

Supplementary Text

Population samples: Discovery

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.¹ 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 primary analysis of the present study, DNA methylation measurements from a representative subsample of 1709 KORA F4 participants were considered, as well as from a sample of 485 smokers and never smokers from the KORA F3 examination.² For the longitudinal analysis, measurements from 1435 KORA S4 participants were also included. For the incident T2D analysis, measurements from a nested case-control study within KORA S3/S4 was used, comprising 200 subjects with newly diagnosed T2D (S3: physician diagnosis, S4: physician diagnosis validated through OGTT) and 200 controls matched for age (± 2 years), sex, cohort and observation time until diagnosis of diabetes. DNA methylation was measured with the Infinium HumanMethylation450K BeadChip® (Illumina, Inc., CA, USA). Sample preparation and measurement have been described in detail elsewhere.²

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 2008.³ 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 -80°C for extraction of genomic DNA. DNA methylation measurement was performed amongst 2,680 participants of the LOLIPOP study free from T2D (physician diagnosis or HbA1c $\geq$ 6.5%), using genomic DNA from peripheral blood collected at enrolment.⁴ The LOLIPOP study is approved by the National Research Ethics Service (07/H0712/150) and all participants gave written informed consent.

Italian cardiovascular component of the European Prospective Investigation into Cancer and Nutrition (EPICOR)

EPICOR is a nested case-cohort study within the EPIC-Italy cohort of ~50,000 participants recruited between 1994-1998.^{5,6} The enrolment assessment included a detailed dietary and lifestyle questionnaire, as well as collection of a non-fasting peripheral blood sample. The cohort is under long-term follow-up for incident disease including cancers and other non-communicable disorders. The Whole genome DNA methylation was measured in genomic DNA from peripheral blood collected at enrolment in 292 EPICOR participants who experienced non-fatal myocardial infarction during follow-up, and 292 healthy controls matched for age and gender. All participants are of European ancestry.

Population samples: Replication

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.^{7,8} 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/>). 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.

As part of the ARIES (Accessible Resource for Integrated Epigenomic Studies, <http://www.ariesepigenomics.org.uk/>) project, the Infinium HM450 BeadChip was used to quantify DNA methylation in genomic DNA from peripheral blood amongst 1,018 mother-offspring pairs from 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). Methylation samples from the latter time point are included in this analysis. The DNA methylation wet-lab and pre-processing analyses were performed at the University of Bristol.

Biobank-based Integrative Omics Studies (BIOS) Consortium

The mission of the BIOS Consortium is to create a large-scale data infrastructure and to bring together researchers focusing on integrative omics studies in Dutch Biobanks. The BIOS Consortium applies a functional genomics approach that integrates genome-wide genetic data with data on the epigenome and transcriptome to elucidate these mechanisms. Over BIOS consortium includes ~4000 samples from the Leiden Longevity Study, the LIFELINES Deep study and the Rotterdam study (RS-BIOS).

The Leiden Longevity Study (<http://www.molepi.nl>) consists of offspring of nonagenarian sibling pairs aged >90 years, and the partners of these offspring as population controls. LifeLines (<http://www.lifelines.nl>) is a population based study of 165,000 participants from the northern provinces of the Netherlands aimed at investigating the relationship between biomarkers and healthy ageing. The Rotterdam Study is a large prospective, population based cohort study in the district of Rotterdam, the Netherlands, investigating the prevalence, incidence, and risk factors of various chronic disabling diseases among elderly Caucasians aged 45 years and over.

Estonian Genome Centre University of Tartu (EGCUT) study.

EGCUT is a population based biobank of 51,530 individuals age ≥18 years in Estonia.⁹ All participants completed a computer assisted interview, including personal data (place of birth, place(s) of living, nationality etc.), genealogical data (family history, three generations), educational and occupational history and lifestyle data (physical activity, dietary habits,

smoking, alcohol consumption, quality of life). Anthropometric and physiological measurements were also recorded. The samples used in this study were selected from the EGCUT Center for Translational Genomics (CTG) cohort of individuals who have been re-contacted for a second time-point sample (N=96), and the EGCUT asthma cohort (N=173) a set of non-obese, non-smoking individuals with early onset asthma, and matched controls. The collection of blood samples and data generation are conducted according to the Estonian Human Gene Research Act and all participants have signed a broad informed consent.

The LOLIPOP Study

LOLIPOP has been described above. A further 656 Indian Asian samples representative of the LOLIPOP participants were available for replication testing. DNA methylation was quantified in genomic DNA from peripheral blood.

The Rotterdam Study (RS)

RS is a large prospective population-based cohort study including men and women of 45 years and over. The design and rationale of this study are described in detail elsewhere.¹⁰ In summary, the Rotterdam Study aims to investigate the determinants, incidence and progression of chronic disabling diseases in the elderly. The first cohort, Rotterdam Study I (RS-I) was initiated in 1989 including 7983 persons aged 55 years and older and was extended in 1999 with 3011 participants in Rotterdam Study II (RS-II). In 2005, Rotterdam Study III (RS-III) added another 3932 individuals aged 45 years and older. All participants were examined in detail at baseline and follow-up visits every ± 5 years. In summary, a home interview was conducted and the subjects underwent an extensive set of examinations at the research center, including cardiovascular and metabolic health. DNA was isolated from whole blood using standard procedures. Epigenome-wide methylation scans were carried out on the DNA collected at the baseline visit of the RS-III cohort (the RS-III dataset), or at the fifth follow-up measurement in all three Rotterdam cohorts (RS-BIOS dataset).

TwinsUK cohort (TwinsUK)

Epigenome wide association scans were performed in 355 whole blood samples and in 542 adipose tissue biopsy samples from TwinsUK subjects. The TwinsUK cohort is a nationwide registry of healthy volunteer twins in the United Kingdom, with about 13,000 registered twins since 1992, with predominately Caucasian female (84%) and equal number of monozygotic and dizygotic twins. Participants completed phenotype questionnaires, including information about their health, self and family disease history, medication use, and habitual behaviours, such as smoking and alcohol consumption. The data include collections of clinical, phenotype, and biochemical measures from biological samples often profiles at multiple time points. DNA methylation profiles were generated using the Illumina 450k array in both blood and adipose tissue samples. The subjects included in the methylation analyses were free from severe diseases, such as cancer. Whole blood aliquots were stored at -80C and DNA was extracted using standard protocols. Adipose tissue biopsies and DNA methylation profiling were performed as previously described.¹¹

LifeLines Deep

Epigenome wide association was done amongst 752 participants of the LifeLines Cohort Study (Supplementary Table 1). The LifeLines Cohort Study is a large population-based cohort study and biobank that was established as a resource for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy ageing. Between 2006 and 2013, inhabitants of the northern part of The Netherlands and their families were invited to participate, thereby contributing to a three-generation design. Participants visited one of the LifeLines research sites for a

physical examination, including lung function, ECG and cognition tests, and completed extensive questionnaires. Baseline data were collected for 167 729 participants, aged from 6 months to 93 years.¹² 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 -80°C for extraction of genomic DNA. Epigenome-wide methylation scans were carried out on the DNA collected at enrolment to the LifeLines study.

Leiden Longevity Study (Leiden Longevity)

Epigenome wide association was done amongst 642 participants of the Leiden Longevity Study (LLS) (Supplementary Table 1). LLS consists of offspring of nonagenarian sibling pairs of which the members are aged over 90 years. The partners of these offspring were recruited as population controls. The study has been designed to investigate biomarkers of healthy ageing and longevity.¹³ Samples for methylation analysis were unrelated individuals selected from the partner and offspring groups.

