Web Extra Material

Epigenome-wide association identifies DNA methylation markers in peripheral blood that predict incident Type-2 diabetes amongst Indian Asians and Europeans.

Chambers et al

Supplementary Appendix.

Population samples

The London Life Sciences Prospective Population Study (LOLIPOP)

LOLIPOP is a prospective cohort study of ~28K Indian Asian and European men and women, recruited from the lists of 58 General Practitioners in West London, United Kingdom between 2003 and 2008.[1] At enrolment all participants completed a structured assessment of cardiovascular and metabolic health, including anthropometry, and collection of blood samples for measurement of fasting glucose, insulin and lipid profile, HbA1c, and complete blood count with differential white cell count. Aliquots of whole blood were stored at -80C for extraction of genomic DNA. Epigenome-wide association was performed amongst 2,680 participants of the LOLIPOP study free from T2D (physician diagnosis or HbA1c≥6.5%), using genomic DNA from peripheral blood collected at enrolment. The LOLIPOP study is approved by the National Research Ethics Service (07/H0712/150) and all participants gave written informed consent.

Cooperative Health Research in the Region of Augsburg (KORA)

KORA (Cooperative Health Research in the Region of Augsburg) is a research platform of independent population-based health surveys and subsequent follow-up examinations of individuals of German nationality resident in the region of Augsburg in Southern Germany. Written informed consent was obtained from all participants and the studies have been approved by the ethics committee of the Bavarian Medical Association. Study design, sampling method and data collection have been described in detail elsewhere.[2] The surveys S3 and S4 were conducted in 1994/1995 and 1999-2001, respectively, and comprised independent samples of 4856 and 4261 subjects aged 25 to 74 years. Both cohorts were reinvestigated in the follow-up examinations F3 and F4 in 2004/2005 and 2006-2008, respectively, with 2974 and 3080 participants. Anthropometric variables and clinical parameters were determined at all examinations. For the present study, DNA methylation measurements from 196 people with newly diagnosed T2D and 196 controls matched for age (±2 years), sex, cohort and observation time till diagnosis of diabetes.

Avon Longitudinal Study of Parents and Children (ALSPAC)

ALSPAC is a large, prospective cohort study based in the South West of England. 14,541 pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were recruited and detailed information has been collected on these women and their offspring at regular intervals.[3, 4] The study website contains details of all the data that is available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/). Weight and height were measured with the participant in light clothing and without shoes. Weight was measured to the nearest 0.1kg with Tanita scales and height to the nearest 0.1cm using a Harpender stadiometer. Body mass index (kg/m^2) was then calculated. A Lunar Prodigy dual energy x-ray absorptiometry (DXA) scanner (GE Medical Systems Lunar, Madison, WI) was used to quantify fat mass (kg), lean mass (kg) in to total, android, gynoid and truncal distributions.

As part of the ARIES (Accessible Resource for Integrated Epigenomic Studies, http://www.ariesepigenomics.org.uk/) project, Infinium HM450 BeadChip data has been generated in 1,018 mother-offspring pairs in the ALSPAC cohort. The ARIES participants were selected based on availability of DNA samples at two time points for the mother (antenatal and at follow-up when the offspring were adolescents) and three time points for the offspring (neonatal, childhood [age 7 years] and adolescence [age 15-17 years]). Methylation samples of the offspring at age 15-17 are included in this analysis.

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

Quantification of DNA methylation

Epigenome wide association

Methylation of genomic DNA was initially quantified in 2,687 samples using the Illumina HumanMethylation450 array (450K array) according to manufacturer's instructions. Bisulfite conversion of genomic DNA was performed using the EZ DNA methylation kit according to manufacturer's instructions (Zymo Research, Orange, CA). Case and control samples were distributed randomly in all experiments. Bead intensity was retrieved using the minfi software package, and a detection P value of $<10^{-16}$ was used for marker calling. Of the 486,427 positions assayed by the array, we excluded markers with call rates <98% (N=4,684), or that do not measure methylation at CpG sites (N=4,006). This left 466,186 autosomal and 11,329 sex chromosome markers for analysis. Markers reported to be cross-hybridising were retained, but flagged. In total 23 samples were excluded; 18 for low marker call rate (<98%), and 5 for gender inconsistency. This left 2,664 discovery samples (1,074 cases and 1,590 controls) for analysis.

Replication and further testing

Replication testing in samples from the LOLIPOP study was carried out by pyrosequencing using biotinylated primers to amplify bisulfite-treated DNA (**Supplementary Table 1**). The biotinylated PCR products were then immobilized on streptavidin-coated Sepharose beads (GE Healthcare, Orsay, France). Pyrosequencing was performed with the PyroMark Q96 MGMT kit (Qiagen, Courtaboeuf, France) on a PSQTM96 MA system (Biotage, Uppsala, Sweden). Replication testing in samples from the KORA study was done using the Illumina HumanMethylation450 BeadChip with genomic DNA (750ng) as described.[5]

The Illumina HumanMethylation450 BeadChip was also used to measure DNA methylation in the ALSPAC cohort, and methylation data pre-processed using R (version 3.0.1), with background correction and subset quantile normalisation performed using the pipeline described by Touleimat and Tost.[6]

Targeted resequencing for regional methylation

The 450K array assays <2% of the estimated ~30M CpG sites in the human genome. To better describe the patterns of regional methylation we carried out re-sequencing of the *TXNIP* locus in 172 samples. We initially used sequence capture and next generation sequencing to assay 99 of the 133 predicted CpG sites within 5kb of the sentinel methylation marker at the *TXNIP* locus (chr 1, bp 145,436,694-145,446,572). We supplemented this with pyrosequencing of 14 CpG sites adjacent to the sentinel marker, both to improve local coverage and to replicate the findings from next-generation sequencing.

Primers were designed using Sequenom EpiDesigner BETA (www.epidesigner.com). Target DNA enrichment was done using the Fluidigm 48.48 Access Array IFC System, followed by PCR to attach sequence-specific adapters and sample barcodes. Pooled sequencing was done using the Illumina MiSeq platform (150bp paired-end runs). We then used the Burrows-Wheeler Aligner to map the directional, paired-end Illumina sequencing reads to the reference genome (hg19 build),[7] then quantified methylation from the frequencies of converted and unconverted cytosine residues observed in reads mapped to each CpG site.

We quantified the pairwise correlation between methylation at the sentinel CpG site with each of the additional CpG sites assayed. We carried forward 8 CpG sites showing r>0.5 with the sentinel marker for pyrosequencing amongst 238 incident T2D cases and 382 controls selected at random from the discovery samples, to quantify their association with T2D (logistic regression) both as single markers and in aggregate (mean methylation across the sites assayed).

Statistical analyses of the Epigenome-wide and replication data

Epigenome-wide data were analysed in R version 2.15 using minfi and other R scripts.[8] Marker intensities were quantile normalised for analysis. Principal component analysis of marker intensities was performed to assess for any cryptic structure in the data. A differential white blood cell count was available for all participants, providing information on total white blood cells, and also lymphocyte, monocyte and granulocyte counts. The epigenome-wide methylation values were used to impute a further 4 lymphocyte subsets (CD4, CD8, NK and B cells).[9] We performed single marker tests using logistic regression to examine the association of each autosomal CpG site with T2D, adjusted for age and gender. We also included intensity values from the 450K array control

probes, bisulfite conversion batch, measured white cells and imputed white cell subsets, and the first 5 principal components as covariates in the regression models (**Supplementary Table 3**), as these progressively reduced test statistic inflation.[10-12] Finally we corrected the association results for the genomic control inflation factor. Markers on the sex chromosomes were tested similarly for association with T2D, but separately in men and women.

We also used logistic regression to test the association of DNA methylation with T2D in the replication stage. For KORA samples, raw methylation data were extracted with Illumina® GenomeStudio Version 2011.1, Methylation Module 1.9.0 and preprocessed using R, version 3.0.1 (R Core Team 2013). Colour bias adjustment and background correction were performed with the R package lumi version 2.12.0.[13] Data were normalized using beta-mixture quantile normalization (BMIQ).[14] Association of DNA methylation and incident diabetes were calculated using a conditional logistic regression using R package survival, version 2.37.4.[15] Results were combined across the discovery and replication stages by inverse variance meta-analysis. Epigenome-wide significance was set at $P<1x10^{-7}$ providing Bonferroni correction for the 466,186 autosomal markers tested. Our choice of threshold was supported by the results of permutation testing. In brief we permuted (randomized) the T2D case-control labels to remove true biological relationships, and reassessed association with DNA methylation. We repeated this 1000 times to generate robust estimates of the distribution of P values, and corresponding confidence intervals. Results of permutation testing show that the associations with permuted T2D case-control status follow the distribution expected under the null hypothesis almost exactly (lambda=1.00, **Supplementary Figure 1**).

To combine information across genetic loci, we calculated a Methylation Score as the sum of the standardised methylation values at each marker, weighted by marker-specific effect size.

Further testing for association with body fat distribution

The association of DNA methylation with body fat distribution was carried out amongst 972 participants of the ALSPAC study. We performed linear regression to examine the association of the identified CpG sites with measures of anthropometry i) adjusted for age and sex ii) adjusted for BMI, age and sex iii) stratified by BMI. Each model was additionally adjusted for batch (bisulphite conversion plate) and the first five principal components from array beta values.

Association of methylation with gene expression in peripheral blood.

The relationship between methylation and gene expression in peripheral blood leucocytes was investigated in samples from three cohorts: LOLIPOP (Indian Asians, N=907), KORA (Europeans, N=703) and the EnviroGenoMarkers (EGM, European, N=591)

LOLIPOP. Details on the LOLIPOP cohort and methylation analysis has been described earlier. Gene expression analysis was performed with the Illumina HumanHT-12 v4 BeadChip array according to manufacturer's protocol. Background correction using negative controls was performed, and subsequently quantile normalised and log2 transformed.[16, 17] Linear models were fitted with log transformed gene expression as response variable, and quantile-normalisd beta values (methylation), age, sex, top 24 control probe PCs from methylation measurement, and technical covariates related to the expression measurement including RIN, RNA extraction batch, RNA conversion batch, scanning batch, array and array position. Calculations were performed using R, version 3.0.1.

KORA. Details on the KORA cohort have been described earlier. For a subset of 703 KORA F4 subjects, both methylation (Illumina 450k) and gene expression data were available. Gene expression analysis was performed by the Illumina HumanHT-12 v3 BeadChip array, with blood sample collection, RNA isolation and preparation as well as gene expression measurement described in detail elsewhere.[18] Linear models were fitted with log transformed gene expression as response variable, and DNA methylation, age, sex, smoking state (categorized as smoker, former smoker and never smoker), physical activity (categorized as active, inactive), alcohol intake, the top 20 control probe PCs from methylation measurement and three technical covariates related to the expression measurement, namely amplification batch, sample storage time and RNA integrity number (RIN) as covariates.[18] Calculations were performed using R, version 3.0.1.

EGM. The EnviroGenoMarkers (EGM) project is a nested case-control study of incident breast cancer and B-cell leukaemia.[19] Methylation and gene-expression were quantified in the baseline blood samples collected 1-17 years prior to disease onset. Transcriptomics profiles were obtained

using the Agilent 4x44K Whole Human Genome Microarray and subjected to extensive quality control procedures. DNA methylation profiles were obtained using the 450K array according to the manufacturer's protocol. Bisulphite conversion was carried out using the Zymo EZ DNA Methylation Kit. Probes that had missing values in more than 20% of the samples were excluded. The final dataset consisted of 29,662 transcripts and 432,633 DNA methylation probes. We used linear regression to determine the association between methylation and expression of nearby genes (1MB). We inferred statistical significance at $P<0.05$ after correction for the number of marker-eQTL comparisons made.

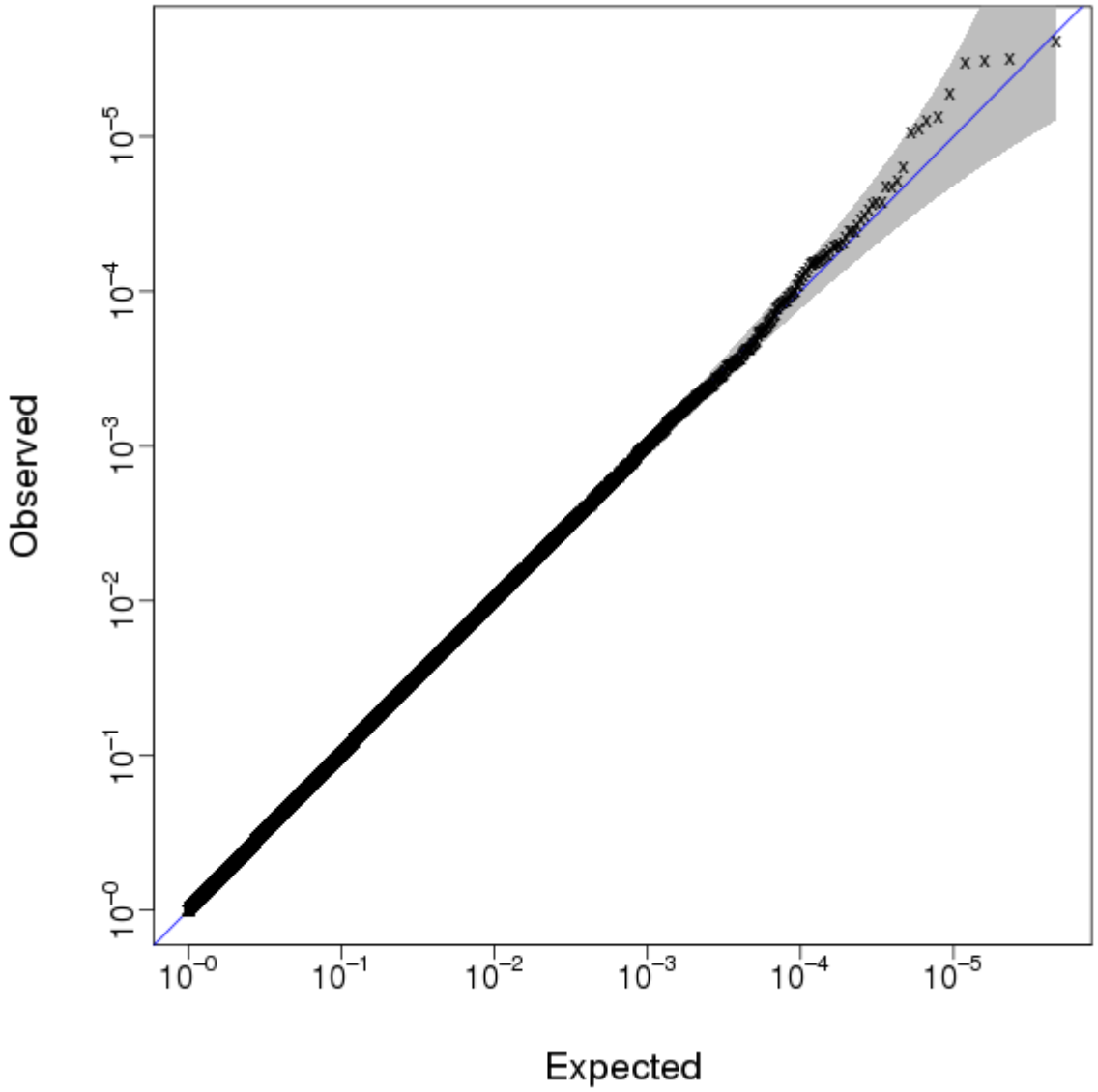
Association of methylation with gene expression in liver.

