Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity

Epigenome-wide association reveals extensive perturbations in DNA methylation associated with adiposity and its adverse metabolic consequences.

Authors

Simone Wahl[1,2,3]*, Alexander Drong[4]*, Benjamin Lehne[5]*, Marie Loh[5,6,7]*, William R Scott[5,8]*, Sonja Kunze[1,2], Pei-Chien Tsai[9], Janina S Ried[10], Weihua Zhang[5,11], Giovanni Fiorito[12,13], Lude Franke[14], Simonetta Guarrera[12,13], Silva Kasela[15,16], Jennifer Kriebel[1,2,3], Rebecca C Richmond[17], Marco Adamo[18], Uzma Afzal[5,11], Benedetta Albetti[19], Ole Ammerpohl[20], Marian Beekman[21], Pier Alberto Bertazzi,[19] Christine Blancher[22], Marc-Jan Bonder[14], Mario Brosch[23], Maren Carstensen-Kirberg[24,25], Anton J.M. De Craen[26], Simon de Lusignan[27], Abbas Dehghan[28], Mohamed Elkalaawy[18], Krista Fischer[15], Oscar Franco[28], Tom R Gaunt[17], Jochen Hampe[23], Majid Hashemi[18], Aaron Isaacs[28], Andrew Jenkinson[18], Sujeet Jha[29], Norihiro Kato[30], Vittorio Krogh[31], Christa Meisinger[2], Thomas Meitinger[32,33,34], Zuan Yu Mok[35], Valeria Motta[19], Hong Kiat Ng[35], Salvatore Panico[36], Natalia Pervjakova[15,16], Holger Prokisch[32,33], Wolfgang Rathmann[37], Michael Roden[24,25,38], Federica Rota[19], Michelle Ann Rozario[35], Johanna K Sandling[39,40], Clemens Schafmayer[41], Katharina Schramm[32,33], Reiner Siebert[20], P Eline Slagboom[21], Richie Soong[35,42], Lisette Stolk[43], Konstantin Strauch[10,44], E-Shyong Tai[45,46,47], Letizia Tarantini[19], Barbara Thorand[2,3], Ettje F Tigchelaar[14], Rosario Tumino[48], Andre G Uitterlinden[49], Cornelia van Duijn[28], Joyce BJ van Meurs[43], Paolo Vineis[12,50], Ananda Rajitha Wickremasinghe[51], Cisca Wijmenga[14], Tsun-Po Yang[39], Wei Yuan[9,52], Alexandra Zhernakova[14], Rachel Batterham[18,53], George Davey Smith[17], Panos Deloukas[39,54,55], Bastiaan T Heijmans[21], Christian Herder[24,25], Albert Hofman[28], Cecilia M Lindgren[4,56], Lili Milani[15], Pim van der Harst[14,57,58], Annette Peters[2,3,34], Thomas Illig[1,2,59], Caroline L Relton[17], Melanie Waldenberger[1,2], Marjo-Ritta Jarvelin[60,61,62,63], Valentina Bollati[19], Tim D Spector[9]*, Jordana T Bell[9]*, James Scott[8]*, Mark I McCarthy[4,64]*, Paul Elliott[5]*, Giuseppe Matullo[12,13]*, Christian Gieger[1,2]*, Jaspal S Kooner[8,11,65]*, Harald Grallert[1,2,3]*, John C Chambers[5,11,65]*.

*: Contributed equally

Correspondence to

John C Chambers
Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK.
john.chambers@ic.ac.uk

Christian Gieger
Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany.
christian.gieger@helmholtz-muenchen.de

Harald Grallert
Research Unit of Molecular Epidemiology, Helmholtz Zentrum München, Neuherberg, Germany
harald.grallert@helmholtz-muenchen.de

Jaspal S Kooner
National Heart and Lung Institute, Imperial College London, London W12 0NN, UK.
j.kooner@ic.ac.uk

Affiliations

1. Research Unit of Molecular Epidemiology, Helmholtz Zentrum Muenchen, German Research Centre for Environmental Health, Neuherberg, Germany.
2. Institute of Epidemiology II, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany.
3. German Center for Diabetes Research (DZD), Neuherberg, Germany.
4. Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK.
5. Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK.
6. Institute of Health Sciences, P.O.Box 5000, FI-90014 University of Oulu, Finland.
7. Translational Laboratory in Genetic Medicine (TLGM), Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, Immunos, Level 5, Singapore 138648.
8. National Heart and Lung Institute, Imperial College London, London W12 0NN, UK.
9. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK, SE1 7EH.
10. Institute of Genetic Epidemiology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Neuherberg, Germany.
11. Ealing Hospital NHS Trust, Middlesex UB1 3HW, UK.
12. Human Genetics Foundation – Torino, Torino, Italy.
13. Medical Sciences Department, University of Torino, Torino, Italy.

14. Univeristy of Groningen, University Medical Center Groningen, Department of Genetics, 9700RB Groningen, The Netherlands.
15. Estonian Genome Center, University of Tartu, Riia 23b, 51010 Tartu, Estonia.
16. Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, 51010 Tartu, Estonia.
17. MRC Integrative Epidemiology Unit (IEU), School of Social and Community Medicine, University of Bristol, Bristol, UK BS8 2BN.
18. UCLH Centre for Weight Loss, Metabolic and Endocrine Surgery, University College London Hospitals, Ground Floor West Wing, 250 Euston Road, London NW1 2PG, United Kingdom.
19. Center of Molecular Epidemiology and Genetics, Department of Environmental and Occupational Health, University of Milan, Via San Barnaba 8, 20122 Milan, Italy.
20. Institute of Human Genetics, University Hospital Schleswig-Holstein, Kiel Campus, Kiel, Germany.
21. Molecular Epidemiology, Leiden University Medical Center, Leiden, 2333 ZC, The Netherlands.
22. High Throughput Genomics - Oxford Genomic Centre, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.
23. Medical Department 1, University Hospital of the Technical University Dresden, Dresden, Germany.
24. Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany.
25. German Center for Diabetes Research (DZD e.V.), Partner Site Düsseldorf, Germany.
26. Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2300 RC, The Netherlands.
27. University of Surrey, Guildford GU2 7PX, UK.
28. Department of Epidemiology, Erasmus Medical Centre, Rotterdam, The Netherlands.
29. Department of Endocrinology, Diabetes and Obesity, Max Healthcare, New Delhi 110 017, India.
30. Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, 1628655, Japan.
31. Epidemiology and Prevention Unit, Fondazione IRCSS Istituto Nazionale Tumori, Milano, Italy.
32. Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
33. Institute of Human Genetics, Technical University Munich, München, Germany.
34. Deutsches Forschungszentrum fur Herz-Kreislauferkrankungen, Munich Heart Alliance, Munich, Germany.
35. Cancer Science Institute of Singapore, National University of Singapore, Singapore.
36. Dipartmento Di Medicina Clinica E Chirurgia Federio II University, Naples, Italy.
37. Institute of Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