Replication testing in liver

The collection and analysis of liver samples has been described previously.¹⁴ In brief, 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. 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, San Diego, CA) 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. The dataset is available via GEO (GSE48325).

Acknowledgements

ALSPAC. We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. This project is supported by the UK Biotechnology and Biological Sciences Research Council (BB/I025751/1 and BB/I025263/1); the UK Medical Research Council and University of Bristol (MC_UU_12013); the Wellcome Trust (WT083431MF to RCR); the Economic and Social Research Council (RES-060-23-0011 to GDS and CLR); and the European Research Council (DEVHEALTH 269874 to GDS).

BIOS consortium. The BIOS Consortium is funded by the BBMRI-NL, a research infrastructure financed by the Netherlands Organization for Scientific Research (NWO project 184.021.007).

EGCUT. This research was supported by grants from the University of Tartu (SP1GVARENG), the Estonian Research Council (IUT20-60), the Estonian Research Roadmap through the Estonian Ministry of Education and Research, the Center of

Excellence in Genomics (EXCEGEN), and the Estonian Science Foundation (ETF9293, ETF9353).

EPICOR. EPIC and EPICOR projects are supported by the Compagnia di San Paolo (SP, VK, RT, PV, GM), by the Human Genetics Foundation (HuGeF; GM, PV), and by the MIUR ex60% grant (GM). EPIC Italy is supported by a generous grant from the Associazione Italiana per la Ricerca sul Cancro (AIRC, Milan). The authors wish to thank all who participated in, or collaborated with EPIC, in particular the AVIS blood donors organization, and the Sicilian Government. The authors are solely responsible for the publication, and the publication does not represent the opinion of the Community. The Community is not responsible for any use that might be made of data appearing in this work.

KORA. KORA was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany and supported by grants from the German Federal Ministry of Education and Research (BMBF), the Federal Ministry of Health (Berlin, Germany), the Ministry of Innovation, Science, Research and Technology of the state North Rhine-Westphalia (Düsseldorf, Germany), and the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ.

This research was supported by a Grant from the GIF, the German-Israeli Foundation for Scientific Research and Development, by the European Union's Seventh Framework Programme (FP7-Health-F5-2012) under grant agreement no. 305280 (MIMOmics), by the Helmholtz-Russia Joint Research Group (HRJRG) 310, and by the German Center for Diabetes Research (DZD). We thank all members of field staffs who were involved in the planning and conduct of the MONICA/KORA Augsburg studies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

LOLIPOP. 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) the NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143) and Action on Hearing Loss (G51). We thank the participants and research staff who made the study possible. PE is Director of the MRC-PHE Centre for Environment and Health and acknowledges support from the Medical Research Council and Public Health England. PE is a National Institute for Health Research senior investigator and acknowledges support from the NIHR Biomedical Research Centre at Imperial College Healthcare NHS Trust and Imperial College London, and the NIHR Health Protection Research Unit on Health Effects of Environmental Hazards

Rotterdam. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of the Illumina 450K methylation array data for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The methylation data were funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure

Netherlands (BBMRI-NL). We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk for their help in creating the methylation database.

TwinsUK. The study was funded by the Wellcome Trust; European Community's Seventh Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for Health Research (NIHR)- funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/CIDR. T.D.S. is a holder of a European Research Council Advanced Principal Investigator award.

LifeLines Deep. This work was supported by the European Research Council Advanced Grant (ERC-671274 to CW), the Dutch Digestive Diseases Foundation (MLDS WO11-30 to CW), the European Union's Seventh Framework Programme (EU FP7) TANDEM project (HEALTH-F3-2012-305279 to CW), the Netherlands Organization for Scientific Research (NWO-VENI grant 916-10135 to LF and NWO VIDI grant 917-14374 to LF). Generation of the methylation data (as part of the Biobank-based Integrative Omics Study (BIOS)) is financially supported by the Biobanking and Biomolecular Research Infrastructure of The Netherlands (BBMRI-NL), funded by the Netherlands Organisation for Scientific Research (NWO).

Leiden Longevity. We thank all participants of the Leiden Longevity Study. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement no. 259679. This study was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology, the Netherlands Consortium for Healthy Ageing (grant 050-060-810), and the Biobank-Based Integrative Omics Studies (BIOS) Consortium for the generation of DNA methylation data funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007), all in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO).

Liver dataset. Gene expression studies in liver were supported by the German Ministry of Education and Research (BMBF) through the Virtual Liver Project and through institutional funds from the Medical Faculty of the Technical University Dresden and the University of Kiel.

Isolated white cells. We acknowledge the support of the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) Imaging and FACS Facility, Hammersmith Campus (Imperial College Healthcare NHS Trust in partnership with Imperial College London).

Supplementary Tables

Supplementary Information Table 1. Characteristics of cohorts and participants. Results are presented as mean (SD) for continuous and % for categorical variables.

Supplementary Information Table 2. Cohort-specific methodological details for quantification and analysis of DNA methylation. PC: principal component; QN: quantile normalization; WBC: white blood cell; SWAN: subset-quantile within array normalization.

Supplementary Information Table 3. Results for the 278 CpG sites markers reaching $P < 1 \times 10^{-7}$ in the epigenome-wide association study of the relationship between DNA methylation in blood and BMI. Results are presented for the discovery and replication phases, and in combined analysis. DNA methylation is quantified on a scale of 0-1, where 1 represents 100% methylation. Sentinel marker: identifies the marker with the lowest P value at each locus. Replicate: indicates that the marker reaches $P < 0.05$ with consistent direction of effect in replication testing. Mean (SD) methylation: is average methylation (standard deviation) at the CpG site in blood. Effect (SE): is change in BMI (kg/m^2) per unit increase in methylation (ie from 0 to 1) from linear regression, and its standard error. The sentinel marker for each locus is indicated. $P_{\text{het EUR/SA}}$: P value for assessment of heterogeneity of effect between Europeans (Eur) and Indian Asians (IA). P_{het} : P value for assessment of heterogeneity of effect between the contributing cohorts. N: sample size. The 'Comments' column annotates probes that have SNPs in probe sequence, or are reported to potentially cross-hybridise.

Supplementary Information Table 4. Cohort specific results for the 207 sentinel CpG sites carried forward from the discovery phase into replication testing. DNA methylation is quantified on a scale of 0-1, where 1 represents 100% methylation. Mean (SD) methylation: is mean and standard deviation for methylation at the CpG site in blood. Effect (SE): is change in BMI (kg/m^2) per unit increase in methylation (ie from 0 to 1) from linear regression, and its standard error. N: sample size. P: cohort specific P value for association of BMI with methylation.

Supplementary Information Table 5. Association between methylation in blood and BMI at CpGs with SNPs in probe sequence. Results are shown for the association of CpG with BMI before and after adjustment for genotype of the SNP in the probe sequence ($P_{\text{CpG vs BMI}}$ and $P_{\text{CpG vs BMI adjusted}}$ respectively). In addition the association of the SNP under the probe with BMI and with methylation of the CpG are provided ($P_{\text{SNP vs BMI}}$ and $P_{\text{SNP vs CpG}}$ respectively). There is no evidence for genetic confounding by SNPs located in Infinium 450k probe sequences.