The relationship between methylation and gene expression in liver was investigated in samples from two European cohorts:

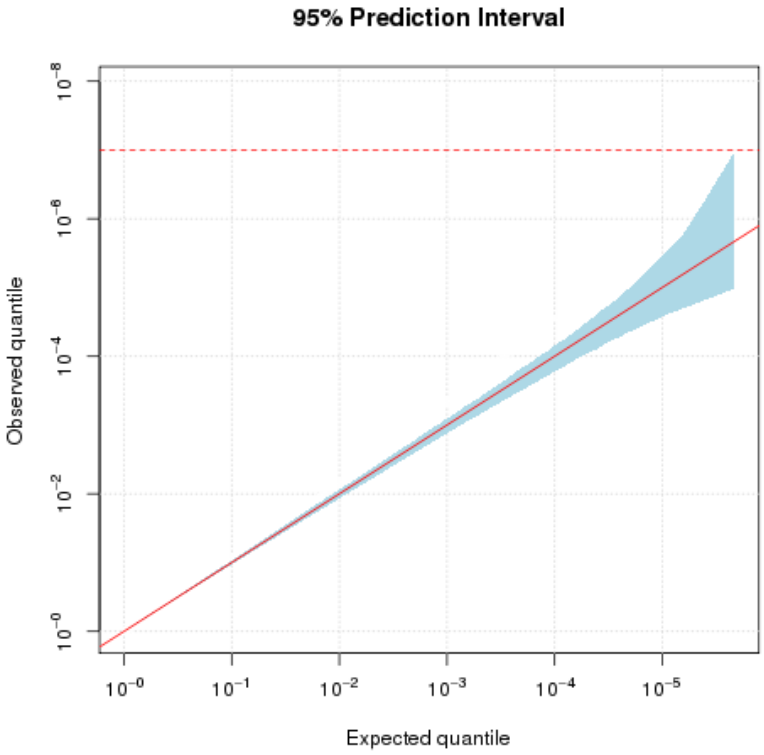
Liver samples 1. Liver samples were obtained percutaneously for patients undergoing liver biopsy for suspected NAFLD or intraoperatively for assessment of liver histology. Normal control samples were recruited from samples obtained for exclusion of liver malignancy during major oncological surgery. None of the normal control individuals underwent pre-operative chemotherapy and liver histology demonstrated absence of both cirrhosis and malignancy Study design, sampling method and data collection have been described in detail elsewhere.[20] For methylation analysis, bisulfite conversion was performed using the Zymo EZ DNA Methylation Kit (Zymo Research, Orange, CA, USA), and hybridization of the Illumina HumanMethylation450 array (Illumina, SanDiego, CA). mRNA expression analysis was performed using the HuGene 1.1 ST gene (Affymetrix, Santa Clara, Ca, USA) according to the manufacturers protocols. Hybridization signals were analyzed using GenomeStudio software (default settings; GenomeStudio ver. 2011.1, Methylation Analysis Module ver. 1.9.0; Illumina Inc) and internal controls for normalization.

Liver samples 2. The "*Atlas Biologique de l'Obésité Sévère*" (ABOS) cohort (ClinicalGov NCT01129297) is a prospective study conducted in the "Département de Chirurgie Générale et Endocrinienne, CHRU de Lille". [21] [20] [19] [18] [17] [16][16] This collection involves obese patients who underwent bariatric surgery and who gave their informed consent for sample collection during their intervention. Selected individuals were unrelated, women, above 35 years of age, of European origin, morbidly obese (body mass index ≥ 40 kg/m^2), non-smoker, non-drinker, without any history of hepatitis, and without any liver damage. RNeasy® Lipid Tissue kit (Qiagen) was used to isolate RNA from liver tissue and subsequently analysed on the HumanHT-12 v4 Expression BeadChips (Illumina). Genome-wide DNA methylation profiling was performed with the Illumina HumanMethylation450 array (Illumina, SanDiego, CA), using 500 ng DNA from human liver cells with bisulfite conversion by the EZ DNA Methylation Kit D5001 (Zymo Research, Orange, CA, USA) according to the manufacturer's instructions. A Beta-mixture quantile normalization of the methylation scores was subsequently applied for correcting probe design bias.[14] Following the QC procedure, averaged methylation scores at each site were compared between T2D cases and controls using an empirical Bayesian method introduced by Smyth and colleagues: the Limma model.[17] The correlation between DNA methylation in blood and liver was also directly determined using samples from the ABOS study.

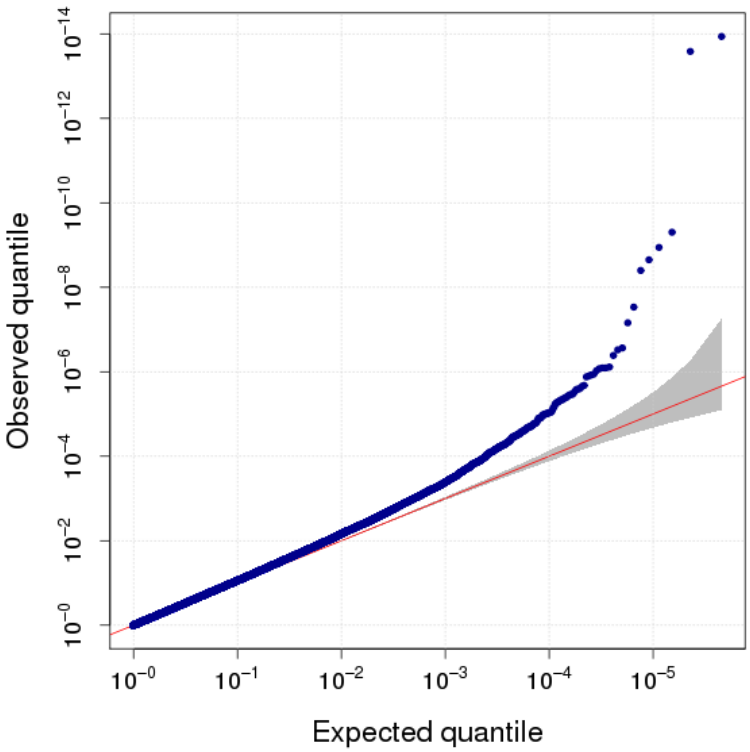
Supplementary Figure 1. Quantile-Quantile (QQ) plot for the comparison of 36 technical duplicates. We used linear regression to test quantify the differences between the first and second repeats of the 36 samples run in duplicate, including illumina control-probe intensities included as covariates in the model. There is no evidence for a batch effect, with no systematic difference between the duplicate samples beyond that expected under the null hypothesis (Genomic Inflation Factor lamda=1.03).



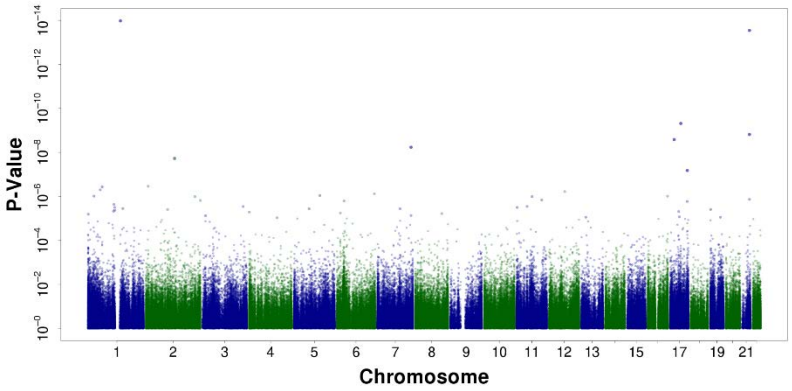
Supplementary Figure 2. Results of permutation testing (N=1000) to establish the expected distribution of P values under the null hypothesis in epigenome wide association using the Illumina 450K methylation array



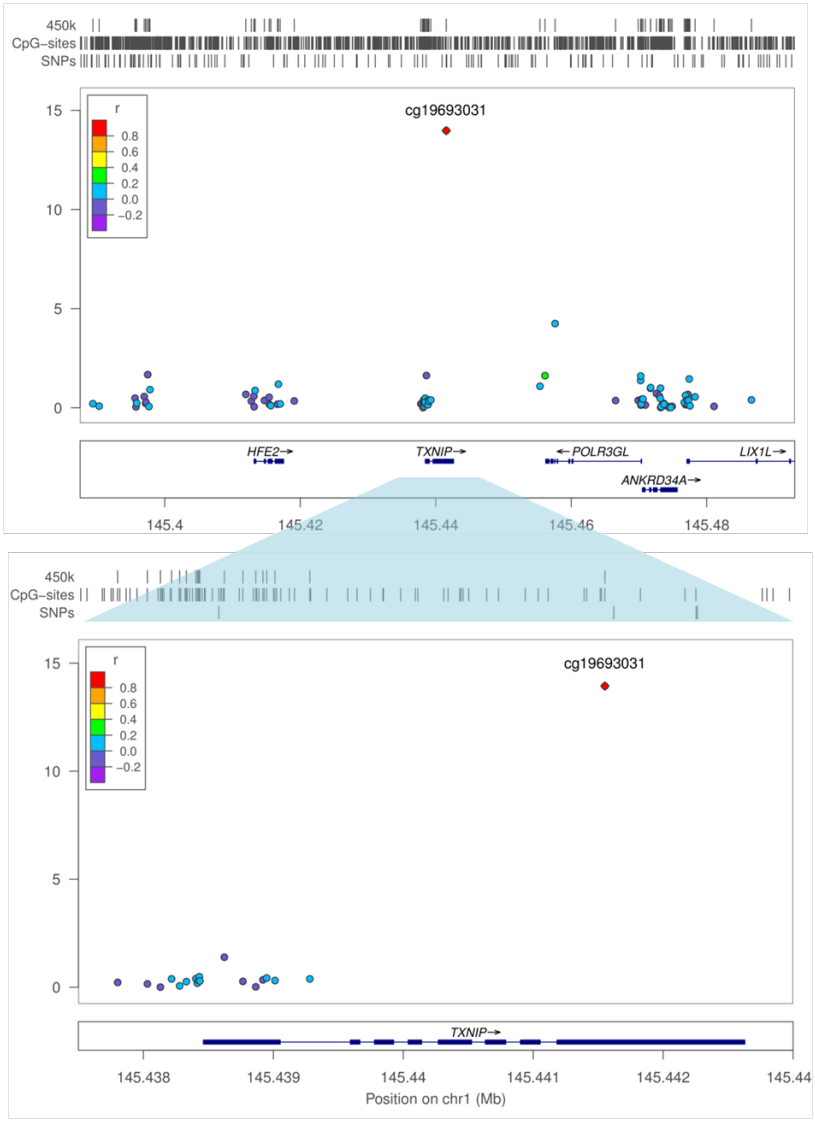
Supplementary Figure 3. QQ plot showing results for the Epigenome-wide association study. Blue dots represent results for individual CpG sites for association with T2D. The red line represents the null hypothesis of no association; the grey shading provides the 95% confidence interval for the null hypothesis.



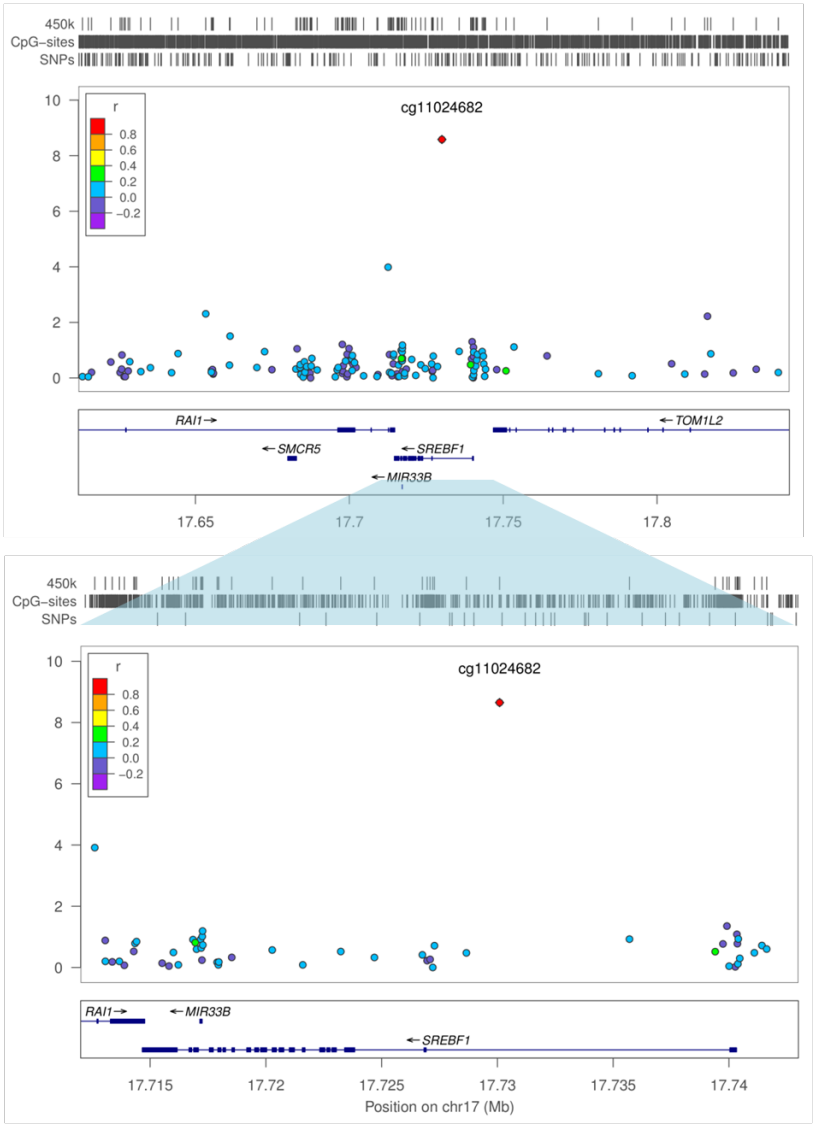
Supplementary Figure 4. Manhattan plot for the Epigenome-wide association study



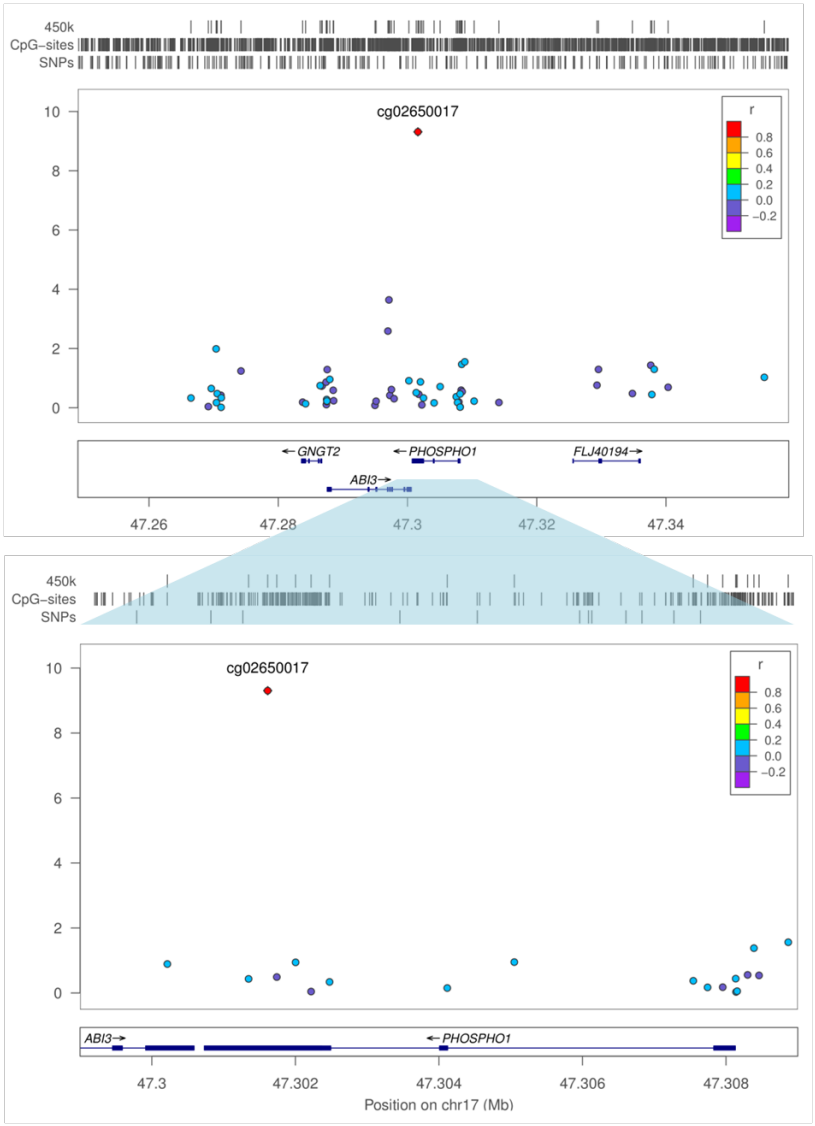
Supplementary Figure 5.1 – EWAS results at the *TXNIP* locus



