Body;TSS200	synaptogyrin 1	Encodes an integral membrane protein associated with presynaptic vesicles in neuronal cells. The exact function is unclear, but may function in synaptic plasticity.	
cg00574958	11	68364198	5.2E-09	*CPT1A*	5'UTR	carnitine palmitoyltransferase 1A (liver)	Initiates mitochondrial oxidation of long-chain fatty acids. CPT I is the key enzyme in the carnitine-dependent transport across the mitochondrial inner membrane. Its deficiency results in a decreased rate of fatty acid beta-oxidation.	**CPT deficiency, hepatic, type IA**: An autosomal recessive metabolic disorder of long-chain fatty acid oxidation characterized by severe episodes of hypoketotic hypoglycemia usually occurring after fasting or illness.
cg06235429	11	67129690	5.5E-09	*NDUFV1*	TSS1500	NADH dehydrogenase (ubiquinone) flavoprotein 1, 51kDa	Encodes a subunit of mitochondrial NADH:ubiquinone oxidoreductase complex I; a large complex that liberates electrons from NADH and channels them to ubiquinone. This subunit carries the NADH-binding site as well as flavin mononucleotide (FMN)- and Fe-S-biding sites.	**Mitochondrial Complex I Deficiency**: The most common enzymatic defect of the oxidative phosphorylation disorders. It causes a wide range of clinical disorders, including myopathies, encephalomyopathies, and neurodegenerative disorders.
cg05778424	17	52524507	7.4E-09	*AKAP1*	5'UTR	A kinase (PRKA) anchor protein 1	The encoded protein binds to type I and type II regulatory subunits of protein kinase A (PKA) and anchors them to the mitochondrion. This protein is speculated to be involved in the cAMP-dependent signal transduction pathway.	
cg11376147	11	57017774	1.3E-08	*SLC43A1*	Body	solute carrier family 43, member 1	Member of the system L family of plasma membrane carrier proteins that transports large neutral amino acids	
cg04816311	7	1033176	1.7E-08	*C7orf50*	Body	chromosome 7 open reading frame 50 (194 aa)		

cg02711608	19	51979804	4.5E-08	*SLC1A5*	1stExon;5'UTR;Body	solute carrier family 1 (neutral amino acid transporter), member 5 (541 aa)	Encodes a sodium-dependent neutral amino acid transporter that can act as a receptor for RD114/type D retrovirus	
cg08309687	21	34242466	4.5E-08	-	-			
cg13514042	7	1158728	4.5E-08	-	-			
cg08994060	10	6254032	5.2E-08	*PFKFB3*	Body	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	The encoded protein catalyses the synthesis and degradation of fructose-2,6-bisphosphate, a regulatory molecule that controls glycolysis in eukaryotes. It regulates cyclin-dependent kinase 1, linking glucose metabolism to cell proliferation and survival in tumor cells.	
cg01676795	7	75424284	6.5E-08	*POR*	Body	P450 (cytochrome) oxidoreductase	Encodes an endoplasmic reticulum membrane oxidoreductase with an FAD-binding domain and a flavodoxin-like domain. The protein binds two cofactors, FAD and FMN, which allow it to donate electrons directly from NADPH to all microsomal P450 enzymes.	**Antley-Bixler syndrome with genital anomalies and disordered steroidogenesis**: a rare craniosynostosis syndrome characterized by radiohumeral synostosis. Other features include midface hypoplasia, choanal stenosis or atresia, multiple joint contractures, visceral anomalies (genitourinary system), and impaired steroidogenesis (specific to POR mutations). **Disordered steroidogenesis due to cytochrome P450 oxidoreductase**: A rare variant of congenital adrenal hyperplasia, due to apparent combined deficiency of P450C17 and P450C21 and accumulation of steroid metabolites. Affected girls are born with ambiguous genitalia, indicating intrauterine androgen excess. After birth, however, virilization does not progress and amounts of circulating androgens are low or normal. Affected boys are sometimes born undermasculinized.

cg25130381	1	27313308	6.7E-08	*SLC9A1*	Body	solute carrier family 9, subfamily A (NHE1, cation proton antiporter 1), member 1	Encodes a Na+/H+ antiporter, a plasma membrane transporter that is expressed in the kidney and intestine. Regulates pH homeostasis, cell migration and cell volume.	**Lichtenstein-Knorr syndrome**: An autosomal recessive neurologic disorder characterized by postnatal onset of severe progressive sensorineural hearing loss and progressive cerebellar ataxia. **Myocardial Stunning**: A state when some section of the myocardium (corresponding to area of a major coronary occlusion) shows a form of contractile abnormality.
cg11183227	15	89256411	7.0E-08	*MAN2A2*	Body	mannosidase, alpha, class 2A, member 2	The encoded protein catalyzes the first committed step in the biosynthesis of complex N-glycans. It controls conversion of high mannose to complex N-glycans; the final hydrolytic step in the N-glycan maturation pathway.	