Supplementary Information Table 6. Association of DNA methylation in blood with BMI at 4 representative CpG sites assayed by pyrosequencing in 990 European and 1,720 Indian Asian participants of the LOLIPOP study, and in combined analysis of the data from the 2 ethnic groups. This was carried out as a technical validation experiment. Effect (SE): is change in BMI (kg/m^2) per unit increase in methylation (ie from 0 to 1) from linear regression, and its standard error. P: is the P value for association between BMI and methylation, from linear regression. Results confirm the association of methylation with BMI at these 4 loci using an alternate method for quantification of methylation.

Supplementary Information Table 7. Characteristics of the 30 obese cases and 30 normal weight controls participating in the study of DNA methylation in isolated white cell subsets.

Results are presented as mean (SD) for continuous and % for categorical variables. P values are for the comparison between obese cases and normal weight controls.

Supplementary Information Table 8. Association of obesity with DNA methylation at the 179 sentinel CpG sites assayed in isolated white cell subsets. Results are provided for sentinel markers in whole blood (from the discovery epigenome-wide association study), and for the 4 isolated white cell subsets (monocytes, neutrophils, CD4+ and CD8+). Mean (SD) methylation: is mean and standard deviation for methylation at the CpG site. Direction: the direction for association between methylation and BMI in the epigenome-wide association of whole blood. P_{BMI}: P value for association between methylation and BMI in epigenome-wide association of whole blood. Difference: the difference in methylation between obese cases and normal weight controls in the respective white cell subset, ([case]-[control]), quantified on a scale of 0-1, where 1 represents 100% methylation). P_{obesity}: P value for association between methylation and obesity case-control status in the respective white cell subset. Consistent: directional consistency between the association of methylation with obesity in isolated white blood cells and the association of methylation with BMI in whole blood ('+' indicates a consistent direction of effect, '-' indicates opposite direction of effect).

Supplementary Information Table 9. Association of the sentinel methylation markers with adiposity in isolated white cell subsets from 30 obese cases and 30 normal weight controls. Results are available for 179 of the 187 sentinel markers and are presented as i. number of loci associated with obesity at nominal significance ($P < 0.05$); ii. number of loci associated with obesity at $P < 2.8 \times 10^{-4}$ (ie $P < 0.05$ after Bonferroni correction for 179 tests), and iii. number of loci showing an association with obesity that is directionally consistent with the association of BMI with methylation in the discovery epigenome-wide experiment. Enrichment P: statistical test for enrichment in directional consistency (Binomial 'sign' test).

Supplementary Information Table 10. DNA methylation in subcutaneous adipose tissue (N=543 samples), and separately in liver tissue (N=55 samples), at the 187 sentinel CpG sites. Mean and standard deviation (SD) methylation levels in adipose tissue are provided for the CpG sites; DNA methylation is quantified on a scale of 0-1, where 1 represents 100% methylation. The association of methylation with BMI are provided separately for adipose and liver tissue: effect is change in BMI (kg/m^2) per unit increase in methylation (ie from 0 to 1) from linear regression, and its standard error (SE); the column 'consistent' compares direction of effect for the association of methylation with BMI in adipose or liver tissue with blood (discovery phase). Adipose-blood correlation: correlation between methylation levels in adipose tissue and blood in paired samples (N=201). NA: not available. The relationships of methylation with BMI at the sentinel CpG sites in blood are shown for comparison (from discovery EWAS, Supplementary Information Table 3).

Supplementary Information Table 11. DNA methylation in isolated adipocytes and association with adiposity. Mean and standard deviation (SD) methylation levels in isolated adipocytes are provided for blood and adipocytes (obese cases and normal weight controls). DNA methylation is quantified on a scale of 0-1, where 1 represents 100% methylation. Difference: the difference in methylation between obese cases and normal weight controls ([Obese cases]-[Normal weight controls]). P_{case-control} is for the comparison of methylation in obese cases with normal weight controls. The 6 markers reaching $P < 2.7 \times 10^{-4}$ (ie $P < 0.05$ after Bonferroni correction for 187 tests) for association with obesity in adipocytes are highlighted. Association results for markers in blood (discovery phase; association with BMI from Supplementary Information Table 3) are provided for comparison. Consistent: consistent direction of effect for association with adiposity in blood and isolated adipocytes.

Supplementary Information Table 12. Causality analysis to investigate whether DNA methylation in blood plays a causal role underlying obesity. Effect_{SNP-CpG}: is change in methylation per unit copy of effect allele (A2, from Supplementary Information Table 12), from linear regression. SE_{SNP-CpG}: standard error of the effect. P_{SNP-CpG}: is for the association of SNP with methylation. P_{SNP-CpG (adj)}: is for the association of SNP with methylation after Bonferroni correction for the number of SNPs in cis (within 1MB). Effect_{CpG-BMI}: is change in BMI per unit increase in methylation (0-1 scale where 1 represent 100%, from discovery epigenome-wide association study [Supplementary Information Table 3]). SE_{CpG-BMI}: standard error for the association of methylation on BMI. P_{CpG-BMI}: is P value for the association of methylation with BMI.

Predicted effect of SNP on BMI via methylation calculated from the product of SNP-CpG and CpG-BMI. Observed effect of SNP on BMI is from the GIANT study. The single locus (*NFATC21P*) where the observed effect of SNP on BMI reaches $P < 2.7 \times 10^{-4}$ (ie $P < 0.05$ after Bonferroni correction for 187 tests) is highlighted in bold.

Supplementary Information Table 13. Association of methylation in blood and BMI in cross-sectional and longitudinal studies as described below. In all analyses DNA methylation is quantified on a scale of 0-1, where 1 represents 100% methylation. 'Concordant direction' is in relation to cross-sectional results from the discovery EWAS.

- Cross-sectional (BMI-Meth) analysis: association of methylation with BMI in the discovery epigenome-wide association study (from Supplementary Information Table 3). Effect (SE)_{Meth-BMI}: is change in BMI (kg/m²) per unit increase in methylation (ie from 0 to 1) and its standard error, from linear regression.
- Longitudinal (Meth-ΔBMI) analysis: association of methylation at baseline with subsequent change in BMI during follow-up, amongst 1,435 Europeans and 1,513 Indian Asians. Effect (SE)_{Meth-ΔBMI}: is longitudinal change in BMI (kg/m²) between baseline and follow-up, per unit increase in methylation (ie from 0 to 1) at baseline, and its standard error, from linear regression.
- Longitudinal (ΔMeth-ΔBMI) analysis: association of change in methylation between baseline and follow-up with change in BMI between baseline and follow-up amongst 1,435 Europeans. Effect (SE)_{ΔMeth-ΔBMI}: is longitudinal change in BMI (kg/m²) between baseline and follow-up, per unit increase in methylation (ie from 0 to 1) between baseline and follow-up, and its standard error, from linear regression.

Results show that the change in BMI associated with change in methylation during follow-up is closely correlated with the cross-sectional relationship of BMI with methylation ($r=0.81$, $P < 0.001$)

Supplementary Information Table 14. Causality analysis to investigate whether DNA methylation in blood is the consequence of obesity. GRS: Genetic Risk Score comprising sum of risk alleles for SNPs associated with BMI in the GIANT study, weighted by the per allele effect size of SNP on BMI. To make results comparable to the GIANT study, BMI is inverse normalised for this specific analysis. The association of GRS with BMI (outcome variable) is provided as Effect_{GRS-BMI}, the change in BMI (inverse normalised) per unit change in GRS, with standard error (SE_{GRS-BMI}) and P value (P_{GRS-BMI}). The association of BMI (inverse normalised) with methylation (outcome variable) is provided as Effect_{BMI-CpG}, the change in methylation (0-1 scale where 1 represent 100%) per unit change in BMI (inverse normalised), along with standard error (SE_{BMI-CpG}) and P value (P_{BMI-CpG}). Predicted effects of GRS on methylation via BMI calculated from the product of GRS-BMI and BMI-CpG associations. Observed effect of GRS on methylation (outcome variable) directly quantified amongst participants of LOLIPOP and KORA studies as described in Online Methods. The

three loci where Observed effect of GRS on methylation reaches $P < 2.7 \times 10^{-4}$ (ie $P < 0.05$ after Bonferroni correction for 187 tests) are highlighted in bold.