38. Department of Endocrinology and Diabetology, University Hospital Düsseldorf, Düsseldorf, Germany.
39. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK.
40. Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, 751 44 Uppsala, Sweden.
41. Department of Visceral and Thoracic Surgery, University Hospital Schleswig-Holstein, Kiel Campus, Kiel, Germany.
42. Department of Pathology, National University Hospital, Singapore.
43. Department of Internal Medicine, Erasmus Medical Centre, Rotterdam, The Netherlands.
44. Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany.
45. Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228.
46. Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597.
47. Duke-National University of Singapore Graduate Medical School, Singapore 169857.
48. Cancer Registry and Histopathology Unit, “Civile – M.P. Arezzo” Hospital, ASP 7, Ragusa, Italy.
49. Departments of Internal Medicine and Epidemiology, Erasmus Medical Centre, Rotterdam, The Netherlands.
50. Epidemiology and Public Health, Imperial College London, UK.
51. Department of Public Health, Faculty of Medicine, University of Kelaniya, P.O. Box 6, Thalagolla Road, Ragama 11010, Sri Lanka.
52. The Institute of Cancer Research, London, Surrey SM2 5NG, UK.
53. Centre for Obesity Research, Rayne Institute, Department of Medicine, University College London, London WC1E 6JJ, United Kingdom.
54. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ London, UK.
55. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah 21589, Saudi Arabia.
56. Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts 02142, USA.
57. University of Groningen, University Medical Center Groningen, Department of Cardiology, 9700RB Groningen, The Netherlands.
58. Durrer Center for Cardiogenetic Research, ICIN - Netherlands Heart Institute, 3511GC Utrecht, The Netherlands.
59. Hannover Medical School, Hannover Unified Biobank, Hannover, Germany.
60. Department of Epidemiology and Biostatistics, MRC Health Protection Agency (HPE) Centre for Environment and Health, School of Public Health, Imperial College London, UK. .
61. Biocenter Oulu, P.O.Box 5000, Aapistie 5A, FI-90014 University of Oulu, Finland.
62. Center for Life Course Epidemiology, Faculty of Medicine, P.O.Box 5000, FI-90014 University of Oulu, Finland.

63. Unit of Primary Care, Oulu University Hospital, Kajaanintie 50, P.O.Box 20, FI-90220 Oulu, 90029 OYS, Finland.
64. Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford, UK.
65. Imperial College Healthcare NHS Trust, London W12 0HS, UK.

Abstract

The biologic pathways linking adiposity to its metabolic complications are poorly understood. We used epigenome-wide association to explore the relationships between DNA methylation, a key regulator of genomic function, and body mass index (BMI), amongst 10,261 individuals of European and Indian Asian ancestry. We identify widespread changes in DNA methylation in blood strongly associated with BMI (187 genetic loci at $P<1x10^{-7}$, range $P=9.2x10^{-8}$ to $6.0x10^{-46}$). Genetic association analyses suggest that altered DNA methylation is predominantly the consequence of adiposity ($P=4.7x10^{-44}$), rather than the cause (P=0.86). At 11 loci methylation is associated with BMI prospectively ($P<2.7x10^{-4}$, N=1,435). Methylation at the sentinel CpG sites is also associated with BMI in subcutaneous adipose tissue (79 loci at $P<2.7x10^{-4}$, N=542), and with obesity in isolated adipocytes (6 loci at $P<2.7x10^{-4}$, N=48). The methylation markers are enriched ~3-fold for association with expression of nearby genes (±500kb, $P=3.0x10^{-4}$), and identify gene expression signatures in blood at 38 loci ($P<9.0x10^{-6}$, range $P=5.5x10^{-6}$ to $6.1x10^{-35}$, N=1,785). The methylation markers point to genes involved in lipid and lipoprotein metabolism, amino acid and small molecule transport, and inflammatory pathways including activation of *NFKB.* Methylation at the sentinel CpG sites in blood is associated with multiple adiposity related clinical traits, including glucose and lipid metabolism, inflammation and blood pressure, and in prospective studies strongly predicts future type-2 diabetes (relative risk per 1SD increase in Methylation Risk Score: 2.3 [2.07-2.56]; $P=1.1x10^{-54}$, N=3,064). Our results provide new insights into the biologic pathways linked to adiposity, and may enable new strategies for prediction and prevention of type-2 diabetes and other adverse consequences of obesity.

Main text

Overweight and obesity are major public health problems affecting ~1.5 billion people worldwide.[1,2] Both overweight and obesity are key risk factors for type-2 diabetes (T2D), cardiovascular disease and their related metabolic and inflammatory disturbances.[1-4] Although many obese individuals are insulin resistant and at high risk of progression to T2D, others remain metabolically healthy with preserved insulin sensitivity.[5] The biological mechanisms linking adiposity to its metabolic disturbances, and determining "healthy" versus "unhealthy" obesity, are poorly understood.

Recent studies suggest that adiposity may influence DNA methylation,[6-10] a key regulator of gene expression and clinical phenotype.[11,12] To extend these observations, we carried out a large-scale epigenome-wide association and replication study of DNA methylation in blood and other metabolically active tissues, to describe the changes in regulatory DNA methylation associated with adiposity, and investigate whether these alterations in DNA methylation might underlie adiposity or link adiposity to its adverse functional and metabolic consequences.

Methods and results

Our study design is summarised in **Supplementary Figure 1**. In brief, we carried out an epigenome-wide association study amongst 5,387 individuals from the EPICOR (N=514), KORA (N=2,193) and LOLIPOP (N=2,680) population studies. Our study design includes people of both European (EPICOR, KORA) and Indian Asian (LOLIPOP) ancestry, both populations known to be at high risk of obesity and related metabolic disturbances.[2,13] DNA methylation in genomic DNA from blood was determined by Illumina Infinium 450K Human Methylation array.[14] Blood was chosen for the analysis as the primary tissue available in large population-based epidemiological studies, and the most widely used sample for clinical diagnostic purposes. Characteristics of participants and analysis details are summarised in the **Supplementary Appendix** and **Supplementary Tables 1** and **2**. The association of DNA methylation with body mass index (BMI, a measure of adiposity) was tested in each cohort separately by linear regression using an analytic strategy validated to reduce batch and other technical confounding effects in quantification of DNA methylation, and to take account of the potential confounding effects arising from cryptic alterations in the white cell composition of blood (see **Online Methods**).[15,16] Inverse variance meta-analysis across the cohorts was done by METAL. We used genomic control correction as a conservative approach to adjust for test statistic inflation both in the individual cohorts (GC_{in}; inflation factors 0.98 to 1.29) and in meta-analysis ($GC_{out;}$; inflation factor 1.11).[17]. There were 466,186 autosomal markers for analysis after quality control. We set the threshold for epigenome-wide significance as $P<1x10^{-7}$, to provide a conservative Bonferroni correction for the number of markers tested.[15] As additional analyses we also investigated the relationship between BMI and DNA methylation amongst the 11,233 X-chromosomal and 417 Y-chromosomal CpG sites assayed.