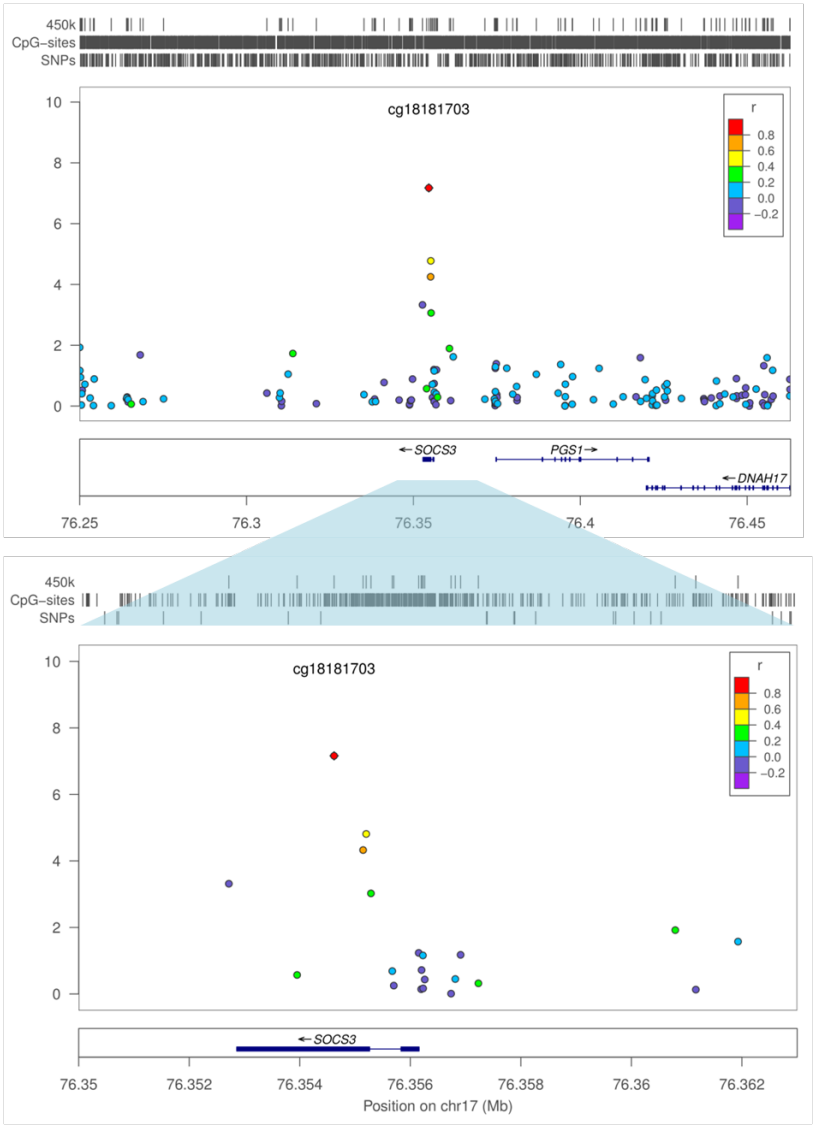
Supplementary Figure 5.2 – EWAS results at the *SREBF1* locus



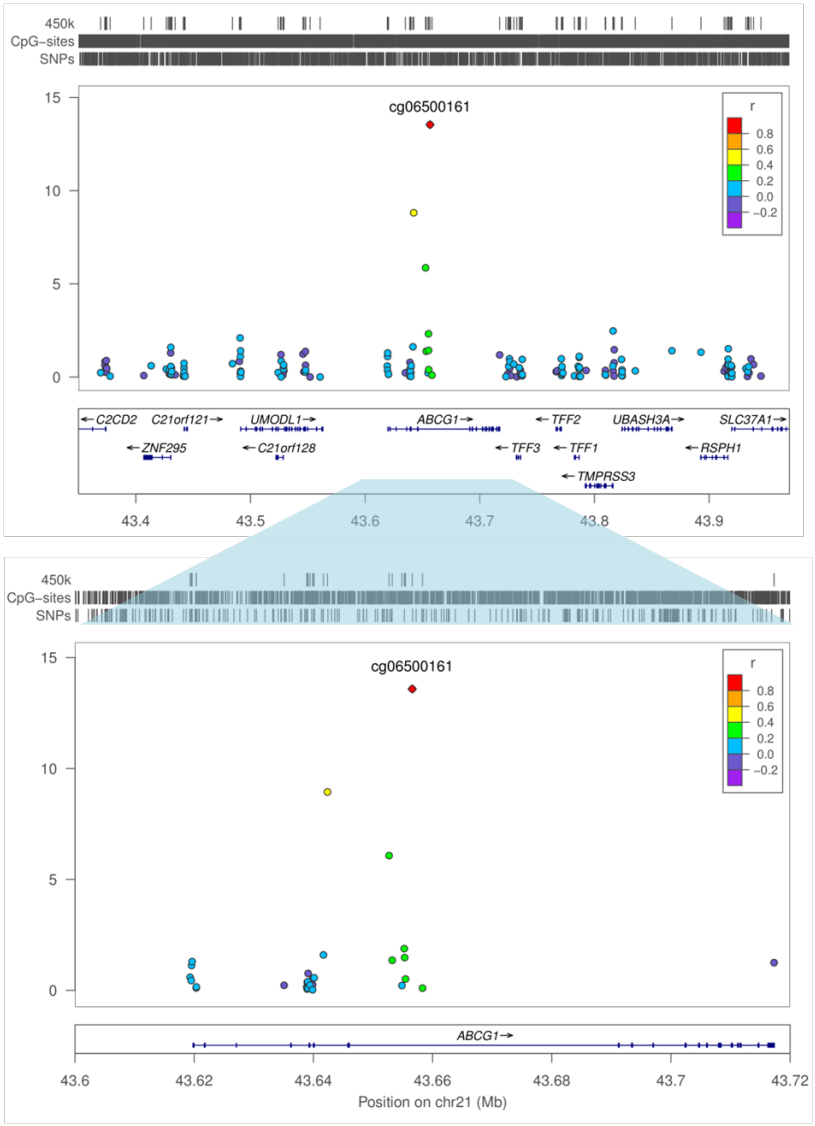
Supplementary Figure 5.3 – EWAS results at the *PHOSPHO1* locus



Supplementary Figure 5.4 – EWAS results at the *SOCS3* locus



Supplementary Figure 5.5 – EWAS results at the *ABCG1* locus



Supplementary Table 1. Primers used for pyrosequencing in replication testing

Locus	Methylation marker	Primer	Sequence
TXNIP	cg19693031	Forward	Bio-ATGTTGTGATATAGGGGTGTTTGT
		Reverse	ACAAAAAACCTCCAAAAAACCTTA
		Sequencing	AAAAACCTTAAAAAACTTAC
PROC	cg09152259	Forward	Bio-GTTTTGAGGATAGTAGAGTGTTTG
		Reverse	AAATAACCTCAAAAACTTCACATT
		Sequencing	CTTCACATTATAAACTACCA
C7orf29	cg04999691	Forward	TTGGGAATTTTGTTTTGGATG
		Reverse	Bio-AATAAAAAAACAACCCTCACCC
		Sequencing	GTTAGTTTGTGGTTGGG
SREBF1	cg11024682	Forward	TTTTTTTTGAAGGTAGATGTAGG
		Reverse	Bio-AAAACAAATAAAAAACTCCCTCTTC
		Sequencing	TTTTTTTTGAAGGTAGATGTAGG
PHOSPHO1	cg02650017	Forward	Bio-GGATATAATAAAGTTAAAGGGAAAAGG
		Reverse	CTCATTCAAAAAACCCAAAAAAC
		Sequencing	ACCTACAACAAATACTAAAA
SOCS3	cg18181703	Forward	TTGGGTGATTTTTTTATAGGAGTT
		Reverse	Bio-TCCCCCCAAAAAAACCTATT
		Sequencing	GAGATGTTGAAGAGTGGTTA
ABCG1	cg06500161	Forward	TAAGGTTTGGGGTTATTTTAGTGG
		Reverse	Bio-CCTAAATTATACCCAAAAAAACTA
		Sequencing	TAGGGTTTTTTTTAGATAT

Supplementary Table 2. Characteristics of participants for samples used in the analysis of the relationship between DNA methylation with gene expression. See Supplementary note for additional sample descriptives. Results are presented as mean (SD) or as %.(N).

	LOLIPOP	**KORA**	**EGM**	**Liver1**	**Liver2**
Population	Indian Asian	European	European	European	European
Tissue	Blood	Blood	Blood	Liver	Liver
N	907	703	591	69	47
Age (yrs)	54.0 (10.3)	68.9 (4.4)	52.1 (7.8)	45.7 (10.8)	48.3 (8.5)
Sex (M)	45.6% (414)	50.5% (355)	37.4% (221)	18.8% (13)	0.0% (0)
T2D (%)	19.5% (177)	12.9% (91)	0.7% (4)	13.0% (9)	48.9% (23)
Body mass index (kg/m^2)	27.1 (4.3)	28.8 (4.5)	25.9 (3.6)	40.9 (11.8)	48.9 (8.5)

Supplementary Table 3. Covariates used in epigenome-wide association regression models.

Category	**Covariate**
Clinical	Age, Gender
White cell subsets	White blood cell count, CD8 T cells, CD4 T cells, Natural Killer cells, B-cells, Monocytes, Granulocytes
Principal components	PC1-5
Technical controls	Bisulfite Batch, BSCI.Green, BSCI.Red, BSCII.Red, stain.Red, stain.Green, extensionA.Red, extensionT.Red, extensionC.Green, extensionG.Green, hybridH.Green, hybridM.Green, hybridL.Green, target.Green, specI.Green, specI.Red, specII.Red, np.A.Red, np.T.Red, np.C.Green, np.G.Green, normA.Red, normT.Red, normC.Green, normG.Green

Supplementary Table 4. Characteristics of Indian Asian incident T2D cases and controls in the LOLIPOP study. Results are presented as mean (SD) or as % (N).

	Europeans	**South Asians**	**P**
N	7,066	13,535	
Follow-up (yrs)	8.8 (1.7)	8.4 (1.7)	0.94
New onset T2D	4.3% (304)	11.9% (1,611)	<0.0001
Age (yrs)	52.2 (11.4)	49.1 (10.9)	<0.0001
Sex (M)	61.2% (4,324)	60.4% (8,175)	0.12
Family history of T2D	17.8% (1,258)	36.1% (4,886)	<0.0001
Physical activity	51.5% (3,639)	31.2% (4,223)	<0.0001
Impaired fasting glucose	6.0% (424)	6.4% (866)	0.28
Fasting glucose (mmol/L)	5.07 (0.52)	5.09 (0.52)	0.01
HbA1c (%)	5.29 (0.43)	5.53 (0.45)	<0.0001
Insulin (IU/L)	9.2 (7.8)	11.9 (8.7)	<0.0001
HOMA-IR	2.2 (2.0)	2.8 (2.2)	<0.0001
HOMA-B	120 (207)	156 (115)	<0.0001
Body mass index (kg/m^2)	27.2 (4.9)	27.0 (4.4)	0.01
Waist circumference (cm)	94.3 (13.3)	95.1 (11.4)	<0.0001
Waist-hip ratio	0.94 (0.07)	0.97 (0.07	<0.0001
Treated hypertension	18.6% (1,314)	21.7% (2937)	<0.0001
Systolic BP (mmHg)	130.4 (19.1)	128.7 (18.7)	<0.0001
Diastolic BP (mmHg)	79.4 (10.6)	80.0 (10.8)	<0.0001
Cholesterol (mmol/L)	5.42 (1.09)	5.29 (1.04)	<0.0001
Triglycerides (mmol/L)	1.44 (1.06)	1.60 (1.08)	<0.0001
HDL cholesterol (mmol/L)	1.44 (0.38)	1.30 (0.32)	<0.0001
Smoking			
Never smoked	42.2% (2,982)	81.8% (11,058)	<0.0001
Ex-smoker	32.1% (2,268)	8.1% (1,096)	
Current smoker	25.7% (1,816)	10.2% (1,381)	

Supplementary Table 5. Characteristics of Indian Asian incident T2D cases and controls in the epigenome-wide association study. Results are presented as mean (SD) or as % (N).

	Incident T2D	**Controls**	**P**
N	1,074	1,590	
Follow-up (yrs)	8.6 (1.7)	8.4 (1.7)	0.94
Age (yrs)	52.5 (10.2)	49.9 (9.8)	<0.0001
Sex (M)	67.3% (722)	68.2% (1,083)	0.61
Impaired fasting glucose	18.9% (207)	3.4% (55)	<0.0001
Fasting glucose (mmol/L)	5.49 (0.59)	5.05 (0.47)	<0.0001
HbA1c (%)	5.77 (0.49)	5.37 (0.48)	<0.0001
Insulin (IU/L)	15.5 (10.9)	10.7 (9.2)	<0.0001
HOMA-IR	3.8 (2.9)	2.4 (2.0)	<0.0001
HOMA-B	164 (115)	147 (207)	0.02
Body mass index (kg/m^2)	28.9 (4.6)	26.7 (3.9)	<0.0001
Waist circumference (cm)	101.0 (11.5)	94.9 (10.3)	<0.0001
Waist-hip ratio	0.97 (0.07	0.94 (0.07)	<0.0001
Treated hypertension (%)	39.2% (421)	23.0% (366)	<0.0001
Systolic BP (mmHg)	134.5 (19.0)	129.6 (18.6)	<0.0001
Diastolic BP (mmHg)	82.9 (11.1)	81.1 (10.4)	<0.0001
Cholesterol (mmol/L)	5.32 (1.08)	5.43 (0.98)	0.01
Triglycerides (mmol/L)	1.90 (1.28)	1.61 (0.96)	<0.0001
HDL cholesterol (mmol/L)	1.21 (0.26)	1.30 (0.29)	<0.0001
Isoleucine (mmol/L)	0.069 (0.018)	0.061 (0.017)	<0.0001
Phenylalanine (mmol/L)	0.101 (0.016)	0.094 (0.014)	<0.0001
Tyrosine (mmol/L)	0.060 (0.012)	0.055 (0.012)	0.0001
Smoking			
Never smoked	80.4% (863)	84.0% (1,338)	0.03
Ex-smoker	10.3% (110)	7.5% (118)	
Current smoker	9.3% (101)	8.5% (134)	
Measured white cell subsets			
Lymphocytes (%)	34.5 (7.8)	34.3 (7.9)	0.61
Monocytes (%)	6.2 (1.9)	6.2 (2.0)	0.99
Neutrophils (%)	54.2 (8.4)	54.2 (8.5)	0.89
Eosinophls (%)	2.4 (3.9)	3.9 (2.6)	0.06
Basophils (%)	0.5 (1.0)	1.0 (0.6)	0.38
Estimated white cell subsets			
CD8T (%)	17.3 (6.1)	17.1 (5.9)	0.47
CD4T (%)	15.3 (5.1)	15.4 (4.9)	0.38
Natural Killer cells (%)	0.2 (2.1)	0.6 (2.2)	<0.0001
B-cells (%)	6.4 (2.4)	6.3 (2.3)	0.15
Monocytes (%)	10.5 (1.9)	10.3 (1.9)	0.01
Granulocytes (%)	51.0 (7.0)	50.8 (7.5)	0.59

Supplementary Table 6: Number of methylation markers associated with T2D across the range of P-value thresholds in the epigenome-wide association study. Expected is estimated as: number of tests (~466,000) x [P value threshold]

P-value threshold	No. of markers	
	Observed	Expected
$P<10^{-7}$	7	0.05
$P<10^{-6}$	10	0.5
$P<10^{-5}$	33	4.7
$P<10^{-4}$	124	47
$P<10^{-3}$	640	466
P<0.01	4892	4663
P<0.05	22975	23309

Supplementary Table 7. Methylation values at the 7 sentinel methylation markers in the discovery and replication datasets. Results presented as mean (SD).

		EWAS discovery (IA)			LOLIPOP replication (EW)			KORA replication (EW)			
CpG ID	**Locus**	**Case**	**Control**	**Diff**	**Case**	**Control**	**Diff**	**Case**	**Control**	**Diff**	**Mean diff**
cg19693031	*TXNIP*	73.1 (4.6)	72.0 (5.1)	1.1 (0.2)	71.0 (5.4)	70.1 (5.5)	0.9 (0.5)	78.2 (5.7)	75.6 (6.9)	2.6 (0.6)	1.2 (0.2)
cg09152259	*PROC*	33.0 (4.6)	31.9 (4.4)	1.1 (0.2)	30.7 (7.9)	30.3 (7.6)	0.4 (0.7)	37.1 (5.8)	36.7 (5.8)	0.4 (0.6)	1.0 (0.2)
cg04999691	*C7orf29*	71.3 (2.6)	70.9 (2.6)	0.4 (0.1)	77.4 (4.7)	77.5 (4.3)	-0.1 (0.4)	76.8 (3.6)	76.4 (3.6)	0.4 (0.4)	0.4 (0.1)
cg11024682	*SREBF1*	45.9 (3.1)	46.8 (3.1)	-0.9 (0.1)	46.0 (3.7)	47.1 (3.6)	-1.1 (0.3)	48.8 (3.7)	49.8 (3.8)	-1.0 (0.4)	-0.9 (0.1)
cg02650017	*PHOSPHO1*	8.9 (1.6)	8.4 (1.4)	0.5 (0.1)	4.3 (1.7)	3.9 (1.5)	0.4 (0.1)	4.2 (1.2)	4.1 (1.3)	0.1 (0.1)	0.4 (0.1)
cg18181703	*SOCS3*	44.2 (4.0)	43.1 (3.9)	1.1 (0.2)	48.4 (7.3)	47.4 (7.4)	1.0 (0.6)	49.7 (4.3)	48.0 (5.4)	1.7 (0.5)	1.1 (0.1)
cg06500161	*ABCG1*	57.9 (2.6)	58.8 (2.6)	-0.9 (0.1)	55.4 (8.2)	56.8 (8.9)	-1.4 (0.7)	61.1 (3.2)	62.4 (3.2)	-1.3 (0.3)	-0.9 (0.1)

Supplementary Table 8. Characteristics of European white incident T2D cases and controls in the replication study. Results are presented as mean (SD) or as % (N).