Supplementary Information Table 15. Characteristics of individuals included in the gene expression studies.

Supplementary Information Table 16. Association of gene expression with DNA methylation in blood. Results are presented for all genes identified to have expression QTLs associated with DNA methylation at $P < 0.05$ after correction for multiple testing ($N = 5,551$ tests), within 500kb of the 187 sentinel CpG sites. Distance: genomic distance between CpG and expressed gene. Mean (SD) expression: is for mean level of transcript in the samples. Effect_{CpG-Tx} and SE_{CpG-Tx}: change in log2-transformed gene expression level per unit change (0 to 100%) in methylation, and standard error, from linear regression. P_{CpG-Tx}: is P value for association of methylation with transcript in respective population sample or in meta-analysis. P_{het}: P value for assessment of heterogeneity of effect between the contributing cohorts. Association of gene expression with BMI in blood is also provided for comparison, where Effect_{BMI-Tx} and SE_{BMI-Tx} are: change in log2-transformed gene expression level per 1kg/m² increase in methylation, and standard error, from linear regression

Supplementary Information Table 17. Association of gene expression with DNA methylation in blood, for genes that are i. nearest, or ii. annotated by Illumina as the target gene, for the 187 sentinel CpG sites. Results are shown for genes reaching $P < 0.05$ after Bonferroni correction for the number of CpG-transcript pairs tested (500kb gene: $P < 9.0 \times 10^{-6}$; nearest gene: $P < 2.0 \times 10^{-4}$; Illumina gene: $P < 2.2 \times 10^{-4}$). The lower threshold required for statistical significance is lower in the 500kb interval approach as a result of the greater burden of multiple testing compared to 'nearest' or 'Illumina' strategies. Mean (SD) is for level of transcript. Effect_{CpG-Tx} and SE_{CpG-Tx}: change in log2-transformed gene expression level per unit change (0 to 100%) in methylation, and standard error, from linear regression. P_{CpG-Tx}: is P value for association of methylation with transcript. The column labelled '500kb gene?' indicates whether the gene listed is identified under the primary strategy for assessment of the relationship between methylation and gene expression over a 500kb genomic interval. The columns labelled 'Nearest gene?' and 'Illumina gene?' indicate whether the gene listed is nearest to the CpG, or annotated to the CpG site in the Illumina manifest, respectively. There are 6 genes identified as having an association of gene expression with methylation by the 'nearest' or 'Illumina' annotation strategy, which are not identified by the more conservative 500kb interval approach.

Supplementary Information Table 18. Association of gene expression with DNA methylation in blood, subcutaneous adipose tissue and liver tissue. Results are shown for the 38 genes (column 'Transcript gene') which show an association of expression with DNA methylation in blood. Mean (SD) expression: is for mean level of transcript. Effect_{CpG-Tx} and SE_{CpG-Tx}: change in log2-transformed gene expression level per unit change (0 to 100%) in methylation, and standard error, from linear regression. P_{CpG-Tx}: is P value for association of methylation with transcript.

Supplementary Information Table 19. Candidate genes at the 187 methylation loci associated with adiposity. Results are shown for i. nearest gene: gene nearest to the sentinel methylation marker; ii. Illumina gene: annotated by Illumina as the target gene; iii. eQTL of nearest gene (from Supplementary Information Table 20); iv. eQTL of Illumina gene (from Supplementary Information Table 20); v. eQTL of any gene within 500kb of sentinel CpG site (from Supplementary Information Table 19).

Supplementary Information Table 20. Pubmed summary for the candidate genes highlighted by the 187 sentinel methylation markers.

Supplementary Information Table 21. Pathway analysis of candidate genes. Results are shown for the primary candidate gene criteria: i. gene nearest to CpG site, or ii. gene within 500kb and showing association between methylation and expression. Column names indicate data base origin and name of enriched pathways. P value: is enrichment P value based on permutation testing (10,000 sets of 187 matched CpGs, see Online Methods). FDR P value is P value corrected for multiple testing using the Benjamini-Hochberg method (to control the false discovery rate [FDR] at 5 %). Rows correspond to genes belonging to the respective pathways.

Supplementary Information Table 22. Association of DNA methylation in blood with metabolic traits linked to adiposity amongst participants of the LOLIPOP and KORA studies. Effect: change in the respective trait per unit increase in methylation (ie from 0 to 1). DBP: diastolic blood pressure; Glc: fasting glucose; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; SBP: systolic blood pressure; TG: triglycerides; WHR: Waist-hip ratio.

Supplementary Information Table 23. Association of methylation in blood at the 187 sentinel CpG sites with incident type-2 diabetes (T2D). Results are provided for single marker tests ('Single marker') and in joint analysis of all 187 markers ('Multi-marker'), with/without adjustment for BMI. Effect (SE): change in log odds for T2D per unit increase (ie 0 to 100%) in methylation and standard error, from logistic regression, with corresponding P value.

Supplementary Information Table 24. DNA methylation in blood as a risk factor for future, incident Type-2 diabetes (T2D) amongst 2,664 Indian Asians (1,074 with incident T2D) and 400 Europeans (200 with incident T2D) and in meta-analysis. Results are presented as relative risk for T2D (with 95% confidence intervals [CI]) per 1SD increase in Methylation Risk Score. Results are shown before and after adjustment for adiposity and glycaemic measures as conventional risk factors for T2D. P_{het} is the P value for heterogeneity of effect between Indian Asians and Europeans. WHR: Waist-hip ratio.

Supplementary Information Table 25. Cohort-specific methods for whole-genome genotyping and imputation of unmeasured genotypes for the Mendelian randomisation experiment, and for generation of principal components from whole-genome genotyping to quantify cryptic genetic population structure. MAF: minor allele frequency; HWE: Hardy-Weinberg Equilibrium.

Supplementary Information Table 26. Primers used for pyrosequencing in replication testing.

Supplementary Information Table 27. All cis-SNPs (within 500kb) associated with methylation of the 187 sentinel CpG sites in blood at $P < 0.05$. A1: is the alternate allele. A2: is the SNP effect allele. EAF: effect allele frequency (A2). Effect_{SNP-CpG}: is change in methylation per unit copy of effect allele (A2), from linear regression. P_{SNP-CpG}: is P value for the association of SNP with methylation.

Supplementary Information Table 28. Association of SNPs with DNA methylation in blood amongst Indian Asians and Europeans and in combined analysis. Results are shown for the single SNP with lowest P value for association with methylation at the sentinel CpG sites.

A1: is the alternate allele. A2: is the SNP effect allele. EAF: effect allele frequency (A2). Effect_{SNP-CpG}: is change in methylation per unit copy of effect allele (A2), from linear regression. SE_{SNP-CpG}: standard error of the effect. P_{SNP-CpG}: is P value for the association of SNP with methylation in combined analysis. P_{het}: is the P value for heterogeneity of effect between the ethnic groups.