Epigenome-wide association and replication testing

Epigenome-wide association identified 278 CpG sites showing an association between methylation and BMI in blood at $P<1x10^{-7}$, and distributed between 207 genetic loci (**Supplementary Tables 3 and 4**). At each locus we identified the sentinel marker (CpG site with lowest P value for association with BMI), and carried out replication testing in separate samples of whole blood from European and Indian Asian men and women in population-based studies (N=4,874, **Supplementary Table 1).** 187 of the 207 markers were associated with BMI at P<0.05 in blood, with directional consistency in replication testing, and at epigenome-wide significance in combined analysis of discovery and replication data (**Figure 1, Supplementary Table 3**). There was enrichment for

methylation-BMI associations at CpG sites with intermediate levels of methylation (214/278 sites with 20-80% methylation, a 3.7 fold enrichment compared to background, $P=1.2x10^{-88}$ Fisher's test), consistent with a relationship of differentially methylated CpG sites with phenotypic variability.

Regional plots for the 187 identified loci are shown in **Supplementary Figures 2** and **3.** Absolute effect sizes ranged from 6.3±0.9 to 40.2±3.1 kg/m^2 change in BMI per unit increase in DNA methylation in blood (scale for methylation 0-1, where 1 represents 100% methylation), with little evidence for heterogeneity between Europeans and Indian Asians (**Supplementary Table 3**). At 7 loci the associations between DNA methylation and BMI are stronger amongst Indian Asians or Europeans (Heterogeneity $P<1.0x10^{-7}$, **Supplementary Table 3**) raising the possibility that some effects may be population specific. We note that CpGs neighbouring the sentinel markers are strongly enriched for association with BMI in blood (**Supplementary Figure 4**). Conditional analyses show multiple (range 2 to 6) additional CpG sites independently associated with BMI at 23 of the 187 confirmed loci ($P<1x10^{-7}$ after conditioning on sentinel CpG site at the locus, **Supplementary Table 5**).

Results of sensitivity analyses show that our findings are robust across a range of analytic decisions (**Supplementary Figure 5**). The associations of DNA methylation in blood with BMI are independent of population stratification caused by DNA sequence variation, or by genetic confounding by SNPs in the probe sequence (**Online Methods, Supplementary Table 6, Supplementary Figures 5 and 6**). In addition, at 4 loci we further replicated the associations of DNA methylation in blood with BMI amongst 990 Europeans and 1,720 Indian Asians (LOLIPOP study), using pyrosequencing as an alternative technical approach to quantification of methylation ($P=1.2x10^{-7}$ to $2.1x10^{-12}$, **Supplementary Table 7**).

Twenty of the 207 markers did not reach $P<0.05$ in replication testing. All 20 showed consistent direction of effect between discovery and replication stages ($P=1.9x10^{-6}$, binomial test, **Supplementary Table 3**), suggesting that the majority are unlikely to be false positives. There were no CpG sites associated with BMI in blood at $P<1x10^{-7}$ on the sex-chromosomes.

Genetic association to investigate relationships between DNA methylation and BMI

We used genetic association and the concept of Mendelian randomisation to investigate the potential causal relationships between DNA methylation in blood and BMI.[18-20]

We first used genome-wide association to identify SNPs influencing DNA methylation in blood in *cis* (1Mb, N=4,034 people, see **Online Methods**). We tested the association of SNPs with DNA methylation in blood, amongst Europeans and Indian Asians Indians separately, followed by meta-analysis (**Supplementary Table 8**). There was little evidence for heterogeneity of effect between the populations ($P>2.7x10^{-4}$ at 177/187 loci, corresponding to P>0.05 after Bonferroni correction for 187 tests, **Supplementary Table 8**), indicating that the genetic influences on DNA methylation are largely shared rather than population specific.

We then tested whether the SNPs that influence methylation in blood are separately associated with BMI, and whether the predicted effects of SNPs on BMI via methylation are consistent with the directly observed association (further details in **Online Methods**). We identify a single CpG (cg26663590: *NFATC2IP*) showing potential evidence for a contribution of methylation to BMI ($P=9.6x10^{-7}$, **Figure 2A** and **Supplementary Table 9**). The *NFATC2IP* locus has previously been identified to be associated with obesity through genome-wide association;[21] the locus contains the gene encoding SH2B1 which is known to be involved in energy and glucose homeostasis.[22] At the other loci evaluated there was little relationship between the effects of the SNPs on BMI predicted via methylation and that directly observed ($R^2=0.00$, P=0.86). These findings suggest that the majority of methylation sites studied are unlikely to directly influence levels of BMI.

Next we investigated whether DNA methylation in blood is the consequence of adiposity. Here we evaluated whether SNPs known to influence BMI also influence DNA methylation in blood, and whether the predicted effect of SNP on methylation through BMI, is consistent with the directly observed effect of SNP on methylation. We used a weighted genetic risk score (GRS) to combine effects across BMI SNPs (**Figure 2B and Supplementary Table 10**). We observe a strong correlation between predicted and observed effects of BMI SNPs on methylation ($R^2=0.81$; $P=4.7x10^{-44}$) across the CpG sites evaluated. In addition, GRS was associated with DNA methylation at the *ABCG1, KLHL18, FTH1P20* loci ($P<2.7x10^{-4}$, corresponding to P<0.05 after Bonferroni correction for 187 tests). An effect of body mass on *ABCG1* methylation is consistent with observations that weight loss influences both *ABCG1* expression in adipose tissue and ABCG1 activity.[23,24] Although the mechanisms remain to be elucidated, our findings support the view that BMI may contribute to altered methylation in blood at the majority of the identified CpG sites.

DNA methylation in blood and adiposity in prospective population studies

We used longitudinal data from prospective population studies to further explore causal relationships between DNA methylation in blood and BMI. First, we studied the association of methylation in blood at baseline, and change in BMI during follow-up of up to 11 years, amongst 1,435 Europeans and 1,513 Indian Asians (**Online Methods**). At cg26663590 (*NFATC2IP* locus, identified by our genetic association studies to have a potential causal effect underlying BMI) there was a positive association of baseline methylation level with change in BMI during follow-up (P=0.02, **Supplementary Table 11**), further supporting the observation that methylation at the *NFATC2IP* locus might causally influence BMI.