	LOLIPOP			KORA		
	Incident T2D	**Controls**	**P**	**Incident T2D**	**Controls**	**P**
N	181	568		196	196	
Age (yrs)	60.7 (8.7)	60.4 (9.7)	0.72	57.8 (8.9)	57.6 (8.9)	0.21
Sex (M)	74.5% (135)	73.2% (416)	0.75	54% (106)	54% (106)	1.00
Impaired fasting glucose	28.3% (51)	4.0% (23)	<0.0001	-	-	-
Fasting glucose (mmol/L)	5.7 (0.6)	5.1 (0.5)	<0.0001	-	-	-
Random glucose (mmol/L)	-	-	-	6.5 (2.2)	5.4 (0.7)	<0.0001
HbA1c (%)	5.7 (0.6)	5.4 (0.4)	<0.0001	5.8 (0.8)	5.3 (0.4)	<0.0001
Body mass index (kg/m^2)	30.8 (5.6)	27.3 (4.1)	<0.0001	30.9 (4.8)	27.5 (4.0)	<0.0001
Waist circumference (cm)	104.9 (13.4)	95.9 (11.5)	<0.0001	101.6 (12.6)	91.8 (12.0)	<0.0001
Waist-hip ratio	0.97 (0.07)	0.93 (0.07)	<0.0001	0.93 (0.08)	0.88 (0.09)	<0.0001
Treated hypertension	48.4% (88)	26.9% (153)	<0.0001	67% (131)	49% (96)	0.0002
Systolic BP (mmHg)	140.8 (19.3)	137.8 (19.6)	0.07	140.3 (19.3)	135.0 (20.2)	0.01
Diastolic BP (mmHg)	83.4 (9.9)	81.5 (10.4)	0.03	84.1 (11.9)	82.1 (10.8)	0.09
Cholesterol (mmol/L)	5.49 (1.16)	5.48 (1.04)	0.87	6.1 (1.1)	6.3 (1.1)	0.25
Triglycerides (mmol/L)	2.00 (1.20)	1.45 (0.87)	<0.0001	2.5 (1.6)	2.0 (2.2)	0.01
HDL cholesterol (mmol/L)	1.27 (0.29)	1.44 (0.37)	<0.0001	1.2 (0.3)	1.4 (0.4)	<0.0001
Smoking						
Never smoked	34.2% (62)	42.4% (241)		40% (78)	51% (100)	
Ex-smoker	45.7% (83)	36.1% (205)	0.06	35% (69)	35% (69)	0.01
Current smoker	20.1% (36)	21.5% (122)		25% (49)	14% (27)	

Supplementary Table 9. Replication testing for association with future T2D amongst 1,141 Europeans. Results are expressed as OR (95%CI) per 1% increase in methylation, 1SD increase in Methylation score, or for Q4 vs Q1 of distribution, where Q1 is the quartile with lowest T2D risk. P is the P-value in combined analysis. P_{het} is for the comparison of effect sizes between the LOLIPOP and KORA samples.

		LOLIPOP (N=749)		KORA (N=392)		Combined		
CpG	**Locus**	**Relative risk (95%CI)**	**P**	**Relative risk (95%CI)**	**P**	**Relative risk (95%CI)**	**P**	**P_{het}**
Relative risk per 1% increase in methylation / 1SD increase in methylation score								
cg19693031	*TXNIP*	0.96 (0.92-1.00)	0.05	0.92 (0.88-0.96)	$4.8x10^{-5}$	0.96 (0.94-0.98)	$2.5x10^{-5}$	0.12
cg09152259	*PROC*	0.99 (0.97-1.01)	0.43	0.99 (0.96-1.02)	0.54	0.99 (0.97-1.01)	0.32	0.96
cg04999691	*C7ORF29*	1.00 (0.97-1.03)	0.87	0.97 (0.92-1.03)	0.34	1.00 (0.98-1.02)	0.71	0.38
cg11024682	*SREBF1*	1.03 (0.99-1.07)	0.12	1.09 (1.03-1.16)	0.006	1.03 (1.01-1.05)	0.005	0.13
cg02650017	*PHOSPHO1*	0.88 (0.81-0.97)	0.007	0.82 (0.66-1.02)	0.07	0.97 (0.95-0.99)	0.001	0.86
cg18181703	*SOC3*	0.98 (0.96-1.01)	0.14	0.91 (0.86-0.96)	0.0005	0.97 (0.95-0.99)	0.002	0.04
cg06500161	*ABCG1*	1.02 (1.00-1.04)	0.06	1.16 (1.08-1.25)	$6.0x10^{-5}$	1.04 (1.02-1.06)	0.0001	0.03
Methylation score		1.63 (1.27-2.09)	0.0001	2.24 (1.70-2.96)	$1.1x10^{-8}$	1.88 (1.56-2.26)	$2.5x10^{-11}$	0.09
Relative risk in Q4 vs Q1 of methylation								
cg19693031	*TXNIP*	1.32 (0.83-2.11)	0.24	5.75 (1.99-16.63)	0.001	1.68 (1.09-2.58)	0.018	0.01
cg09152259	*PROC*	1.19 (0.73-1.95)	0.48	1.50 (0.72-3.11)	0.28	1.28 (0.85-1.93)	0.23	0.06
cg04999691	*C7ORF29*	1.00 (0.60-1.66)	1.00	1.25 (0.49-3.17)	0.64	1.05 (0.68-1.64)	0.82	0.68
cg11024682	*SREBF1*	1.54 (0.92-2.56)	0.10	1.57 (0.61-4.05)	0.35	1.54 (0.99-2.42)	0.06	0.97
cg02650017	*PHOSPHO1*	2.16 (1.26-3.69)	0.005	1.50 (0.53-4.21)	0.44	2.00 (1.24-3.22)	0.004	0.54
cg18181703	*SOC3*	1.59 (0.99-2.56)	0.05	4.75 (1.62-13.96)	0.005	1.90 (1.23-2.94)	0.004	0.07
cg06500161	*ABCG1*	1.13 (0.71-1.79)	0.60	4.00 (1.50-10.66)	0.006	1.42 (0.94-2.16)	0.10	0.02
Methylation score		2.16 (1.27-3.66)	0.0043	20.00 (2.68-149.0)	0.004	2.49 (1.50-4.15)	0.0005	0.04

Supplementary Table 10. Summary of current knowledge for candidate genes at the identified loci.

ABCG1	**ATP-binding cassette sub-family G member 1**. The protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters. ABCG1 is involved in macrophage cholesterol and phospholipid transport, and promotes cholesterol efflux to HDL. ABGG1 regulates cellular lipid homeostasis in other cell types, including pancreatic beta cells. ABCG1 is present in insulin granules; loss of ABCG1 leads to altered granule morphology, reduced granule cholesterol levels and impaired insulin secretion in mice. *Abcg1*$^{-/-}$ mice have impaired glucose tolerance and insulin secretion with normal insulin sensitivity.[22] ABCG1 expression is downregulated in humans with diabetes and upregulated by the insulin sensitizing agents such as thiazolidenediones.[23]
PHOSPHO1	**Phosphatase, orphan 1**. Encodes a phosphatase involved in the generation of inorganic phosphate for bone mineralization. It is highly expressed at sites of mineralization in bone and cartilage. The bones of *Phospho1*$^{-/-}$ mice are hypomineralized leading to spontaneous fractures.[24] PHOSPHO1 has not previously been associated with obesity, insulin action or diabetes.
SOCS3	**Suppressor of Cytokine Signaling 3**. Encodes a STAT-induced STAT inhibitor, a cytokine-inducible negative regulator of cytokine signaling. *SOCS3* expression is induced by various cytokines, and also by obesity induced Hypoxia-inducible Factor 1-alpha.[25] SOCS3 binds and inhibits the activity of JAK2 kinase. SOCS3 is also a major negative regulator of insulin signaling, and is implicated in the pathogenesis of obesity and associated metabolic abnormalities. *SOCS3* expression is increased in skeletal muscle in the setting of diet-induced and genetic obesity, inflammation, and hyperlipidemia.[26] Muscle-specific overexpression of *SOCS3* impairs systemic and muscle-specific glucose homeostasis and insulin action despite similar body weight.[26] In contrast *SOCS3*$^{-/-}$ mice are protected against obesity induced hyperinsulinemia and insulin resistance.[27]
SREBF1	**Sterol regulatory element binding transcription factor 1**. This gene 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. The protein is synthesized as a precursor that is attached to the nuclear membrane and endoplasmic reticulum. Following cleavage, the mature protein translocates to the nucleus and activates transcription by binding to the SRE1. SREBPF1 is the master transcriptional regulator of hepatic lipogenesis, capable of inducing the entire complement of genes necessary for the synthesis of monounsaturated fatty acids. Insulin activates *SREBPF1* by at least two mechanisms: it increases SREBP-1 transcription, and increases the processing of SREBPF1 from an inactive membrane-bound precursor to a soluble fragment capable of translocating to the nucleus to activate transcription. SREBPF1 is decreased in insulin-deficient states, such as fasting, but increased in feeding, obesity and insulin resistance.[28] Inhibition of SREBPF1 reverts hepatic steatosis.[29]

TXNIP	Thioredoxin Interacting Protein. *TXNIP* is a recognised to be a key component of pancreatic beta cell biology, nutrient sensing, energy metabolism and regulation of cellular redox. *TXNIP* expression is highly induced by glucose though activation of the carbohydrate response element-binding protein (ChREBP) which binds the *TXNIP* promoter.[30] TXNIP downregulates GLUT1, the major transmembrane glucose transporter, thereby acting as a negative feedback loop to regulate glucose entry and mitochondrial oxidative stress. *TXNIP* is one of the most glucose responsive genes expressed in human islets; in animal models *Txnip* is a mediator of glucotoxic beta cell death, while *Txnip* downregulation protects against and obesity-induced diabetes by preventing beta cell apoptosis and preserving beta cell mass.[31] *TXNIP* may also contribute to regulation of adiposity and energy expenditure through hypothalamic pathways.[32] Overexpression of *TXNIP* in Agouti-related peptide (Agrp) neurons predisposes to diet-induced obesity and adipose tissue storage by decreasing energy expenditure and spontaneous locomotion, without affecting food intake. Conversely, Agrp neuronal *TXNIP* deletion protects against diet-induced obesity and adipose tissue storage and improves fasting glucose and glucose tolerance.[33]

Supplementary Table 11. Combined analysis of discovery (epigenome-wide) and replication results in samples with incident T2D. Results are expressed as relative risk (RR) for T2D in Q4 vs Q1 of distribution, where Q1 is the quartile with lowest T2D risk. P is the P-value in combined analysis.

CpG	Locus	Indian Asians (discovery) RR (95%CI)	Europeans (replication) RR (95%CI)	Combined RR (95%CI)	P
cg19693031	*TXNIP*	2.08 (1.67-2.60)	1.68 (1.09-2.58)	1.99 (1.63-2.42)	8.2E-12
cg11024682	*SREBF1*	1.95 (1.57-2.44)	1.54 (0.99-2.42)	1.87 (1.53-2.28)	7.6E-10
cg02650017	*PHOSPHO1*	1.95 (1.56-2.44)	2.00 (1.24-3.22)	1.96 (1.60-2.40)	6.8E-11
cg18181703	*SOC3*	1.73 (1.39-2.16)	1.90 (1.23-2.94)	1.77 (1.45-2.15)	1.7E-08
cg06500161	*ABCG1*	2.41 (1.93-3.02)	1.42 (0.94-2.16)	2.14 (1.76-2.61)	3.5E-14
Methylation score		3.51 (2.79-4.42)	2.49 (1.50-4.15)	3.31 (2.69-4.09)	5.7E-29

Supplementary Table 12. Pearson correlation co-efficients between clinical measures and sentinel methylation markers amongst Indian Asians in the Epigenome-wide association study.

	Correlation co-efficients					**P values**				
	TXNIP	*SREBF1*	*PHOSPHO1*	*SOCS3*	*ABCG1*	*TXNIP*	*SREBF1*	*PHOSPHO1*	*SOCS3*	*ABCG1*
Age	-0.03	0.14	-0.02	-0.09	0.08	0.09	$7.8x10^{-14}$	0.20	$1.8x10^{-6}$	$2.0x10^{-5}$
Systolic BP	-0.16	0.15	-0.01	-0.01	0.07	$2.9x10^{-17}$	$1.0x10^{-14}$	0.75	0.48	$1.5x10^{-4}$
Diastolic BP	-0.16	0.11	-0.01	0.02	0.07	$2.5x10^{-17}$	$3.0x10^{-8}$	0.48	0.35	$3.7x10^{-4}$
Body mass index	-0.05	0.13	-0.14	-0.15	0.17	$8.3x10^{-3}$	$5.0x10^{-12}$	$1.4x10^{-12}$	$2.7x10^{-15}$	$1.2x10^{-17}$
Waist	-0.10	0.17	-0.12	-0.13	0.16	$4.1x10^{-7}$	$8.3x10^{-19}$	$6.8x10^{-10}$	$4.4x10^{-11}$	$1.2x10^{-16}$
Waist hip ratio	-0.13	0.17	-0.06	-0.06	0.12	$3.7x10^{-12}$	$8.8x10^{-18}$	$1.3x10^{-3}$	$1.1x10^{-3}$	$1.7x10^{-10}$
Glucose	-0.20	0.17	-0.04	-0.03	0.16	$4.2x10^{-26}$	$1.4x10^{-18}$	$4.2x10^{-2}$	0.14	$9.3x10^{-16}$
HbA1c	-0.03	0.05	-0.06	-0.10	0.11	0.07	$5.2x10^{-3}$	$1.0x10^{-3}$	$1.2x10^{-7}$	$2.0x10^{-8}$
Insulin	-0.10	0.16	-0.10	-0.09	0.18	$1.5x10^{-6}$	$1.8x10^{-16}$	$1.4x10^{-7}$	$6.7x10^{-6}$	$3.5x10^{-20}$
HOMAIR	-0.12	0.18	-0.10	-0.09	0.19	$9.3x10^{-10}$	$1.1x10^{-19}$	$1.4x10^{-7}$	$7.2x10^{-6}$	$3.9x10^{-23}$
HOMAB	0.01	0.08	-0.09	-0.07	0.10	0.69	$1.4x10^{-4}$	$1.0x10^{-5}$	$3.3x10^{-4}$	$3.8x10^{-7}$
Cholesterol	-0.12	0.04	0.08	0.09	-0.09	$1.9x10^{-9}$	$2.8x10^{-2}$	$7.5x10^{-5}$	$7.3x10^{-6}$	$2.9x10^{-6}$
Triglycerides	-0.20	0.20	0.01	-0.01	0.20	$8.6x10^{-26}$	$1.1x10^{-25}$	0.58	0.59	$1.3x10^{-24}$
HDL cholesterol	0.03	-0.08	0.09	0.05	-0.21	1.00	$1.8x10^{-5}$	$4.9x10^{-6}$	$1.3x10^{-2}$	$1.2x10^{-27}$
LDL cholesterol	-0.05	-0.03	0.06	0.09	-0.14	$8.5x10^{-3}$	0.18	$4.8x10^{-3}$	$5.5x10^{-6}$	$8.4x10^{-13}$
Isoleucine	-0.13	0.18	-0.02	-0.05	0.19	$6.6x10^{-9}$	$1.4x10^{-15}$	0.45	$2.8x10^{-2}$	$4.3x10^{-18}$
Phenylalanine	-0.02	0.07	-0.12	-0.11	0.12	0.46	$1.1x10^{-3}$	$2.0x10^{-7}$	$5.6x10^{-7}$	$1.2x10^{-7}$
Tyrosine	-0.07	0.11	-0.08	-0.11	0.10	$1.3x10^{-3}$	$2.1x10^{-6}$	$2.6x10^{-4}$	$4.8x10^{-7}$	$5.9x10^{-6}$

Supplementary Table 13. Characteristics of APLSAC participants.