Supplementary Information Table 29. Sensitivity analysis for the pathway analysis of candidate genes. Results are shown for the primary candidate gene criteria (nearest gene or gene expression association) and for alternate proximity selection criteria: i. Illumina: gene annotated by Illumina; ii. 10kb: all genes within 10kb of CpG site; iii. 20kb: all genes within 20kb of CpG site; iv. 40kb: all genes within 40kb of CpG site; v. 100kb: all genes within 100kb of CpG site. The 38 genes showing an association between methylation and expression in blood are included in all models.

Supplementary Information Table 30. Published GWA studies used for investigation of causal relationships between 187 sentinel methylation markers and metabolic traits linked to adiposity.

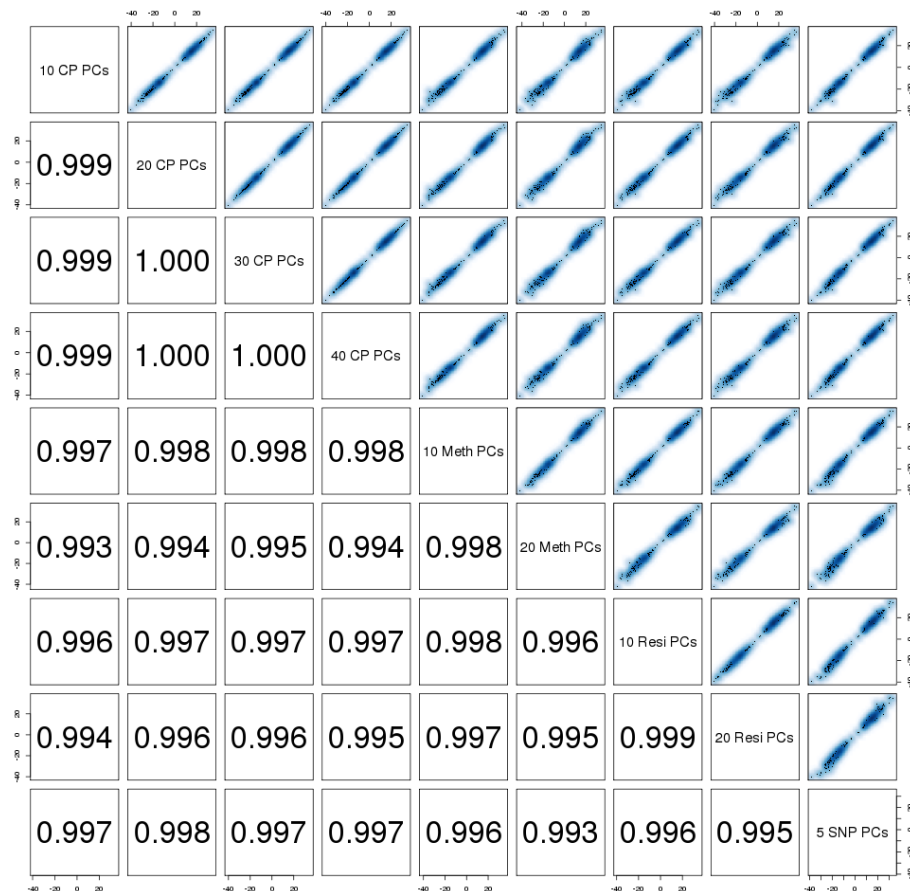
Supplementary Figures

Supplementary Information Figure 1. The accompanying pdf file provides 1Mb regional plots for the 187 sentinel CpG sites associated with BMI. Regional plots are ordered by chromosome and position and show 500kb either side of the sentinel marker (see separate file).

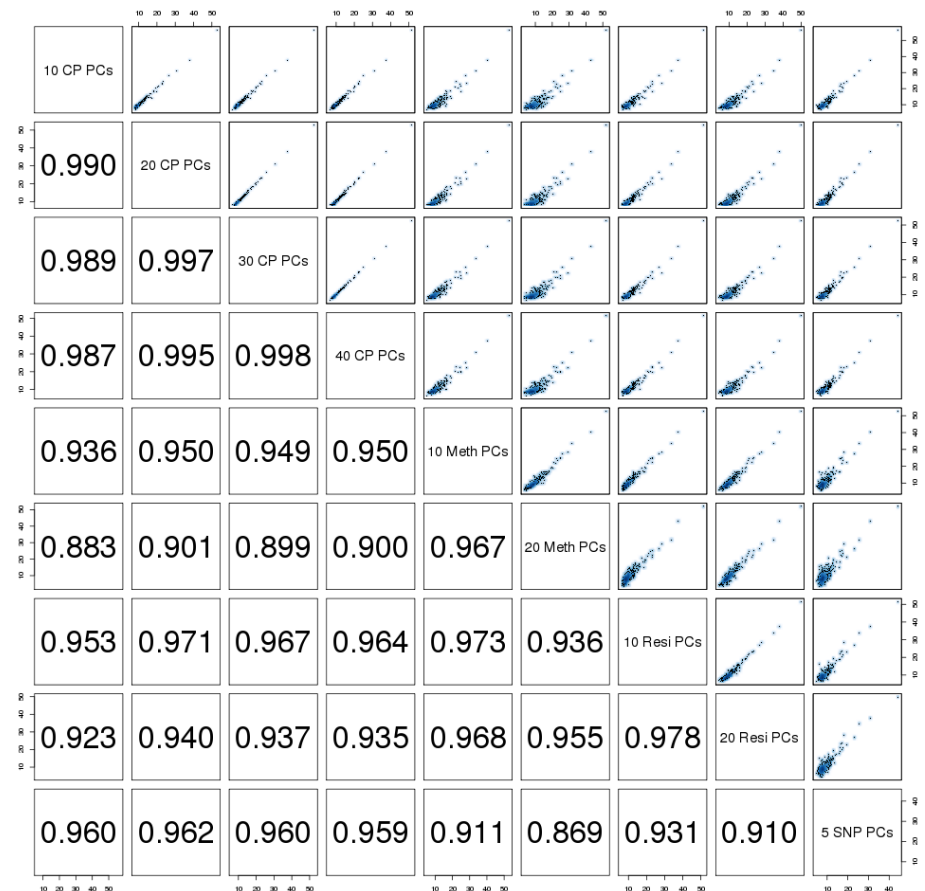
Supplementary Information Figure 2. The accompanying pdf file provides 20kb regional plots for the 187 sentinel CpG sites associated with BMI. Regional plots are ordered by chromosome and position and show 10kb either side of the sentinel marker (see separate file).

Supplementary Information Figure 3. Scatterplot matrix of effect sizes (SF4A) and $-\log_{10}(P)$ (SF4B), for association of methylation with BMI in blood derived from different models in a sensitivity analysis on the discovery epigenome-wide association study. Models evaluated are defined as follows: *20 CP PCs*: original discovery model which includes 20 control probe PCs as covariates. *10 CP PCs*, *30 CP PCs*, *40 CP PCs*: models including 10, 30 and 40 control probe PCs. *10 Meth PCs*, *20 Meth PCs*: models including 10, 20 methylation PCs in addition to the original discovery covariates. *10 Resid PCs*, *20 Resid PCs*: models including 10, 20 methylation PCs after adjusting methylation data for the discovery covariates and BMI, in addition to the original discovery covariates. *5 SNP PCs*: models including 5 SNP PCs in addition to the original discovery covariates.