Next we studied the relationship between change in BMI and DNA methylation amongst 1,435 Europeans with paired measurements of adiposity and methylation in blood at baseline and follow-up. At 11 CpG sites, we observed a directionally consistent and significant association between change in BMI and change in methylation ($P<2.7x10^{-4}$, corresponding to P<0.05 after Bonferroni correction for 187 tests, **Supplementary Table 11**), including the *ABCG1* locus ($P=1.9x10^{-9}$). Furthermore, the relationship between BMI and methylation in blood at the 187 loci is closely correlated (R=0.81, $P=4.9x10^{-45}$) and directionally consistent (178 of 187 loci, $P=6.8x10^{-42}$, binomial test) in the longitudinal and cross-sectional data (**Supplementary Table 11** and **Figure 3**). These findings provide further evidence that BMI contributes to methylation changes in blood at the majority of the identified methylation sites.

DNA methylation in adipose tissue and isolated adipocytes

We next investigated whether our findings in blood might also reflect processes in adipose tissue. We quantified DNA methylation at the 187 sentinel methylation markers in genomic DNA from subcutaneous adipose tissue (N=542 people). We found that 120 of the CpG sites showed directional consistency for association with BMI between adipose tissue and blood ($P=1.3x10^{-4}$, binomial test), while 91 sites were associated with BMI in adipose tissue ($P<2.7x10^{-4}$, corresponding to P<0.05 after Bonferroni correction for 187 tests, **Supplementary Table 12**). Paired samples of blood and adipose tissue were available for 201 people, enabling direct comparison of DNA methylation levels between blood and adipose tissue. Methylation levels were positively correlated at 149 of the CpG sites studied (**Supplementary Table 12**, $P=9.8x10^{-17}$ for directional consistency, binomial test). Our findings demonstrate that many of the relationships between methylation and BMI in blood are shared by adipose tissue, but also identify effects that are tissue specific.

Alterations in tissue levels of DNA methylation may represent changes in cell composition.[16] Since histological assessment of cell composition was not available for the adipose samples, we used Principal Components Analysis (PCA) to assess for cryptic structure in the methylation data (**Online Methods**). Including principal components as covariates in regression models did not materially influence the association of DNA methylation with BMI in adipose tissue, suggesting that the associations observed are unlikely to be the result of differences in the composition of canonical cell-types (**Supplementary Figure 7**).

To obtain further evidence for the relationship of adiposity with DNA methylation in adipose tissue, independent of cell composition, we quantified the DNA methylation at the 187 loci in isolated adipocytes from subcutaneous adipose tissue collected from morbidly obese (BMI>40kg/m2, N=24) and normal weight (N=24) individuals (**Online Methods**). Despite the sample size, 6 markers were associated with obesity at $P<2.7x10^{-4}$ ($P<0.05$ after Bonferroni correction for 187 markers, **Supplementary Table 13**), while 108 markers show a relationship with obesity that is directionally consistent with the discovery EWAS ($P=0.04$).

Finally we used genetic association and the concept of Mendelian randomisation to investigate the causal relationships between BMI and DNA methylation in adipose tissue (**Online Methods**). Results replicate our findings in blood and confirm that, in adipose tissue as well as in blood, the differences in methylation observed are primarily the consequence of adiposity ($R=0.73$, $P=1.6\times10^{-32}$; **Supplementary Figure 8**).

Cross-tissue patterns of DNA methylation

To further describe the relationship of DNA methylation in blood with other metabolically relevant tissues, we compared methylation levels at the 187 loci in blood, subcutaneous and omental fat, liver, muscle, spleen and pancreas using 41 samples from 6 individuals.[25] We find that 30.6% of the loci are consistently hypo- or hypermethylated (average methylation value <0.5 or >0.5) across the 7 tissues, while the other 69.4% show tissue specificity. Mean methylation levels at the 187 loci correlate moderately to strongly between the tissues ($R=0.37$ to 0.93, $P=8.9x10^{-8}$ to $1.9x10^{-82}$ for the 21 tissue pairs, **Supplementary Figures 9** and **10**). Correlations are highest between blood, subcutaneous and omental fat, suggesting the highest similarity in methylation levels for these tissues. Although phenotypic data are not available for these samples to assess for confounding effects, our findings provide support for the view that methylation levels in blood are related to methylation patterns in other tissues at the CpG sites examined.

Functional genomic and gene expression studies

The 187 sentinel CpGs sites identified are enriched for location within CpG shores and 'open sea', and depleted in CpG islands and transcription start sites, compared with expectations under the null hypothesis (P<0.05 by permutation testing, **Supplementary Table 14**). The CpG sites are also strongly enriched in active chromatin sites, including at DNase hypersensitivity sites and the activating histone marks H3K4me1 and H3K27ac in a wide range of cell lines (P<0.05, **Supplementary Figure 11**) suggesting that adiposity-related methylation changes occur more likely at *cis*-regulatory regions.[11,26]

We therefore examined the relationships between DNA methylation at the 187 identified CpG sites and expression of *cis*-genes (500kb) in blood from 878 Europeans and 907 Indian Asians (**Supplementary Tables 15 and 16**). The choice of 500kb was guided by published studies,[27] and by the results of permutation testing (**Supplementary Figure 12**) both of which support an association of methylation with gene expression over this distance. We found 44 transcripts of 38 annotated genes that are associated with DNA methylation at $P<9.0\times10^{-6}$ (ie P<0.05 after Bonferroni correction for 5,551 CpG-transcript tests, **Supplementary Table 16**); this represents an ~3-fold enrichment compared to expectations under the null hypothesis ($P=3.0\times10^{-4}$, **Supplementary Figure 12**)**.** Sensitivity analyses limiting assessment of the relationship between methylation and gene expression to nearest gene, or to Illumina annotated gene, reveal five additional potential expression associations (**Supplementary Table 17**).

The majority of the observed associations between methylation and gene expression are inverse (82%, **Supplementary Table 16**). The strongest *cis*-signals observed are for cg09315878 with *TNFRSF4* transcription ($P=7.2\times10^{-86}$), cg14476101 with *PHGDH* transcription ($P=1.0\times10^{-64}$) and cg09152259 with *MAP3K2* transcription ($P=1.6\times10^{-67}$). Interestingly, both *TNFRSF4* and *MAP3K2* encode proteins involved in activation of NF-KB, a key regulator of inflammation and immune response.[28]

We also examined the relationship between DNA methylation and gene expression in adipose tissue (N=499 people) and liver (N=70 people). Amongst the 38 methylation-gene expression associations observed in blood, 3 replicated in adipose tissue *(HOXA5, BBS2, SELM)* and 3 in liver (*ANXA1, LGALS3BP, PHGDH*) at $P<1.3\times10^{-3}$ (ie P<0.05 after Bonferroni correction for 38 tests), all with consistent direction of effect (**Supplementary Table 18**). Although limited by small sample size, our findings show that the relationships between methylation and gene expression are in part shared between blood, adipose and liver tissue.