Variable	Mean	SD
Age (yrs)	17.1	1
Female (%)	51.3	
Body mass index (kg/m2)	22.3	3.9
Height (m)	1.7	0.1
Weight (kg)	66.2	13.3
Total fat mass (kg)	17.1	9.9
Total lean mass (kg)	46.1	9.8
Android fat mass (kg)	1.2	0.9
Android lean mass (kg)	2.7	0.6
Gynoid fat mass (kg)	3.4	1.7
Gynoid lean mass (kg)	6.5	1.5
Trunk fat mass (kg)	8.5	5.4
Trunk lean mass (kg)	21.7	4.5
cg19693031	65.9%	8.6%
cg11024682	14.1%	4.0%
cg02650017	3.1%	1.1%
cg18181703	21.6%	6.7%
cg06500161	46.3%	9.6%

Supplementary Table 14. Relationships between DNA methylation and quantitative measures of adiposity amongst participants of the APLSAC study.

	TXNIP (cg19693031)		*SREBF1* (cg11024682)		*PHOSPHO1* (cg02650017)		*SOCS3* (cg18181703)		*ABCG1* (cg06500161)	
	Effect	**P**	**Effect**	**P**	**Effect**	**P**	**Effect**	**P**	**Effect**	**P**
Adjusted for Age and Sex										
Body mass index (kg/m2)	-5.2E-05	9.4E-01	6.0E-04	5.0E-02	-1.2E-05	8.9E-01	-8.1E-04	1.3E-01	2.1E-03	4.2E-03
Total fat mass (kg)	3.7E-05	8.9E-01	3.0E-04	1.9E-02	-1.9E-05	6.0E-01	-3.0E-04	1.8E-01	7.4E-04	1.5E-02
Total lean mass (kg)	1.4E-04	7.7E-01	-1.7E-04	4.4E-01	1.7E-04	5.7E-03	4.2E-04	2.6E-01	9.2E-04	7.2E-02
Android fat mass (kg)	-5.4E-04	8.6E-01	3.3E-03	1.7E-02	-2.4E-04	5.3E-01	-3.8E-03	1.2E-01	9.8E-03	3.1E-03
Android lean mass (kg)	6.2E-03	3.7E-01	-1.2E-03	7.2E-01	1.7E-03	6.7E-02	4.2E-03	4.6E-01	2.2E-02	4.4E-03
Gynoid fat mass (kg)	9.4E-04	5.8E-01	1.5E-03	5.1E-02	-7.9E-05	7.2E-01	-1.7E-03	2.3E-01	3.9E-03	4.0E-02
Gynoid lean mass (kg)	1.2E-03	6.8E-01	-8.8E-04	5.0E-01	9.6E-04	9.3E-03	1.7E-03	4.5E-01	5.5E-03	7.8E-02
Trunk fat mass (kg)	-3.5E-06	9.9E-01	5.6E-04	1.3E-02	-3.7E-05	5.6E-01	-5.6E-04	1.6E-01	1.5E-03	5.4E-03
Trunk lean mass (kg)	3.8E-04	6.9E-01	-7.2E-04	1.0E-01	3.3E-04	8.8E-03	8.5E-04	2.8E-01	2.1E-03	5.1E-02
Adjusted for age, sex and BMI										
Total fat mass (kg)	6.1E-04	3.8E-01	3.8E-04	2.4E-01	-1.1E-04	2.2E-01	7.3E-05	9.0E-01	-4.6E-04	5.5E-01
Total lean mass (kg)	2.4E-04	6.4E-01	-4.4E-04	6.3E-02	2.1E-04	1.7E-03	8.5E-04	4.2E-02	3.5E-04	5.4E-01
Android fat mass (kg)	6.2E-04	9.3E-01	4.3E-03	2.0E-01	-1.3E-03	1.7E-01	-2.7E-03	6.5E-01	5.7E-03	4.7E-01
Android lean mass (kg)	1.0E-02	2.3E-01	-7.0E-03	6.9E-02	2.4E-03	2.7E-02	1.3E-02	5.6E-02	1.4E-02	1.3E-01
Gynoid fat mass (kg)	6.3E-03	9.9E-02	5.6E-04	7.5E-01	-3.7E-04	4.6E-01	1.0E-03	7.5E-01	-5.3E-03	2.1E-01
Gynoid lean mass (kg)	1.7E-03	5.9E-01	-2.5E-03	8.1E-02	1.2E-03	3.2E-03	4.3E-03	9.5E-02	2.0E-03	5.7E-01
Trunk fat mass (kg)	6.4E-04	6.0E-01	8.5E-04	1.3E-01	-2.1E-04	1.8E-01	-1.1E-04	9.1E-01	3.5E-04	7.9E-01
Trunk lean mass (kg)	5.0E-04	6.2E-01	-1.1E-03	1.6E-02	3.8E-04	4.5E-03	1.4E-03	8.1E-02	1.2E-03	3.0E-01

Supplementary Table 15. BMI stratitified relationships between DNA methylation and quantitative measures of adiposity (APLSAC study). Associations reaching P<0.05 are highlighted

	TXNIP (cg19693031)		*SREBF1* (cg11024682)		*PHOSPHO1* (cg02650017)		*SOCS3* (cg18181703)		*ABCG1* (cg06500161)	
	Effect	**P**	**Effect**	**P**	**Effect**	**P**	**Effect**	**P**	**Effect**	**P**
Normal (BMI<25kg/m^2, N=777)										
Total fat mass (kg)	6.2E-04	3.3E-01	6.0E-04	4.4E-02	5.0E-05	5.5E-01	8.8E-04	8.9E-02	4.2E-04	5.4E-01
Total lean mass (kg)	2.4E-04	6.7E-01	-1.8E-04	4.9E-01	2.1E-04	3.5E-03	1.2E-03	9.3E-03	8.1E-04	1.8E-01
Android fat mass (kg)	5.7E-03	4.4E-01	7.3E-03	3.2E-02	4.9E-04	6.1E-01	7.2E-03	2.3E-01	8.7E-03	2.8E-01
Android lean mass (kg)	1.3E-02	1.6E-01	-2.3E-03	5.7E-01	2.6E-03	2.9E-02	1.8E-02	1.3E-02	2.2E-02	2.4E-02
Gynoid fat mass (kg)	5.2E-03	1.4E-01	2.9E-03	8.2E-02	3.6E-04	4.4E-01	4.3E-03	1.4E-01	8.6E-04	8.3E-01
Gynoid lean mass (kg)	1.2E-03	7.1E-01	-1.1E-03	5.0E-01	1.3E-03	3.9E-03	5.9E-03	3.0E-02	4.3E-03	2.4E-01
Trunk fat mass (kg)	8.8E-04	4.6E-01	1.2E-03	2.8E-02	9.3E-05	5.6E-01	1.7E-03	7.5E-02	1.6E-03	2.1E-01
Trunk lean mass (kg)	4.7E-04	6.8E-01	-7.1E-04	1.7E-01	4.2E-04	4.9E-03	2.1E-03	2.4E-02	2.3E-03	5.9E-02
Overweight (BMI>=25 and <30 kg/m^2, N=96)										
Total fat mass (kg)	-5.1E-04	6.7E-01	-5.2E-04	4.0E-01	1.2E-04	5.3E-01	-8.7E-04	3.9E-01	-1.5E-03	3.8E-01
Total lean mass (kg)	6.8E-05	9.7E-01	-4.3E-06	1.0E+00	2.4E-04	3.4E-01	-1.5E-04	9.1E-01	-2.2E-03	3.3E-01
Android fat mass (kg)	-1.4E-02	2.4E-01	-7.0E-03	2.5E-01	7.7E-04	6.9E-01	-4.8E-03	6.3E-01	-1.2E-03	9.5E-01
Android lean mass (kg)	7.2E-03	7.5E-01	-6.9E-03	5.7E-01	2.6E-03	4.9E-01	1.4E-02	4.8E-01	-3.9E-02	2.4E-01
Gynoid fat mass (kg)	3.6E-03	6.1E-01	-5.0E-03	1.8E-01	1.1E-03	3.4E-01	-4.9E-03	4.2E-01	-1.5E-02	1.5E-01
Gynoid lean mass (kg)	1.6E-03	8.6E-01	-1.4E-03	7.6E-01	1.1E-03	4.6E-01	5.9E-03	4.4E-01	-1.7E-02	1.9E-01
Trunk fat mass (kg)	-1.5E-03	4.5E-01	-1.0E-03	3.4E-01	1.5E-04	6.5E-01	-6.5E-04	7.1E-01	-2.0E-03	5.0E-01
Trunk lean mass (kg)	-2.5E-04	9.3E-01	4.4E-04	7.8E-01	2.6E-04	5.9E-01	1.0E-03	6.8E-01	-6.1E-03	1.5E-01
Obese (BMI>=30 kg/m^2, N=45)										
Total fat mass (kg)	-5.7E-03	3.4E-01	-5.3E-04	6.0E-01	3.5E-04	4.5E-01	-6.2E-04	7.9E-01	2.3E-03	4.7E-01
Total lean mass (kg)	-1.8E-02	1.9E-01	-2.8E-03	2.4E-01	1.4E-03	1.9E-01	-6.2E-03	2.4E-01	-1.1E-03	8.9E-01
Android fat mass (kg)	-5.1E-02	5.6E-01	7.3E-04	9.6E-01	4.5E-03	4.9E-01	-7.4E-03	8.2E-01	4.1E-02	3.4E-01
Android lean mass (kg)	-1.8E-01	1.4E-01	-2.9E-02	1.4E-01	1.2E-02	2.0E-01	-5.6E-02	2.3E-01	-3.1E-02	6.5E-01
Gynoid fat mass (kg)	-3.5E-02	2.5E-01	-4.6E-03	3.7E-01	2.2E-03	3.4E-01	-1.0E-03	9.3E-01	3.9E-03	8.2E-01
Gynoid lean mass (kg)	-7.5E-02	3.1E-01	-1.7E-02	1.4E-01	4.6E-03	4.1E-01	-1.7E-02	5.5E-01	-3.2E-02	4.1E-01
Trunk fat mass (kg)	-1.0E-02	3.9E-01	-1.7E-03	3.8E-01	4.6E-04	6.0E-01	-2.9E-03	5.0E-01	1.7E-03	7.8E-01
Trunk lean mass (kg)	-2.5E-02	4.4E-01	-6.2E-03	2.1E-01	1.6E-03	4.9E-01	-1.4E-02	2.3E-01	-7.4E-03	6.5E-01

Supplementary Table 16. Association of DNA methylation with incident T2D amongst the 2,664 Indian Asians in the epigenome-wide association study, after adjustment for measures of adiposity, insulin action and concentrations of amino acids (isoleucine, tyrosine and phenylalanine). Results are presented as relative risk of T2D per 1% change in methylation. Results reaching $P<10^{-7}$ are highlighted.

	Relative risk of T2D	P
Body mass index and waist-hip ratio		
TXNIP	0.93 (0.91 - 0.95)	8.6×10^{-12}
SREBF1	1.08 (1.04 - 1.12)	2.3×10^{-5}
PHOSPHO1	0.86 (0.81 - 0.92)	4.2×10^{-6}
SOC3	0.96 (0.94 - 0.98)	5.1×10^{-4}
ABCG1	1.12 (1.08 - 1.16)	4.7×10^{-9}
HOMA-IR		
TXNIP	0.93 (0.91 - 0.95)	1.8×10^{-10}
SREBF1	1.07 (1.03 - 1.11)	5.5×10^{-4}
PHOSPHO1	0.86 (0.81 - 0.92)	1.5×10^{-5}
SOC3	0.96 (0.93 - 0.98)	2.8×10^{-4}
ABCG1	1.09 (1.05 - 1.14)	7.7×10^{-6}
HOMA-B		
TXNIP	0.92 (0.90 - 0.94)	2.7×10^{-14}
SREBF1	1.11 (1.07 - 1.15)	4.2×10^{-8}
PHOSPHO1	0.83 (0.78 - 0.89)	1.6×10^{-8}
SOC3	0.94 (0.92 - 0.96)	9.0×10^{-7}
ABCG1	1.14 (1.10 - 1.18)	9.6×10^{-12}
Amino acids		
TXNIP	0.93 (0.91 - 0.95)	7.7×10^{-9}
SREBF1	1.08 (1.04 - 1.13)	1.9×10^{-4}
PHOSPHO1	0.85 (0.79 - 0.92)	2.8×10^{-5}
SOC3	0.95 (0.92 - 0.97)	1.1×10^{-4}
ABCG1	1.11 (1.07 - 1.16)	1.1×10^{-6}

Supplementary Table 17. Associations with T2D amongst Indian Asians in the epigenome-wide association study. Effects are per 1% increase in respective methylation markers and are adjusted for age and sex, in multivariable analysis including all 5 confirmed methylation markers simultaneously.

	Relative risk of T2D (95% CI)	**P**
TXNIP	0.94 (0.92-0.95)	3.8×10^{-12}
SREBF1	1.04 (1.01-1.07)	3.5×10^{-3}
PHOSPHO1	0.88 (0.84-0.93)	2.9×10^{-6}
SOCS3	0.96 (0.94-0.98)	5.1×10^{-5}
ABCG1	1.08 (1.04-1.11)	1.2×10^{-5}

Supplementary Table 18. Association of Methylation score with incident T2D amongst Indian Asians in the Epigenome-wide association study. Results are presented as risk (95% confidence interval) per 1 SD increased in Methylation score, and as risk between the highest and lowest quartiles of Methylation score. All associations are adjusted for age and sex. Amino acids: isoleucine, phenylalanine, tyrosine).

Adjustment	Relative Risk per SD	P	Relative Risk Q4 vs Q1	P
All participants				
None	1.68 (1.55-1.83)	$1.1x10^{-33}$	3.51 (2.79-4.42)	$1.3x10^{-26}$
Physical activity	1.68 (1.54-1.82)	$8.4x10^{-33}$	3.45 (2.74-4.35)	$1.0x10^{-25}$
Family history of T2D	1.68 (1.54-1.83)	$4.4x10^{-33}$	3.48 (2.77-4.39)	$3.4x10^{-26}$
Body mass index & waist hip ratio	1.52 (1.39-1.66)	$1.3x10^{-20}$	2.67 (2.09-3.40)	$2.5x10^{-15}$
HOMA-IR	1.48 (1.35-1.62)	$3.7x10^{-17}$	2.56 (1.99-3.29)	$2.3x10^{-13}$
Glucose	1.52 (1.39-1.67)	$6.3x10^{-20}$	2.70 (2.11-3.46)	$2.3x10^{-15}$
HbA1c	1.62 (1.48-1.78)	$2.6x10^{-25}$	3.27 (2.55-4.21)	$1.7x10^{-20}$
Amino Acids	1.54 (1.39-1.70)	$2.0x10^{-17}$	2.71 (2.06-3.56)	$8.0x10^{-13}$
Family history of T2D, body mass index, waist-hip ratio, HbA1c, glucose	1.41 (1.28-1.55)	$8.5x10^{-12}$	2.25 (1.72-2.97)	$5.4x10^{-9}$

Supplementary Table 19. Characteristics of the 1,932 Indian Asian incident T2D cases and controls in the epigenome-wide association study without pre-diabetes at baseline (fasting glucose<6mmo/l and HbA1c<6%.). Results are presented as mean (SD) or as % (N).