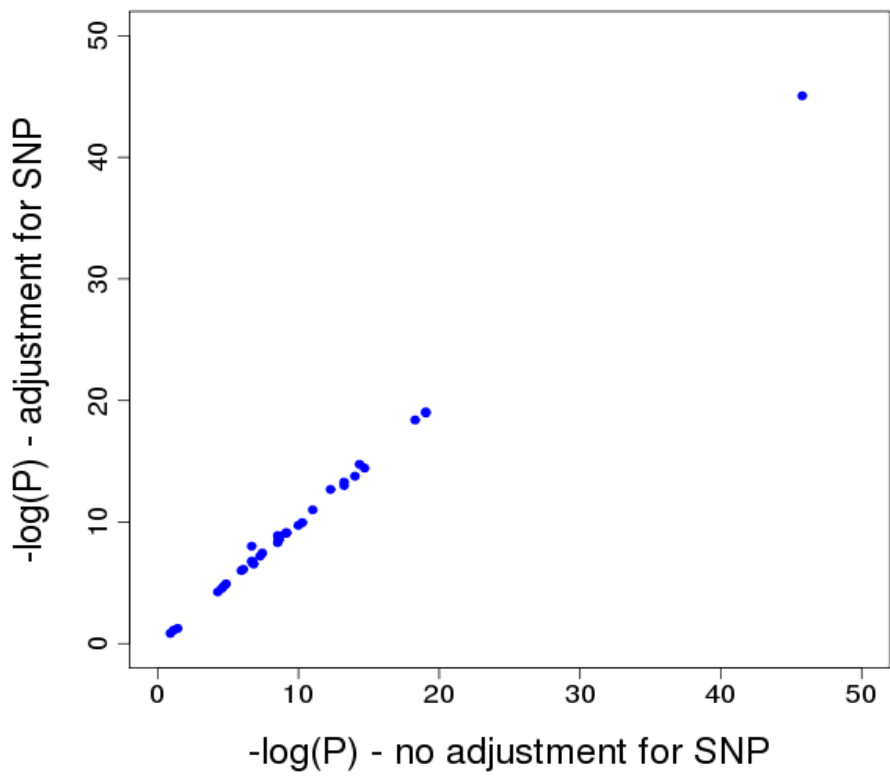
ST3A: Effect size



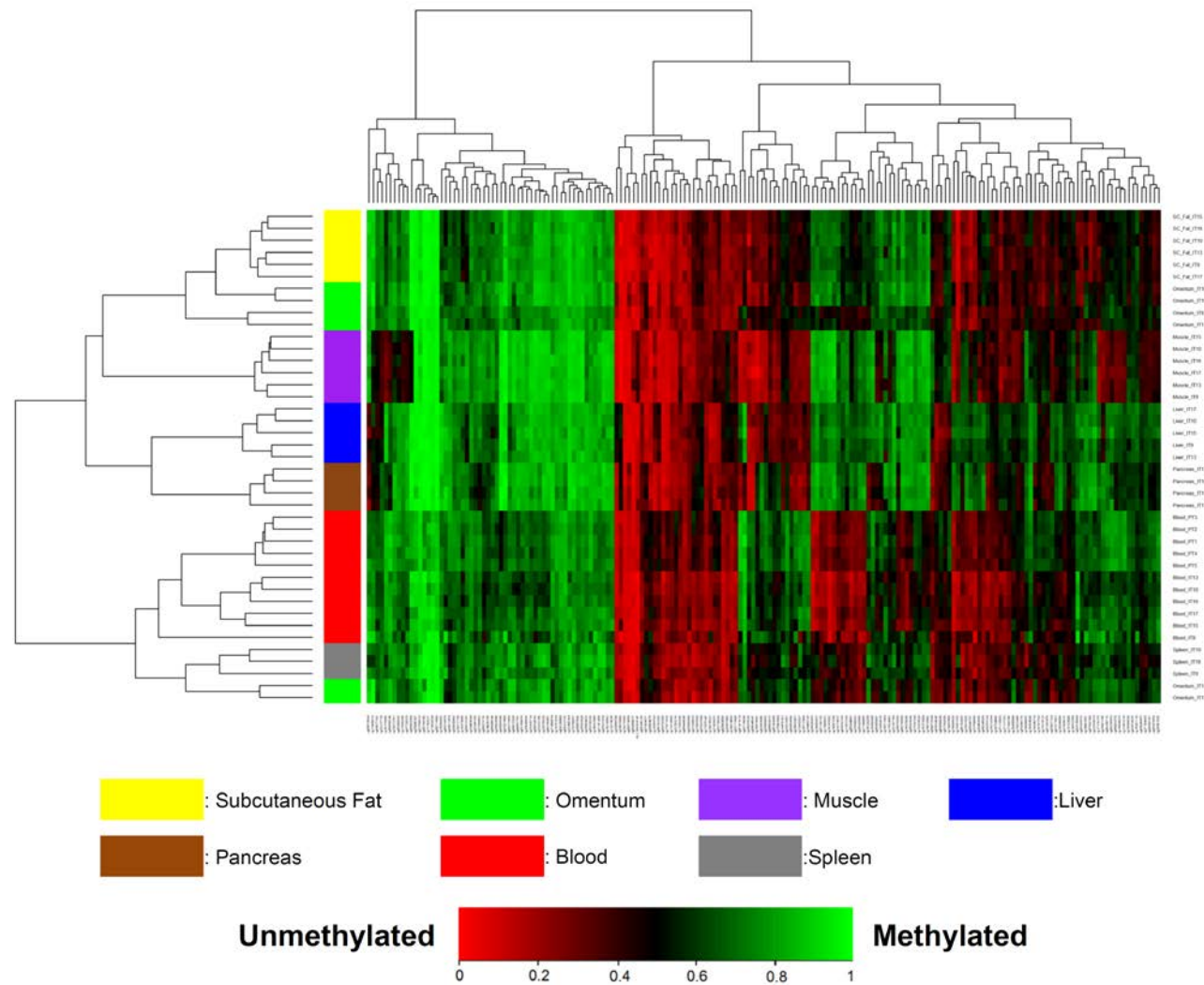
ST3B: $-\log_{10}(P)$



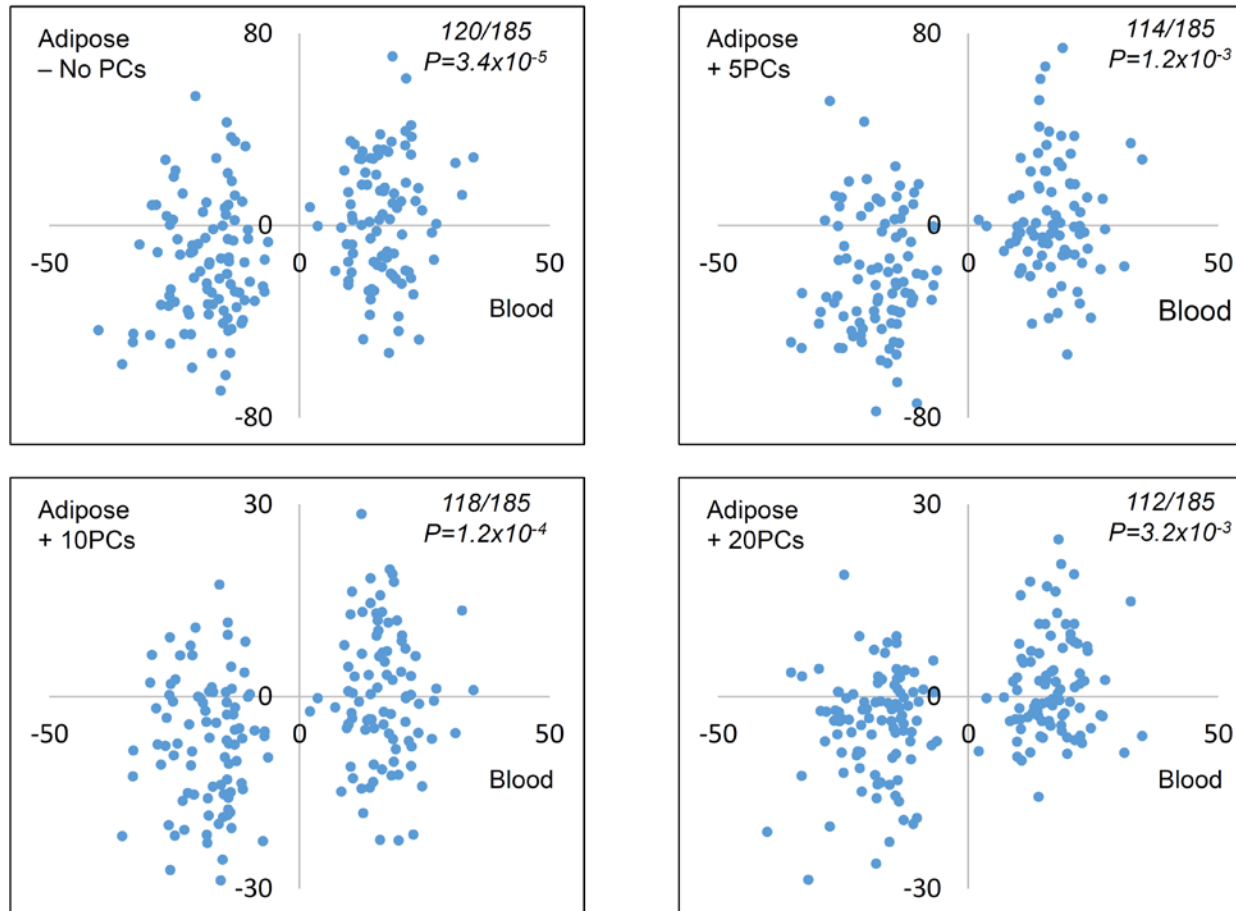
Supplementary Information Figure 4. Genetic confounding by SNPs located in Infinium 450k probe sequences. Adjustment for the SNPs in probe sequence has no material impact on the P values for association in blood between methylation and BMI at the respective CpG sites.



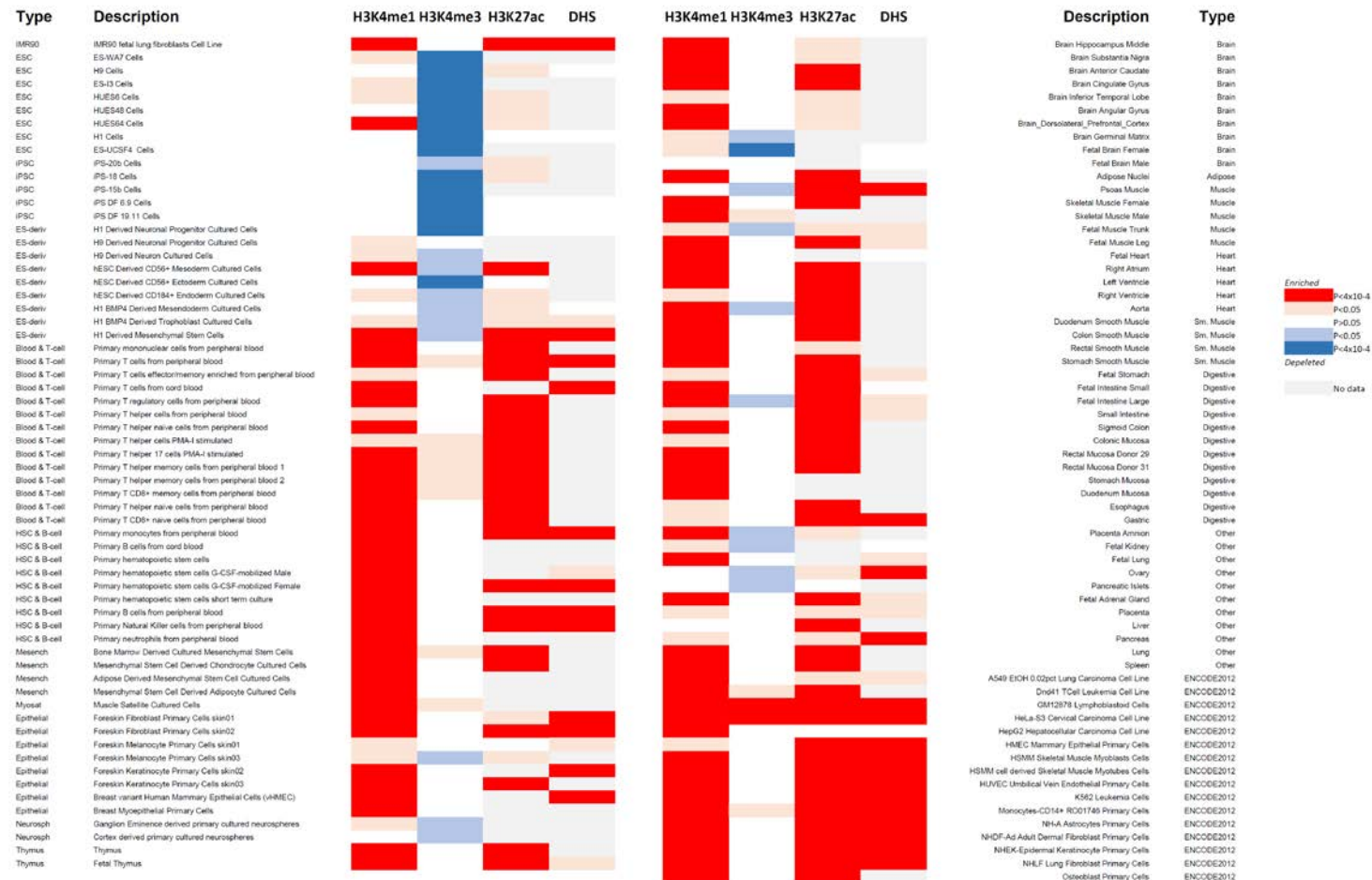
Supplementary Information Figure 5. Mean methylation levels at the 187 sentinel CpG sites across 7 tissue types. Samples (rows, N=41) and CpG sites (columns, N=187) are ordered by hierarchical clustering of methylation levels with Euclidean distance as similarity measure. Methylation values range from 0 (red) to 1 (green).



Supplementary Information Figure 6. Sensitivity analysis on the association of methylation with BMI in adipose tissue. Comparison of effect sizes for association of the 187 sentinel CpG sites with BMI in blood (x-axis) and adipose tissue (y-axis), with / without correction for up to 20 adipose tissue methylation Principal Components. The P value is for directional consistency between blood and adipose tissue at the 187 sentinel CpG sites in each model (binomial test).