Ontology analysis of the loci associated with BMI

Next we sought to identify the genetic pathways that may be influenced by adiposity associated DNA methylation, and thus potentially contributing to the adverse metabolic consequences of obesity. We prioritised potential candidates for genes involved in the association between BMI and DNA methylation at the 187 genetic loci based on two criteria: i. Proximity: gene nearest to the sentinel methylation marker, and ii. Functional genomics: genes within 500kb of the sentinel methylation marker showing association of gene expression with methylation (38 genes, **Supplementary Table 19**). These criteria identified 210 unique genes, that also overlapped 136 (94%) of the 144 genes annotated by Illumina at these loci (**Supplementary Table 19)**. Current knowledge on all 210 candidate genes is summarised in **Supplementary Table 20**.

Gene-set enrichment analysis on the set of 210 candidate genes highlights genes involved in lipid and lipoprotein metabolism, amino acid and small molecule transport, and inflammatory pathways involving *NFKB*, *MAPK*, *TAK1*, *IRAK2* and *TRAF6* (**Supplementary Table 21**). Sensitivity analyses show that results of gene-set enrichment analyses are similar with prioritization of candidate genes based on Illumina 450K array annotation files; in contrast results become less statistically significant when gene selection includes more permissive criteria for the definition of proximity (**Supplementary Table 22**).

DNA methylation and metabolic disturbances associated with adiposity

We then tested the cross-sectional relationship of DNA methylation in blood at the 187 sentinel CpG sites, with clinical traits including fasting glucose, insulin, HDL cholesterol, triglycerides, and HbA1c. The associations of DNA methylation in blood with clinical traits were tested amongst Europeans (KORA, N=1,697) and Indian Asians (LOLIPOP, N=2,462) separately, and results combined across studies by inverse variance meta-analysis. We found that 878 methylation-clinical trait pairs tested were significant at $P<2.1x10^{-5}$ (ie $P<0.05$ after Bonferroni correction for the 2,431 tests performed) a 7-fold enrichment compared to expectations under the null hypothesis ($P<10^{-100}$; binomial test; **Supplementary Figure 13, Supplementary Table 23**). These findings are consistent with recent studies reporting close relationships of DNA methylation with blood lipids and glucose traits.[29,30]

We used genetic association to investigate the potential causal relationships between DNA methylation in blood and clinical traits (**Online Methods**). We found that SNPs influencing methylation markers in blood showed little evidence for association with the

respective clinical traits (**Supplementary Figure 14**). Our findings argue against a causal role for the identified methylation markers as mediators of the relationship between BMI and the clinical traits examined.

We therefore tested whether the methylation markers might be the consequence of alterations in the metabolic phenotypes. For each clinical trait we calculated weighted genetic risk scores based on published genome-wide association studies (GWAS, **Supplementary Table 24**). We find that the predicted effect of GRS on DNA methylation in blood via clinical trait is correlated with the directly observed effect of GRS on methylation for HbA1c, HDL cholesterol, triglycerides and insulin ($P=2x10^{-3}$ to $P=2x10^{-12}$, **Supplementary Figure 14**). Although the mechanisms remain to be determined, our findings suggest that the alterations of methylation associated with BMI in blood may, at least in part, be a consequence of changes in lipid and glucose metabolism.

Adiposity associated methylation predicts Type-2 diabetes

Finally we tested whether DNA methylation levels in blood at the 187 sentinel CpG sites predict new onset, incident T2D, the major clinical consequence of obesity, amongst participants of the LOLIPOP study (N=2,664, see **Online Methods**). In single marker tests, 62 of the 187 methylation markers are associated with incident T2D at $P<2.7x10^{-4}$ (ie $P<0.05$ after Bonferroni correction for 187 tests), while in a fully saturated multivariate model 18 CpG sites remain associated with incident T2D at $P<0.05$ (**Supplementary Table 25**). The strongest association was observed for the *ABCG1* locus, a gene known to be involved in insulin secretion and pancreatic β-cell function.[31-33]

For each participant we calculated a weighted Methylation Risk Score (MRS) as the sum of the standardised methylation values in blood, at each of the markers associated with T2D at $P<0.05$ in the multivariate model, weighted by marker-specific effect size. MRS was predictive of incident T2D (relative risk 2.29 [95% CI 2.06-2.55] per 1SD change in MRS; $P=4.2x10^{-52}$). The association of MRS with incident T2D replicated in an independent sample of 200 T2D cases and 200 controls of European ancestry from the KORA study (relative risk 2.51 [95% CI 1.49-4.23] per 1SD change in MRS; $P=5.7x10^{-4}$), with no evidence for heterogeneity of effect ($P=0.74$).

We find that MRS predicts T2D beyond traditional risk factors including BMI and waist-hip ratio (**Supplementary Table 26**). Furthermore, we find that DNA methylation identifies obese and overweight individuals at high risk of future T2D: for example amongst obese individuals the relative risk for T2D is 7.3 (4.1-12.9, $P=8.2x10^{-12}$) in the top vs the lowest quartile of DNA methylation (**Figure 4**). Although genetic association does not support a

causal role for methylation in blood at the identified CpG sites underlying the development of T2D (**Supplementary Figure 15**), our findings raise the possibility that DNA methylation may help identify individuals with metabolically unfavourable adiposity at increased risk of future T2D.

Conclusions

Our large-scale epigenome-wide association study identifies and replicates extensive changes in DNA methylation in blood and adipose tissue associated with BMI in two population groups. Genetic association in both blood and adipose tissue supports the view that the changes in DNA methylation are a consequence and not the cause of adiposity, at the majority of the identified CpG sites. The methylation markers highlight sites of active chromatin, and identify genes known to be involved in lipid metabolism, amino acid and small molecule transport, and inflammation. Our findings provide new insight into the regulatory pathways potentially underlying the adverse metabolic effects of adiposity. In addition, our prospective population studies amongst both Europeans and Indian Asians show that DNA methylation in blood identifies people at high risk of incident T2D, independent of conventional risk factors. DNA methylation in blood thus distinguishes metabolically unhealthy obesity, and may enable development of new approaches to risk stratification and personalized medicine, to help tackle the current global epidemic of obesity and its associated cardiovascular and metabolic disturbances.

Figures

Figure 1. Circos plot of the epigenome-wide association of DNA methylation in blood with BMI. Results are presented as CpG specific association test results [-log10(P)] ordered by genomic position. Green and blue symbols: CpG sites at loci reaching epigenome wide significance ($P<1x10^{-7}$); grey symbols: CpG sites at loci not reaching epigenome-wide significance. Chromosome numbers are shown on the inner ring. Tick marks on the outer ring identify the genomic loci reaching epigenome-wide significance. The genes nearest to the sentinel methylation markers at each of the 187 loci are listed around the circos plot.