	Incident T2D	**Controls**	**P**
N	1,412	520	
Age (yrs)	49.7 (9.8)	52.2 (10.1)	<0.0001
Sex (M)	67.9% (959)	74.4% (387)	0.13
Impaired fasting glucose	0% (0)	0% (0)	<0.0001
Fasting glucose (mmol/L)	5.01 (0.41)	5.29 (0.43)	<0.0001
HbA1c (%)	5.27 (0.85)	5.74 (0.48)	<0.0001
Insulin (IU/L)	10.2 (9.1)	14.0 (10.5)	<0.0001
HOMA-IR	2.3 (1.9)	3.3 (2.6)	<0.0001
HOMA-B	144 (210)	162 (122)	0.02
Body mass index (kg/m^2)	26.7 (3.9)	28.9 (4.8)	<0.0001
Waist circumference (cm)	94.6 (11.1)	100.7 (12.2)	<0.0001
Waist-hip ratio	0.93 (0.08)	0.97 (0.08)	<0.0001
Systolic BP (mmHg)	129.2 (18.6)	133.4 (18.5)	<0.0001
Diastolic BP (mmHg)	80.9 (10.4)	82.3 (10.6)	0.001
Cholesterol (mmol/L)	5.42 (0.98)	5.29 (1.08)	0.002
Triglycerides (mmol/L)	1.61 (0.92)	1.86 (1.32)	<0.0001
HDL cholesterol (mmol/L)	1.30 (0.29)	1.21 (0.26)	<0.0001
Smoking			
Never smoked	84.5% (1,193)	86.7% (399)	0.12
Ex-smoker	7.2% (102)	11.9% (62)	
Current smoker	8.3% (117)	11.3% (59)	
Measured white cell subsets			
Lymphocytes (%)	34.4 (7.8)	34.3 (7.9)	0.74
Monocytes (%)	6.2 (1.9)	6.2 (2.0)	0.92
Neutrophils (%)	54.3 (8.5)	54.2 (8.5)	0.83
Eosinophls (%)	3.7 (2.4)	3.9 (2.7)	0.11
Basophils (%)	0.9 (0.5)	1.0 (0.6)	0.19
Estimated white cell subsets			
CD8T (%)	17.0 (5.9)	17.2 (6.1)	0.40
CD4T (%)	15.4 (4.9)	15.3 (5.1)	0.64
Natural Killer cells (%)	0.6 (2.2)	0.2 (2.1)	<0.0001
B-cells (%)	6.3 (2.3)	6.4 (2.3)	0.06
Monocytes (%)	10.3 (1.9)	10.5 (1.8)	0.03
Granulocytes (%)	50.8 (7.5)	50.9 (6.9)	0.88

Supplementary Table 20. Association of DNA methylation with incident T2D amongst the 1,932 Indian Asians without pre-diabetes (fasting glucose<6mmo/l and HbA1c<6%) in the epigenome-wide association sample. Results are presented as risk (95% confidence interval) per 1 % increase in Methylation or per 1SD increase in methylation score, adjusted for age and sex.

Marker	Relative Risk of T2D (95%CI)	P
TXNIP	0.94 (0.92-0.97)	7.0×10^{-7}
SREBF1	1.07 (1.03-1.11)	7.4×10^{-5}
PHOSPHO1	0.89 (0.84-0.95)	3.8×10^{-4}
SOCS3	0.95 (0.93-0.97)	8.1×10^{-5}
ABCG1	1.11 (1.06-1.15)	9.4×10^{-7}
Methylation score	1.65 (1.48-1.84)	2.3×10^{-19}

Supplementary Table 21. DNA methylation levels amongst a representative sample of 186 Indian Asian and 192 European participants of the LOLIPOP study, without T2D.

	Europeans		Indian Asians		
	Mean	**SD**	**Mean**	**SD**	**P**
Age (years)	51.3	6.4	52.4	1.3	0.02
Body mass index (kg/m^2)	27.5	4.15	26.8	3.8	0.10
Waist-hip ratio	0.93	0.06	0.96	0.07	<0.0001
Systolic BP (mmHg)	134.3	17.94	136.9	22.0	0.20
Diastolic BP (mmHg)	83.2	10.95	86.5	13.0	0.01
Glucose (mmol/L)	5.11	0.51	5.25	0.54	0.02
Insulin (IU/L)	9.31	6.06	11.8	6.9	<0.0001
HbA1c (%)	5.27	0.44	5.57	0.56	<0.0001
Cholesterol (mmol/L)	5.59	1.05	5.48	0.95	0.30
HDL cholesterol (mmol/L)	1.36	0.32	1.27	0.37	0.01
Triglycerides (mmol/L)	1.61	1.14	1.63	0.73	0.87
TXNIP (%)	70.4	4.8	71.0	5.9	0.32
SREBF1 (%)	48.4	7.7	52.0	5.3	<0.0001
PHOSPHO1 (%)	4.4	1.9	4.1	1.2	0.03
SOC3 (%)	49.5	7.3	46.1	5.6	<0.0001
ABCG1 (%)	55.1	8.9	66.7	4.2	<0.0001

Supplementary Table 22. Pearson correlation co-efficient between DNA methylation in blood and liver in humans (N=175)

Marker	Locus	Correlation co-efficient	P
cg19693031	*TXNIP*	0.18	0.02
cg11024682	*SREBF1*	0.00	1.00
cg02650017	*PHOSPHO1*	0.03	0.68
cg18181703	*SOCS3*	0.31	5.3×10^{-5}
cg06500161	*ABCG1*	-0.02	0.75

Supplementary Table 23. Association of DNA methylation with gene expression in peripheral blood amongst Indian Asians and Europeans. Methylation markers associated with gene expression at P<0.05 after Bonferroni correction for the number of probes for the respective gene on the expression array are highlighted.

		Indian Asians LOLIPOP (n=907)			Europeans EGM (n=591)			Europeans KORA (n=703)		
Marker	**Locus**	Probes	Beta	P	Probes	Beta	P	Probes	Beta	P
cg19693031	*TXNIP*	1	2.20	0.03	1	0.12	0.84	1	-0.03	0.34
cg11024682	*SREBF1*	3	-3.73	2.0E-04	2	-1.58	8.8E-05	3	-0.15	3.8E-03
cg02650017	*PHOSPHO1*	1	1.69	0.09	1	3.29	0.02	1	0.08	0.55
cg18181703	*SOCS3*	2	-1.55	0.12	2	2.58	1.9E-03	1	0.09	0.03
cg06500161	*ABCG1*	6	-9.73	3.8E-21	1	-2.51	1.4E-06	4	-5.20	1.5E-20

Supplementary Table 24. Association of DNA methylation with gene expression in liver (2 datasets). Methylation markers associated with gene expression at P<0.05 after Bonferroni correction for the number of probes for the respective gene on the expression array are highlighted.

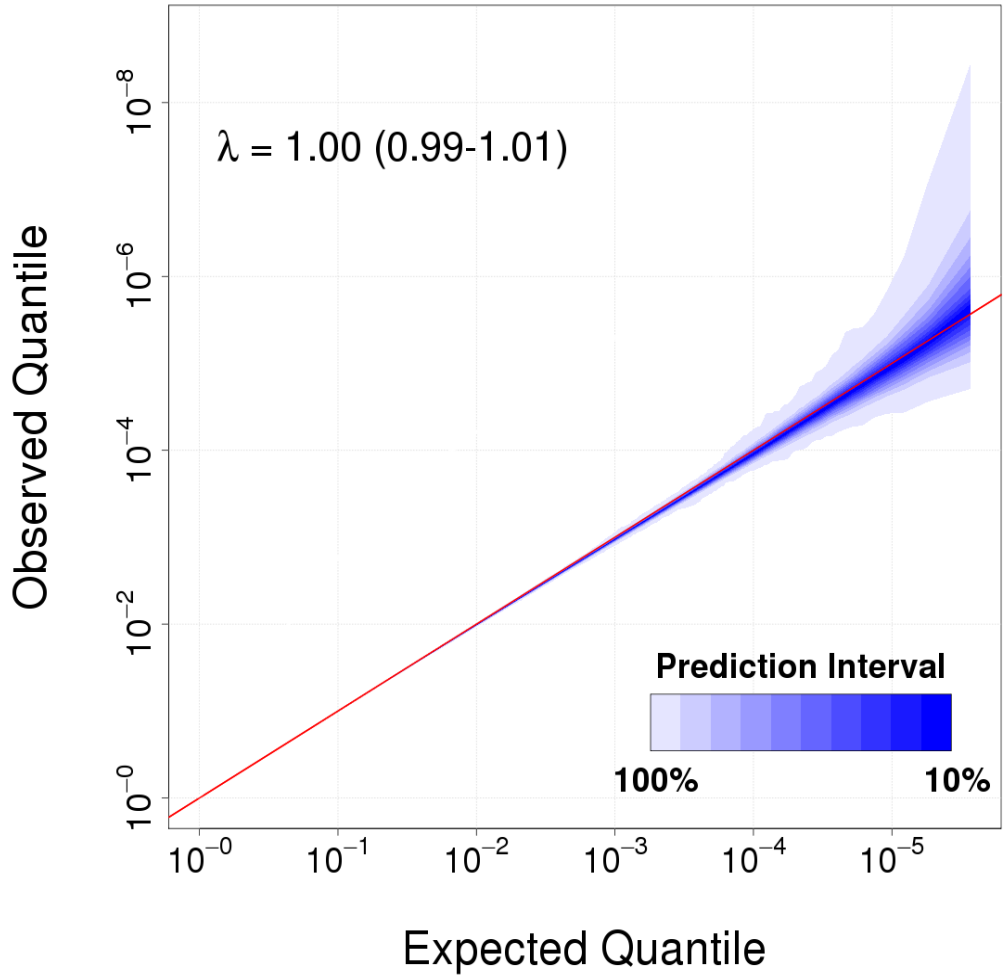
		Liver1 (n=70)			Liver2 (n=47)		
Marker	**Locus**	Probes	Beta	P	Probes	Beta	P
cg19693031	*TXNIP*	10	0.40	7.4E-04	1	0.44	3.9E-02
cg11024682	*SREBF1*	21	-0.38	4.0E-02	2	-0.96	3.9E-01
cg02650017	*PHOSPHO1*	5	0.66	7.2E-02	0	NA	NA
cg18181703	*SOCS3*	4	0.06	6.1E-01	1	-0.86	5.4E-01
cg06500161	*ABCG1*	23	0.25	4.8E-02	2	0.95	2.3E-01

Supplementary references

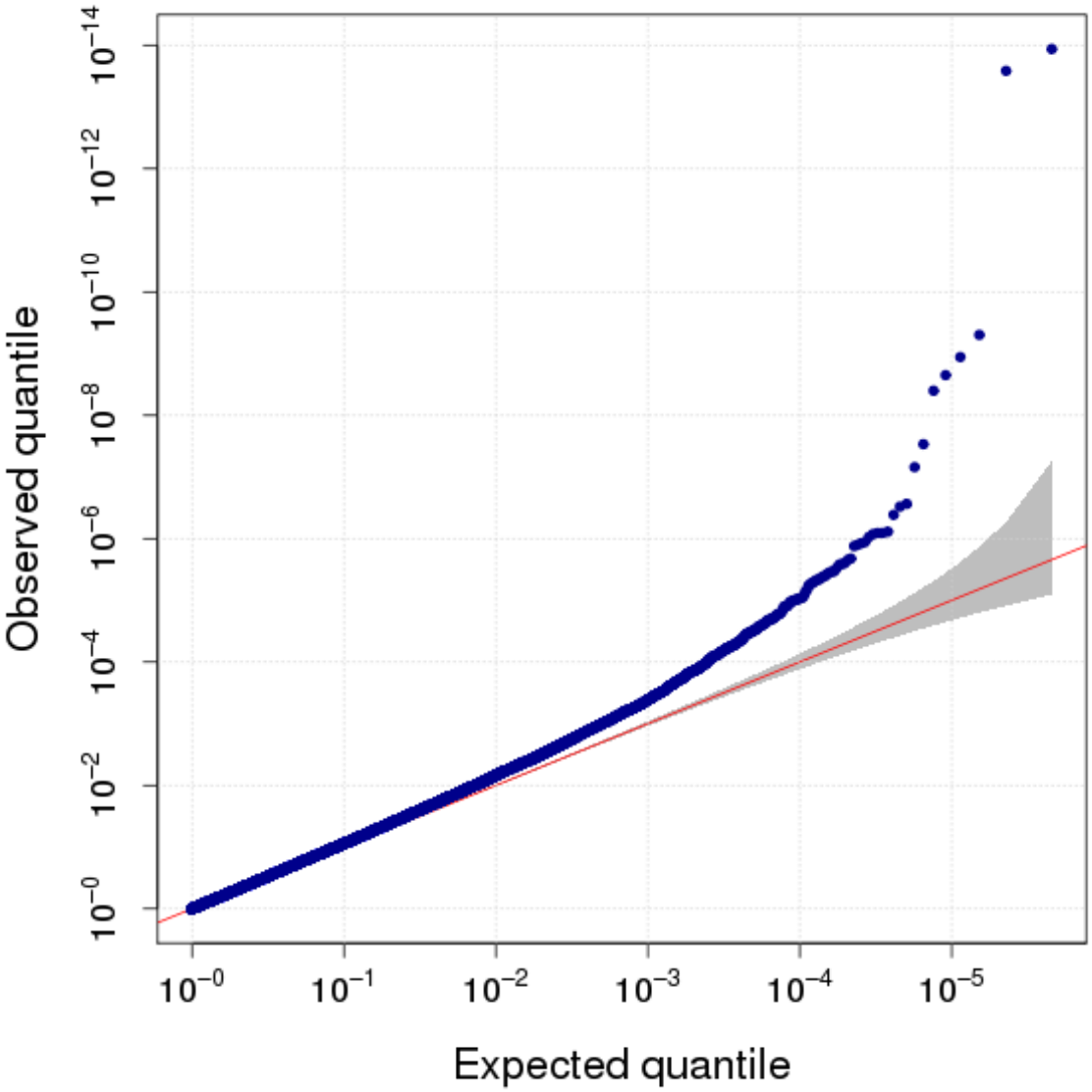
1. Chambers JC, Zhang W, Sehmi J, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature genetics* 2011; **43**(11): 1131-8.
2. Holle R, Happich M, Lowel H, Wichmann HE, Group MKS. KORA--a research platform for population based health research. *Gesundheitswesen* 2005; **67 Suppl 1**: S19-25.
3. Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *International journal of epidemiology* 2013; **42**(1): 111-27.
4. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International journal of epidemiology* 2013; **42**(1): 97-110.
5. Zeilinger S, Kuhnel B, Klopp N, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PloS one* 2013; **8**(5): e63812.
6. Touleimat N, Tost J. Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* 2012; **4**(3): 325-41.
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**(14): 1754-60.
8. Hansen K, Aryee M. Minfi: Analyze Illumina's 450k Methylation Arrays. R package version 1.2.0.
9. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics* 2012; **13**: 86.
10. Loh M, Lehne B, Drong A, et al. Importance of Batch and White Blood Cell Subtypes Correction in Analysis of Illumina Infinium 450K Methylation Arrays. American Society of Human Genetics; 2013 October 22-26, 2013; Boston, MA; 2013.
11. Drong A, Lehne B, Loh M, et al. Quality Control and Data Normalisation in large Illumina Infinium HumanMethylation450 Datasets. American Society of Human Genetics; 2013 October 22-26, 2013; Boston, MA; 2013.
12. Lehne B, Drong A, Loh M, et al. Correlation and Null Hypothesis in Epigenome-wide Association Studies (EWAS). American Society of Human Genetics; 2013 October 22-26, 2013; Boston, MA; 2013.
13. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 2008; **24**(13): 1547-8.
14. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 2013; **29**(2): 189-96.
15. Therneau T. A Package for Survival Analysis in S. R package version 2.37-4.; 2013.
16. Ritchie ME, Silver J, Oshlack A, et al. A comparison of background correction methods for two-colour microarrays. *Bioinformatics* 2007; **23**(20): 2700-7.
17. Smyth GK. Limma: linear models for microarray data. Bioinformatics and Computational Biology Solutions using R and Bioconductor. New York: Springer; 2005: 397-420.
18. Schurmann C, Heim K, Schillert A, et al. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. *PloS one* 2012; **7**(12): e50938.
19. Kelly RS, Lundh T, Porta M, et al. Blood erythrocyte concentrations of cadmium and lead and the risk of B-cell non-Hodgkin's lymphoma and multiple myeloma: a nested case-control study. *PloS one* 2013; **8**(11): e81892.