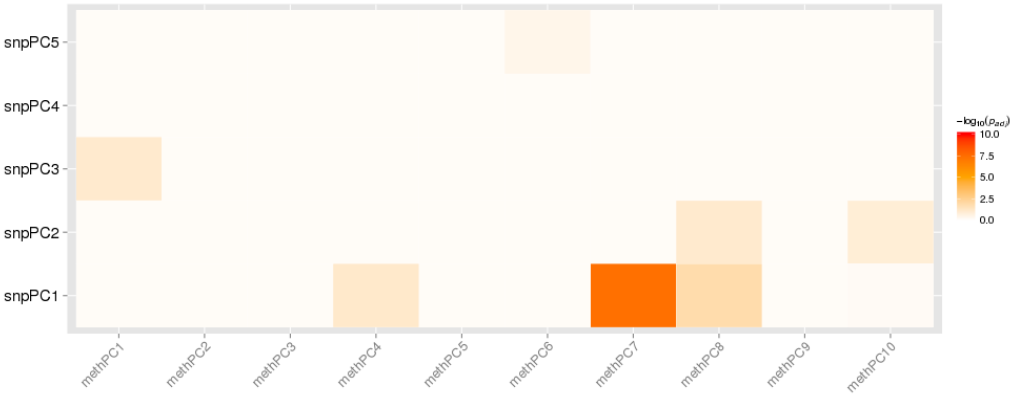


Supplementary Information Figure 7. Heatmap showing enrichment for DNase hypersensitivity sites (DHS) and histone marks at the 187 sentinel CpG sites, across 127 cell lines from the Roadmap and ENCODE projects (compendium release 9). Red boxes: enriched at $P < 4 \times 10^{-4}$ (ie $P < 0.05$ after Bonferroni correction for 127 tests); Blue boxes: depleted at $P < 4 \times 10^{-4}$ (ie $P < 0.05$ after Bonferroni correction for 127 tests); Grey boxes: no data. ESC: embryonic stem cell; iPSC: Induced Pluripotent stem cell; ES-deriv: embryonic stem cell derived; Mesench: Mesenchymal Stem cells; Neurosp: Neurospheres; ENCODE12: data from encode 2012 release.



Supplementary Information Figure 9. Relationship between Principal Components (PCs) derived from methylation data in blood (methPC, x-axis) and genetic data (snpPC, y-axis) for LOLIPOP (Panel A) and the KORA (Panel B) cohorts. P-values are derived using linear regression and adjusted for the number of tests performed.

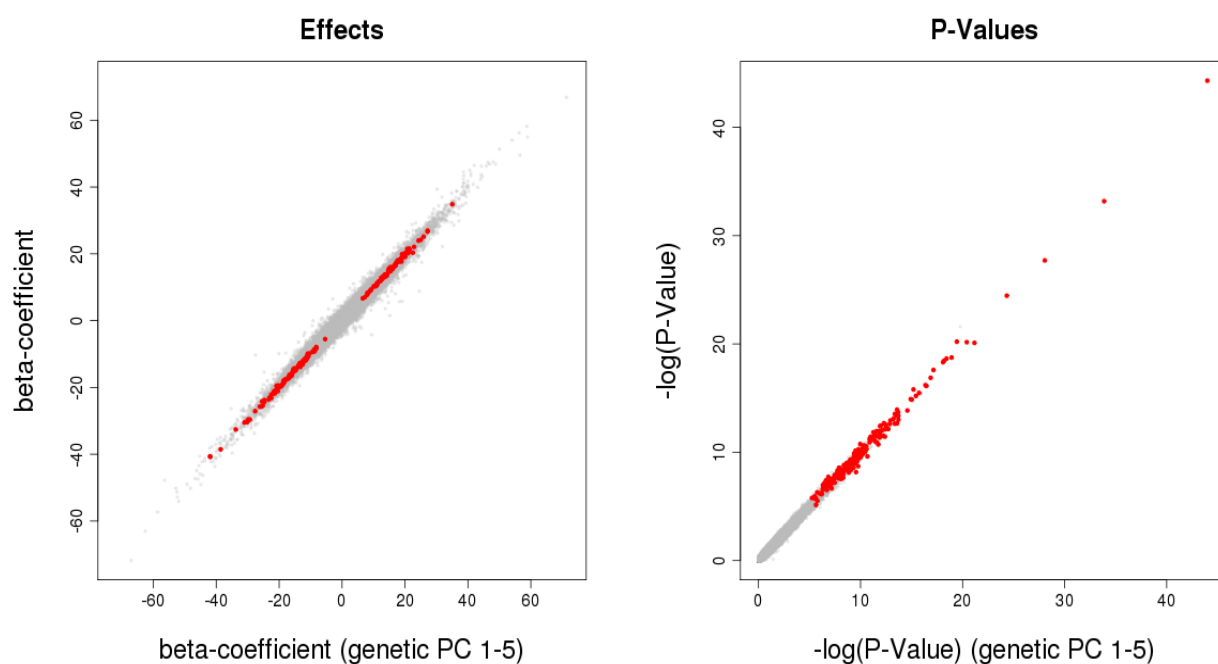
A. LOLIPOP



B. KORA



Supplementary Information Figure 10. Comparison of effects (beta-coefficients) and $-\log_{10}(P)$ for association of methylation in blood with BMI, with and without adjustment for genetic PCs 1-5. Linear regression was performed as described for the discovery epigenome-wide association study, but separately within each GWAS dataset followed by meta-analysis. Red dots represent the 187 significantly associated sentinel CpGs, grey points represent all other CpGs.



References

- 1 Holle, R., Happich, M., Lowel, H., Wichmann, H. E. & Group, M. K. S. KORA--a research platform for population based health research. *Gesundheitswesen* **67 Suppl 1**, S19-25, doi:10.1055/s-2005-858235 (2005).
- 2 Zeilinger, S. *et al.* Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PloS one* **8**, e63812, doi:10.1371/journal.pone.0063812 (2013).
- 3 Chambers, J. C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature genetics* **43**, 1131-1138, doi:10.1038/ng.970 (2011).
- 4 Chambers, J. C. *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. *The lancet. Diabetes & endocrinology* **3**, 526-534, doi:10.1016/S2213-8587(15)00127-8 (2015).
- 5 Bendinelli, B. *et al.* Fruit, vegetables, and olive oil and risk of coronary heart disease in Italian women: the EPICOR Study. *The American journal of clinical nutrition* **93**, 275-283, doi:10.3945/ajcn.110.000521 (2011).
- 6 Fiorito, G. *et al.* B-vitamins intake, DNA-methylation of One Carbon Metabolism and homocysteine pathway genes and myocardial infarction risk: the EPICOR study. *Nutrition, metabolism, and cardiovascular diseases : NMCD* **24**, 483-488, doi:10.1016/j.numecd.2013.10.026 (2014).
- 7 Boyd, A. *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* **42**, 111-127, doi:10.1093/ije/dys064 (2013).
- 8 Fraser, A. *et al.* Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International journal of epidemiology* **42**, 97-110, doi:10.1093/ije/dys066 (2013).
- 9 Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *International journal of epidemiology*, doi:10.1093/ije/dyt268 (2014).
- 10 Hofman, A. *et al.* The Rotterdam Study: 2014 objectives and design update. *European journal of epidemiology* **28**, 889-926, doi:10.1007/s10654-013-9866-z (2013).
- 11 Grundberg, E. *et al.* Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *American journal of human genetics* **93**, 876-890, doi:10.1016/j.ajhg.2013.10.004 (2013).
- 12 Scholtens, S. *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *International journal of epidemiology*, doi:10.1093/ije/dyu229 (2014).
- 13 Westendorp, R. G. *et al.* Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: The Leiden Longevity Study. *Journal of the American Geriatrics Society* **57**, 1634-1637, doi:10.1111/j.1532-5415.2009.02381.x (2009).
- 14 Ahrens, M. *et al.* DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell metabolism* **18**, 296-302, doi:10.1016/j.cmet.2013.07.004 (2013).