Figure 2. Genetic association studies to investigate the potential relationships between BMI and DNA methylation in blood. **2A. Causal analysis** shows results for a causality analysis investigating whether DNA methylation in blood at the sentinel CpG sites influences BMI. Units are change in BMI per copy of effect allele. For each sentinel CpG site we identified the *cis*-SNP (1Mb) most closely associated with DNA methylation levels. For each SNP we then determined i. the effect of SNP on BMI predicted via methylation (x-axis), ii. the directly observed effect of SNP on BMI (y-axis). Grey points represent CpGs not significantly associated with a SNP; blue points represent CpGs significantly associated with a SNP. For a single CpG (*NFATC2IP*) the associated SNP is also associated with BMI and 95% confidence interval error bars are shown. At the other loci there was little relationship between the effects of the SNPs on BMI predicted via methylation and that directly observed (R^2=0.00, P=0.86). **2B**. **Consequential analysis** shows results for a causality analysis investigating whether DNA methylation in blood at the sentinel CpG sites is the consequence of BMI. Units are change in methylation per unit change in weighted genetic risk score (GRS). We identified the SNPs reported to influence BMI in GWAS meta-analysis,[21] and calculated a weighted GRS (see **Online Methods**). For each sentinel CpG site we then determined i. the effect of GRS on methylation predicted via BMI (x-axis) and ii. the directly observed effect of GRS on CpG (y-axis). Three CpGs (*ABCG1, KLHL18, FTH1P20*) are associated with the GRS at $P<2.7x10^{-4}$ ($P<0.05$ after Bonferroni correction for 187 tests). 95% confidence interval error-bars shown. The overall correlation between observed and predicted effects (R^2=0.81; P=4.7 x 10^{-44}) suggesting that methylation in blood at the majority of CpG-sites is consequential to BMI.

Figure 3. Relationship between DNA methylation in blood and BMI amongst 1,435 participants of the KORA S4/F4 population cohort. Cross-sectional results (x-axis) are for the relationship between methylation in blood and BMI at each of the 187 sentinel CpG sites in the baseline samples; longitudinal results are for the relationship between change in methylation (in blood) and change in BMI after 7 year follow-up. Units for both axes are kg/m^2 change in BMI per unit change in methylation (scale 0-1, where 1 represents 100% methylation).

Figure 4. Relative risk of incident T2D by quartile of Methylation Risk Score amongst normoglycaemic Indian Asians (HbA1c<6% and fasting glucose<6mmol/l) with normal weight (BMI 18.5-24.9kg/m^2), overweight (BMI 25.0-29.9kg/m^2) and obese (BMI $\geqslant$ 30.0kg/m^2). The P value is for the interaction between adiposity and DNA methylation on risk of T2D.

Online Methods

Population samples.

Details of the population samples are provided in the Supplementary appendix.

Quantification of DNA methylation

Quantification of DNA methylation

DNA methylation was quantified in bisulfite converted genomic DNA from whole blood, using the Illumina Infinium HumanMethylation450 array in all samples.[34] Cohort specific methods are summarised in **Supplementary Table 2**. DNA methylation was quantified on a scale of 0-1, where 1 represents 100% methylation.

Preprocessing and quality control criteria are summarised in **Supplementary Table 2**. Briefly, in the LOLIPOP and KORA studies, raw signal intensities were retrieved using the function *readIDAT* of the R package *minfi*, version 1.6.0[34], from the Bioconductor open source software (http://www.bioconductor.org/), followed by background correction with the function *bgcorrect.illumina* from the same R package. Detection P values were derived using the function *detectionP* as the probability of the total signal (methylation + unmethylated) being detected above the background signal level, as estimated from negative control probes. Signals with detection P values ≥ 0.01 were removed. Similarly, signals summarized from less than three functional beads on the chip were removed. Observations with less than 95% CpG sites providing a signal were subsequently excluded from the data set. To reduce non-biological variability between observations, data were quantile normalized with the function *normalizeQuantiles* of the R package *limma*, version 2.12.0[35], from Bioconductor, separately in six probe categories based on probe type and colour channel.[13,14] If not stated otherwise, this preprocessing pipeline was used for all data used in downstream analyses.

In order to account for technical effects during the experiment, we performed principal component analysis (PCA) on the signal intensities for the 235 positive control probes on the 450k array, which assess multiple steps in the laboratory processing.[13] The resulting principal components (PCs) are thought to capture technical variability in the experiment and the first 20 control probe PCs were included as covariates in the model to remove technical biases.[14]

To estimate proportions of white blood cell types, we used the method by Houseman *et al.*[36] They provide 500 CpG sites showing the most pronounced cell type specific methylation levels in an experiment based on purified cells. Of these, 473 CpGs were

available on the 450k array. Following the proposed procedure and using the R code provided with the manuscript (R function *projectWBC*), we used these 473 CpG sites to infer white blood cell proportions (i.e., proportion of granulocytes, monocytes, B cells, CD4+ T cells, CD8+ T cells and natural killer cells) in our samples. These proportions were subsequently used as covariates in the model to avoid cell type confounding.

Epigenome-wide association

We performed single marker tests separately in each cohort using linear regression to examine the association of each autosomal CpG site with BMI. We adjusted for age, gender, smoking status, physical activity index and alcohol consumption, as well as for the first 20 control probe PCs and for the estimated white blood cell proportions; this set of covariates is henceforth referred to as "discovery covariates". Finally we corrected the association results for the genomic control inflation factor (GC_{in}), in order to account for population stratification and other forms of cryptic structure in the data, which can for instance arise from unobserved confounding. Markers on the sex chromosomes were tested similarly for association with BMI, but separately in men and women. Results were combined across cohorts by inverse variance meta-analysis using METAL version 2011-03-25 (http://www.sph.umich.edu/csg/abecasis/Metal/). The resulting P values where then corrected for in a second round of genomic control (GC_{out}).

To assess the stability of discovery results towards the analytic choices made, we performed sensitivity analyses to determine the impact of control probe PCs, methylation PCs, and genetic PCs as covariates. Specifically, we compared results from the discovery meta-analysis when the first 10, 20, 30 and 40 control probe PCs were included as covariates, 10 or 20 PCs derived from a PCA on the matrix of methylation β-values, 10 or 20 PCs derived from a PCA on the matrix of methylation values adjusted for the discovery covariates and BMI, or 5 PCs derived from a PCA on SNP data were included as covariates. PCA of the methylation data was performed separately for each cohort based on quantile normalised beta-values of autosomal probes without missing data. Genetic PCs (SNP PCs) were generated separately for each cohort and genotyping platform (**Supplementary Tables 27**). The correlation between SNP PCs and methylation PCs was assessed using linear regression (**Supplementary Figure 16**). Discovery results are very stable towards the considered variations in covariates, with correlations of effect sizes between the models varying between 0.99 and 1.0 (**Supplementary Figures 5 and 17**). In addition, SNPs in the probe sequences did not materially affect the observed associations (**Supplementary Figure 6, Supplementary Table 6**).