20. Ahrens M, Ammerpohl O, von Schonfels W, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell metabolism* 2013; **18**(2): 296-302.
21. Lien F, Berthier A, Bouchaert E, et al. Metformin interferes with bile acid homeostasis through AMPK-FXR crosstalk. *The Journal of clinical investigation* 2014; **124**(3): 1037-51.
22. Kruit JK, Wijesekara N, Westwell-Roper C, et al. Loss of both ABCA1 and ABCG1 results in increased disturbances in islet sterol homeostasis, inflammation, and impaired beta-cell function. *Diabetes* 2012; **61**(3): 659-64.
23. Sturek JM, Castle JD, Trace AP, et al. An intracellular role for ABCG1-mediated cholesterol transport in the regulated secretory pathway of mouse pancreatic beta cells. *The Journal of clinical investigation* 2010; **120**(7): 2575-89.
24. Huesa C, Yadav MC, Finnila MA, et al. PHOSPHO1 is essential for mechanically competent mineralization and the avoidance of spontaneous fractures. *Bone* 2011; **48**(5): 1066-74.
25. Jiang C, Kim JH, Li F, et al. Hypoxia-inducible factor 1alpha regulates a SOCS3-STAT3-adiponectin signal transduction pathway in adipocytes. *The Journal of biological chemistry* 2013; **288**(6): 3844-57.
26. Yang Z, Hulver M, McMillan RP, et al. Regulation of insulin and leptin signaling by muscle suppressor of cytokine signaling 3 (SOCS3). *PloS one* 2012; **7**(10): e47493.
27. Jorgensen SB, O'Neill HM, Sylow L, et al. Deletion of skeletal muscle SOCS3 prevents insulin resistance in obesity. *Diabetes* 2013; **62**(1): 56-64.
28. Haas JT, Miao J, Chanda D, et al. Hepatic insulin signaling is required for obesity-dependent expression of SREBP-1c mRNA but not for feeding-dependent expression. *Cell metabolism* 2012; **15**(6): 873-84.
29. Vitto MF, Luz G, Luciano TF, et al. Reversion of steatosis by SREBP-1c antisense oligonucleotide did not improve hepatic insulin action in diet-induced obesity mice. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme* 2012; **44**(12): 885-90.
30. Cha-Molstad H, Saxena G, Chen J, Shalev A. Glucose-stimulated expression of Txnip is mediated by carbohydrate response element-binding protein, p300, and histone H4 acetylation in pancreatic beta cells. *The Journal of biological chemistry* 2009; **284**(25): 16898-905.
31. Minn AH, Hafele C, Shalev A. Thioredoxin-interacting protein is stimulated by glucose through a carbohydrate response element and induces beta-cell apoptosis. *Endocrinology* 2005; **146**(5): 2397-405.
32. Hand LE, Saer BR, Hui ST, et al. Induction of the metabolic regulator Txnip in fasting-induced and natural torpor. *Endocrinology* 2013; **154**(6): 2081-91.
33. Blouet C, Liu SM, Jo YH, Chua S, Schwartz GJ. TXNIP in Agrp neurons regulates adiposity, energy expenditure, and central leptin sensitivity. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 2012; **32**(29): 9870-7.

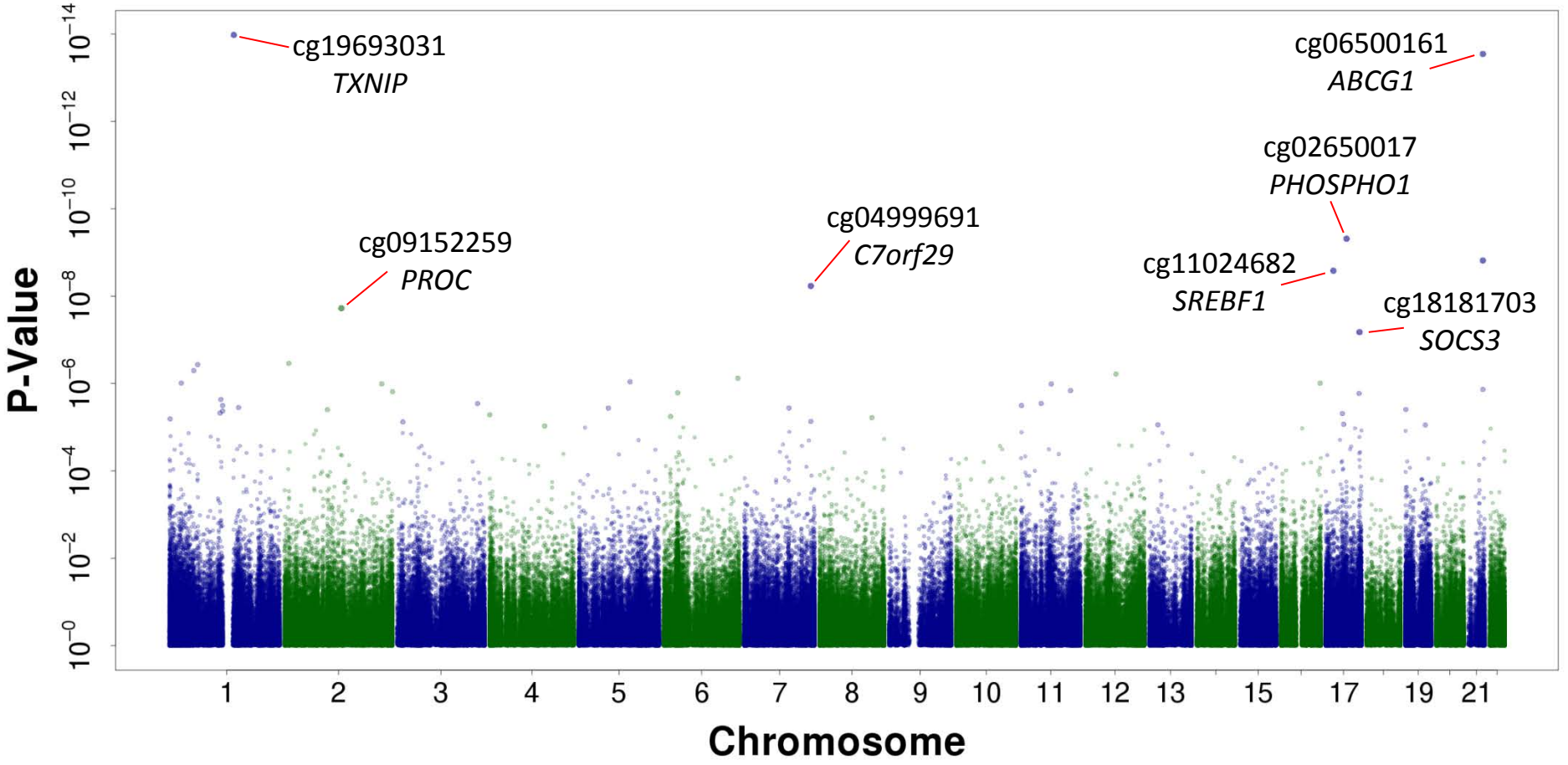
Supplementary Figure 1. Results of permutation testing (N=1000) to establish the expected distribution of P values under the null hypothesis in epigenome wide association using the Illumina 450K methylation array



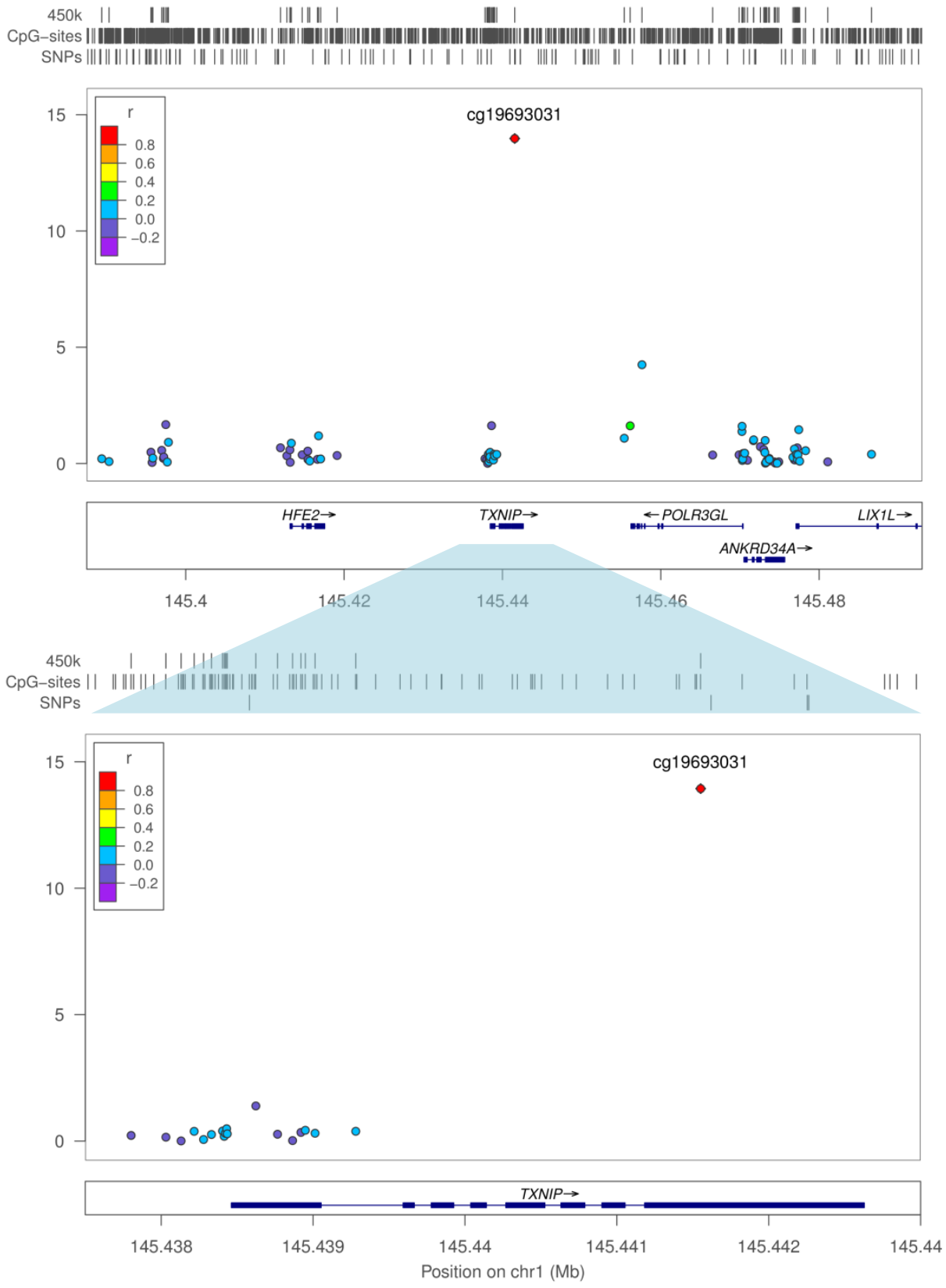
Supplementary Figure 2. QQ plot showing results for the Epigenome-wide association study. Blue dots represent results for individual CpG sites for association with T2D. The red line represents the null hypothesis of no association; the grey shading provides the 95% confidence interval for the null hypothesis.



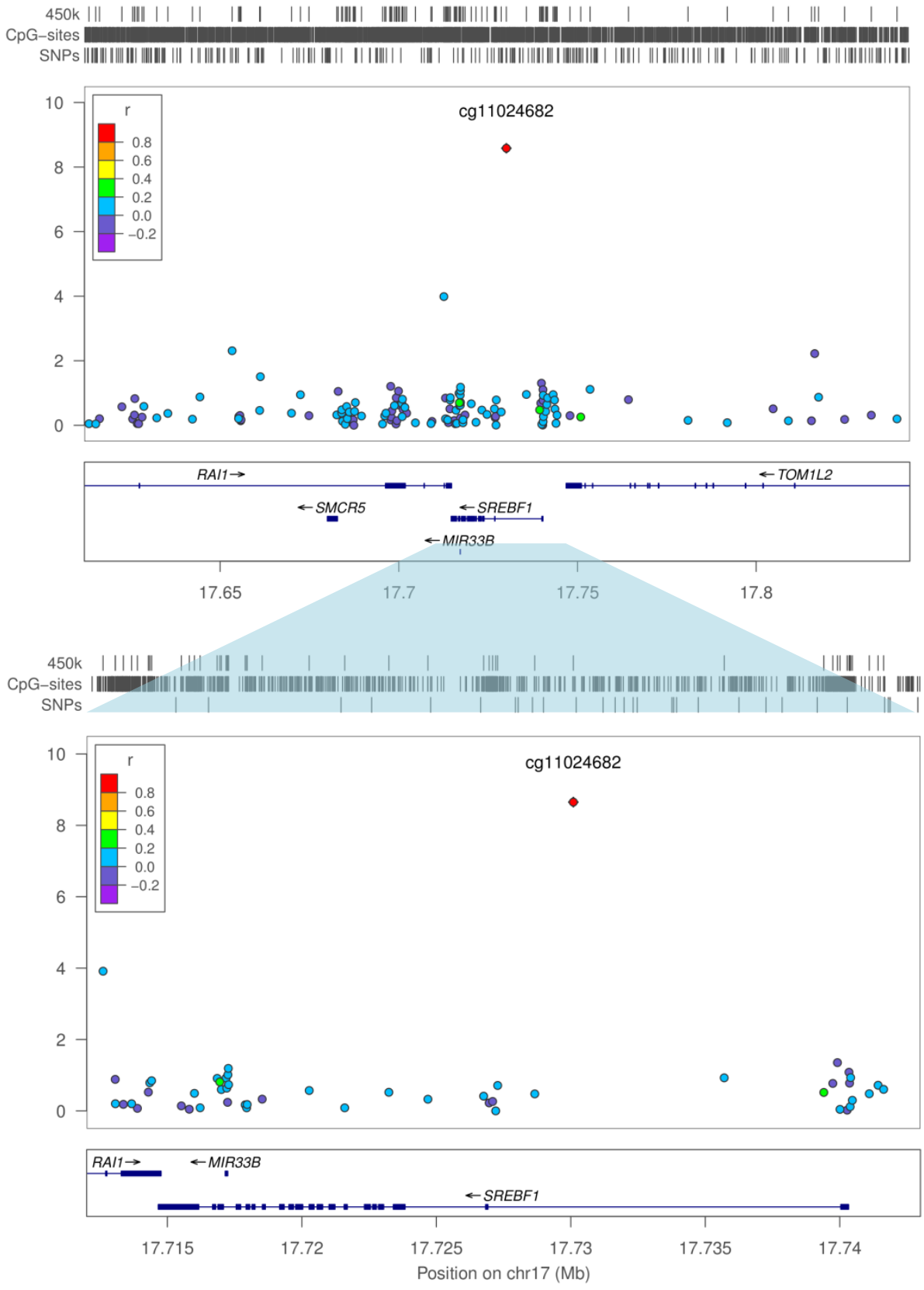
Supplementary Figure 3. Manhattan plot for the Epigenome-wide association study



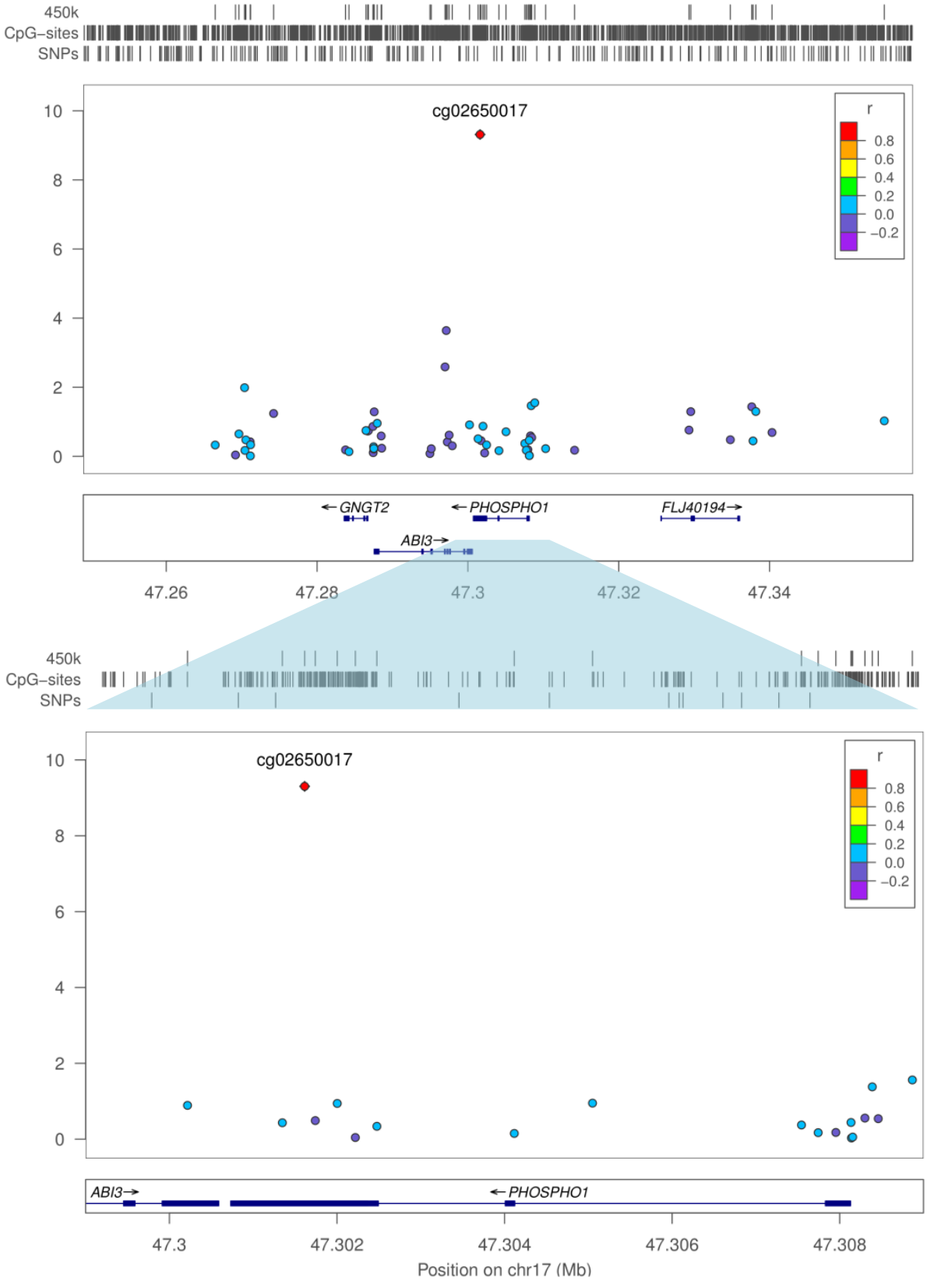
Supplementary Figure 4.1 – EWAS results at the *TXNIP* locus