Replication testing

Markers associated with BMI at $P<1x10^{-7}$ in the discovery experiment as within ±500 kb of each other were considered as a single genetic region. At each locus we identified the CpG sites with lowest P value for association with BMI (sentinel marker). Our choice of 1Mb to define a genetic locus was made to take account of long-range enhancers. To ensure a complete description of methylation markers associated with BMI, we then used conditional analysis to examine for secondary signals at the locus; association testing for secondary markers was carried out using the model and covariates described previously, but in addition including methylation at the respective sentinel marker as covariate.

The 207 sentinel CpG sites were analysed similarly in the replication samples; cohort-specific details of analysis pipelines are described in **Supplementary Table 2**. Results were combined across discovery and replication by weighted z 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 is supported by the results of permutation testing.[15].

To assess whether the 187 identified sentinel CpGs were enriched for intermediately methylated CpGs (sites with 20-80% average methylation), we randomly generated 100,000 sets of 187 CpGs and determined the number of intermediately methylated CpGs for each of them in order to derive an expected distribution under the null hypothesis of no enrichment. We then compared the observed number of intermediately methylated CpGs for the 187 sentinel CpGs against the null distribution to calculate an empirical P value.

An exact binomial test (R function *binom.test*) was used to test whether consist direction of effect between discovery and replication was observed more often than expected by chance amongst the 20 non-replicating CpG sites.

Replication by pyrosequencing

As a technical validation we used pyrosequencing to carry out replication testing of the relationship between DNA methylation and BMI at 4 loci, using samples of whole blood from 990 Europeans and 1,720 Indian Asians participating in the LOLIPOP study. Pyrosequencing was carried out using biotinylated primers to amplify bisulfite-treated DNA (**Supplementary Table 28**). 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).

Genetic association studies

We used genetic association and the concept of Mendelian randomisation to investigate for potential causal relationships between DNA methylation and adiposity.[18-20] Briefly, Mendelian Randomisation goes back to the more general *instrumental variable* concept. As an instrumental variable, it uses a genetic variant (or a combination of genetic variants) Z associated with a variable X in order to show causal relation between X and another variable Y. It relies on the fact that the alleles of a genetic variant are inherited randomly from parents to offspring, so that the relation of a genetic variant with a phenotype should not be confounded (with exceptions including population stratification). Thus, if the effect of X on Y is causal and the study has enough power, Z should also associate with Y. Specifically, the predicted association of Z with Y can be calculated as follows, assuming linear relationships and assuming that Z is unrelated to Y given X and unrelated to any unobserved confounders U[22]:

(1) $X = \alpha_1 + \beta_1 Z + \gamma_1 U$, where $\gamma_1 U$ plays the role of the error term that is per assumption unrelated to Z

(2) $Y = \alpha_2 + \beta_2 X + \gamma_2 U = \alpha_2 + \beta_2(\alpha_1 + \beta_1 Z + \gamma_1 U) + \gamma_2 U = \alpha_2 + \beta_2\alpha_1 + \beta_2\beta_1 Z + (\beta_2\gamma_1 + \gamma_2)U = \alpha_3 + \beta_3 Z + \gamma_3 U$

➔ Predicted effect of Z on Y: $\beta_3 = \beta_2\beta_1$

Unbiased estimation and formal inference on the causal effect β_1 of X on Y (where X and Y represent a CpG-phenotype-pair) heavily relies on strong genetic effects and typically requires tens of thousands of samples for adequate power.[35,36] Since these sample sizes are currently not available for epigenomic datasets we instead explored consistency of the predicted effect of Z on Y versus the actually observed effect, thereby obtaining some indication on the plausibility of a causal effect of X on Y. This was done in two directions, studying causality of the effect of DNA methylation (X) on BMI (Y) and of BMI (X) on DNA methylation (Y).

DNA methylation as determinant of BMI (causal analysis)

To address the question of DNA methylation being a determinant of BMI (whereby X=DNA methylation, Y=BMI) we used data on genetic variants from 4,034 participants of the KORA and LOLIPOP studies (**Supplementary Table 27**) to identify *cis* (1Mb) SNPs (Z) influencing methylation in blood at the 187 sentinel CpG sites. The associations between SNPs and methylation were tested in each data set separately using linear

models with methylation as response and SNP as independent variable, adjusting for the discovery covariates, and then combined by inverse variance meta-analysis using METAL, version 2011-03-25. Results for all 173,367 pairs reaching $P<5x10^{-8}$ (conventional genome-wide significance) are provided in **Supplementary Table 29**. We excluded three CpGs that shared no *cis*-SNPs across all data sets, and a further 9 CpGs because they had SNPs within their probe-binding sequence. For the remaining 175 CpG sites, the single SNP with the lowest P value for association with methylation was chosen as an instrumental variable (**Supplementary Table 8**). As mentioned above, to be an appropriate instrument, a SNP must not be directly associated with BMI (Y) but only through the respective CpG (X). For this purpose we removed six CpG-SNP pairs from the analysis because the corresponding SNPs remained associated with BMI after adjustment for the sentinel CpG (cg07136133, cg08548559, cg09152259, cg12484113, cg18120259, cg26403843). Statistical significance was inferred at $P<2.9x10^{-4}$ (corresponding to $P<0.05$ after Bonferroni correction for 175 tests).

To enable comparison with the observed effect of SNPs on BMI obtained from published data (see below), we next reassessed the relationship between DNA methylation and adiposity in linear models, using an inverse-normal transformation of BMI as the outcome variable to be consistent with the GIANT GWAS.[21] The associations between DNA methylation and inverse-normal transformed BMI were quantified in the LOLIPOP and KORA cohorts separately, followed by inverse variance meta-analysis using METAL, version 2011-03-25. We then calculated the predicted effect sizes and standard errors (β_{pred} and SE_{pred}) as follows:

$$\beta_{pred} = \beta_{CpG\sim SNP} \times \beta_{BMI\sim CpG}$$

$$SE_{pred} = \sqrt{SE^2_{CpG\sim SNP} \times SE^2_{BMI\sim CpG} + SE^2_{CpG\sim SNP} \times \beta^2_{BMI\sim CpG} + SE^2_{BMI\sim CpG} \times \beta^2_{CpG\sim SNP}}$$

The predicted effect sizes were compared against the observed effects of SNPs on BMI, whereby the latter were obtained from large published GWAS to increase power.[21] Statistical significance for individual SNPs was again inferred at $P<2.9x10^{-4}$. We used correlation analysis to examine the global relationship between predicted and observed effect on BMI for the SNPs influencing DNA methylation across the sentinel CpG sites.

DNA methylation as consequence of BMI (consequential analysis)

To test the hypothesis of DNA methylation being a consequence of BMI (whereby X=BMI, Y=DNA methylation), we followed a similar procedure as described above for the opposite direction with minor differences.