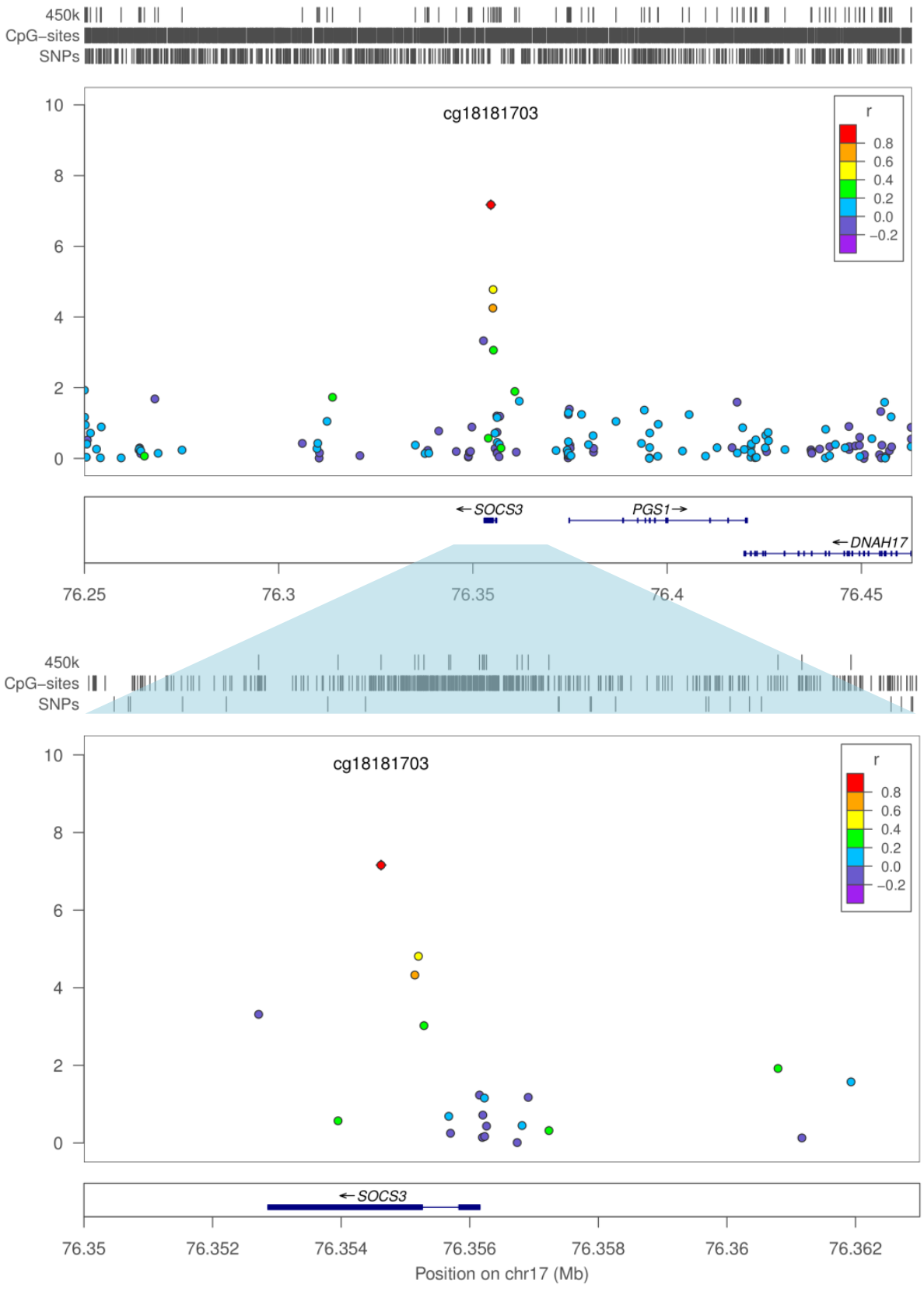
Supplementary Figure 4.2 – EWAS results at the *SREBF1* locus



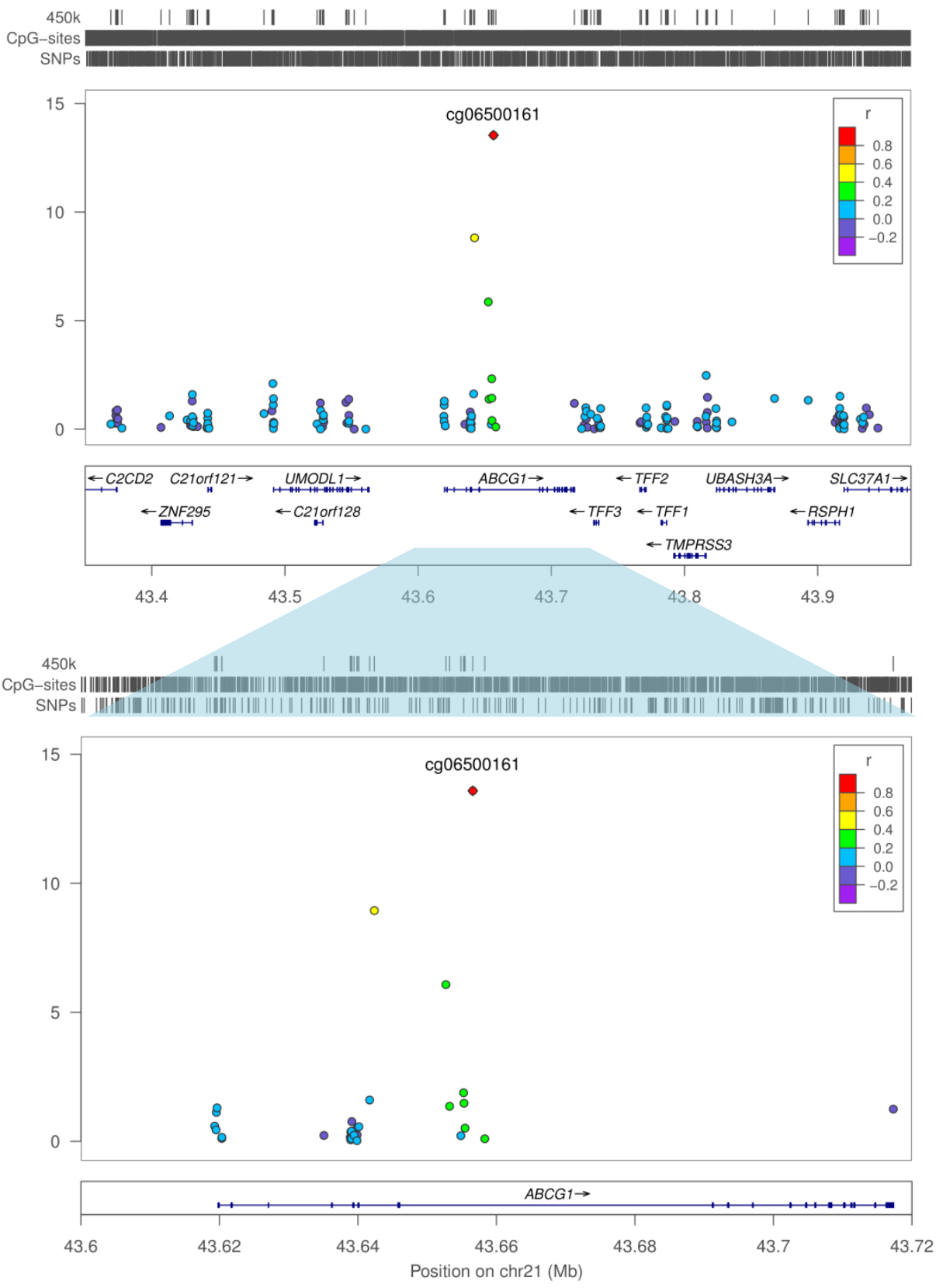
Supplementary Figure 4.3 – EWAS results at the *PHOSPHO1* locus



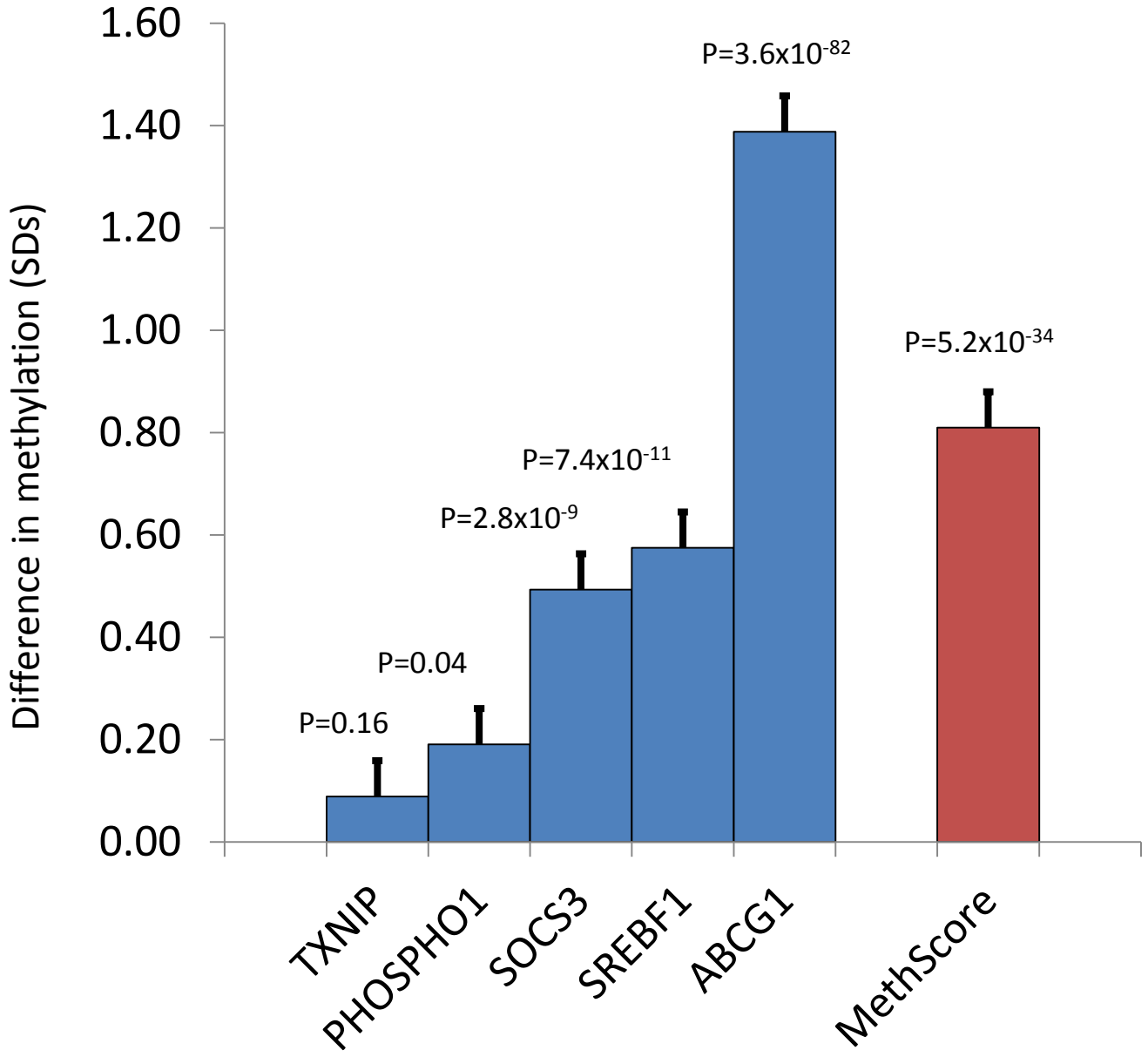
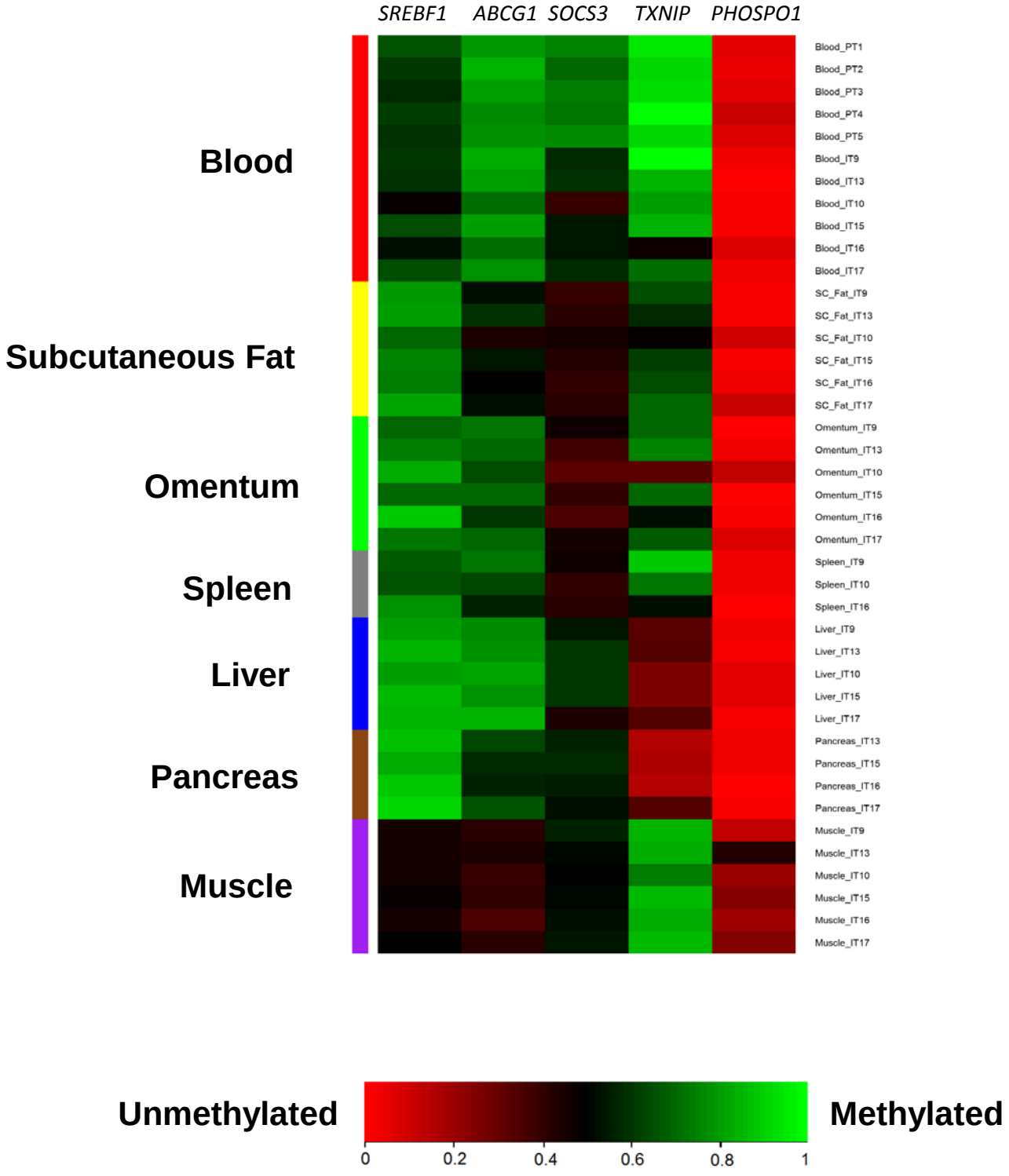
Supplementary Figure 4.4 – EWAS results at the *SOCS3* locus



Supplementary Figure 4.5 – EWAS results at the *ABCG1* locus



Supplementary Figure 5. Difference in methylation between Indian Asians and Europeans without T2D. Results presented as the SD increase in methylation amongst Indian Asians compared to Europeans, aligned to the direction of effect for increased T2D risk, adjusted for age and gender.



Supplementary Figure 6. DNA methylation levels at the 5 CpG sites associated with T2D in 41 samples assayed on the 450K methylation array