First, instead of using a single SNP as instrumental variable, we calculated a weighted genetic risk score (GRS) comprising SNPs reported to influence BMI. [21]Again, for the GRS to provide a valid instrument, the included SNPs must not show direct association with the CpG (Y) but only through BMI (X). For this purpose we removed three SNPs (rs12444979, rs10968576, rs7359397) which remained significantly associated at $P<8.4x10^{-6}$ (corresponding to P<0.05 after Bonferroni correction for the 187 x 32 tests performed) with at least one of the sentinel CpGs after adjusting for BMI. The final GRS was calculated as the sum of risk allele dosage of the remaining 29 SNPs previously reported to associate with BMI, weighted by the reported effect sizes.[21]

Second, the observed effects of GRS on DNA methylation were quantified using linear models as described above adjusted for the discovery covariates amongst participants of the KORA and LOLIPOP studies (**Supplementary Table 27**). Regression analysis was carried out in the KORA and LOLIPOP cohorts separately and results combined by inverse variance meta-analysis using METAL, version 2011-03-25.

DNA methylation in blood and adiposity in prospective population studies

We used data from the KORA (N=1,435 Europeans) and LOLIPOP (N=1513 Indian Asians) to examine the prospective, longitudinal association between DNA methylation at baseline and subsequent change in BMI during follow-up. We carried out linear regression with change in BMI during follow-up as response variable, and technically adjusted baseline methylation as the predictor variable, with age, sex, physical activity, smoking, alcohol intake, estimated white blood cell proportions and BMI at baseline, as well as follow-up time as additional covariates. Data were analysed in KORA and LOLIPOP separately, followed by inverse variance meta-analysis using METAL, version 2011-03-25.

We studied the longitudinal relationship between change in BMI and change in DNA methylation amongst 1,435 participants of the KORA S4/F4 cohort with methylation data available both at baseline and at the 7-year follow-up timepoint. To ensure comparability of methylation measurements from the two time points measured in two batches, methylation β-values were jointly adjusted for the first 20 PCs obtained from a PCA on the positive control probes, and residuals were subsequently used as adjusted methylation values. Linear models were used with change in BMI during follow-up as response variable, and

change in technically adjusted methylation as independent variable, including age, sex, physical activity, smoking, alcohol intake and estimated white blood cell proportions both at baseline and

DNA methylation in other tissues

DNA methylation in adipose tissue

We investigated whether the observed methylation markers in blood are representative of BMI-associated methylation changes in adipose tissue. We used a data set of 542 adipose tissue samples from the TwinsUK study to test association of the 187 identified methylation markers with BMI. The association of BMI with methylation was quantified using a linear mixed-effects model adjusting for chip, for bisulfite conversion level and bisulfite conversion efficiency, smoking state (3 categories: current, former and never smokers), alcohol intake (in g/d) and age, with zygosity and family as random effects.

We carried out sensitivity analyses to assess the potential contribution of cryptic structure arising from differences in cell composition of the adipose tissue samples. We used PCA to quantify latent structure in the adipose tissue methylation data, and included the top 5 components as covariates in the regression model.

We separately compared DNA methylation between paired samples of blood and subcutaneous adipose tissue (available for the same N=201 individuals, TwinsUK). Blood methylation values were first adjusted for age, chip and chip position, smoking state, alcohol intake, and estimated white blood cell subsets by taking the residuals from a linear model with these as covariates. Similarly, adipose tissue methylation values were adjusted for age, chip, bisulfite conversion level, bisulfite conversion efficiency, smoking state, alcohol intake, and the top 5 PCs from the adipose methylation data. Pearson's correlation was then determined between the adjusted methylation values.

Finally, we used genetic association to carry out causality analyses on the association between BMI and DNA methylation in adipose tissue, as described above for blood. We studied a subset of 325 adipose tissue samples from the Twins UK cohort with genotype data available. Regression analyses in adipose tissue between BMI, SNPs/GRS and CpGs were carried out using the R package lme4, and with smoking, alcohol intake, age, zygosity (random effect), family (random-effect), beadchip, bisulphite conversion batch and bisulphite conversion efficiency as covariates.

DNA methylation in isolated adipocytes

Subcutaneous adipose tissue samples were obtained intraoperatively in 24 morbidly obese individuals (BMI >40kg/m^2) undergoing laparoscopic bariatric surgery and 24 healthy controls (BMI <30kg/m^2) undergoing non-bariatric laparoscopic abdominal surgery. Participants were unrelated, between 18-60 years of age, from a multi-ethnic background, and free from type-2 diabetes. Controls were matched to cases by age, sex, and ethnicity. All participants gave informed consent (Ethics committee reference 13/LO/0477).

Adipose samples were processed immediately to isolate populations of primary human adipocyte cells using established protocols.[37-40] Polypropylene plastic ware was used to minimise adipocyte cell lysis. Adipose tissue samples were minced into 1-2mm^3 pieces and washed in Hank's buffered salt solution (HBSS), before digestion using type 1 collagenase (1mg/ml, Worthington) in a water bath at 37C shaking at 100rpm for ~45min. Digested samples were filtered through a 300 micron nylon mesh to remove debris, and the filtered solution centrifuged at low speed (500-g; 5min; 4 degrees), to leave four layers: top to bottom – (1) oil, (2) mature adipocytes, (3) supernatant, and (4) stromovascular pellet. After removal of the oil layer, the mature adipocyte layer was collected by pipette, washed in ~5x volume of HBSS and recentrifuged. After 3 washes the adipocyte cell suspension was collected for snap freezing and storage at -80C.

Genomic DNA and RNA were extracted from the isolated adipocytes using the Qiagen AllPrep DNA/RNA/miRNA Universal Kit according to manufacturer's protocol for lipid-rich samples. Methylation of genomic DNA was quantified using the Illumina HumanMethylation450 array in a single batch according to manufacturer's specifications. Raw methylation data were preprocessed using R, version 2.15. Bead intensity was retrieved using the R package *minfi*, version 1.6.0[34]. Marker intensities were quantile normalised for analysis. PCA of control probe intensities was performed to quantify cryptic structure in the data arising from technical factors. Logistic regression was used to examine the association of each CpG site with morbid obesity compared to normal weight, adjusting for age, sex and ethnicity, and the first 5 control probe PCs.

Cross-tissue methylation

For extended cross-tissue correlation analyses, publicly available data (GSE48472) were downloaded from the Gene Expression Omnibus (GEO) database.[25] Briefly, the dataset consists of 41 samples from six individuals of blood, liver, muscle, pancreas, subcutaneous fat, omentum and spleen analysed on the 450K methylation array. Data from the 187 CpG sites of interest were extracted and plotted using the *heatmap.2* function

in the R package *gplots* (version 2.17.0). Mean methylation levels for each CpG site across all samples within each tissue type were used to test for pairwise correlation between tissue types.

Functional genomics

Genomic annotation analyses

To test for functional enrichment of the 187 CpG sites associated with BMI, we used annotations of genomic context provided by Illumina,[14] and of histone modification ChIP peaks (H3K4me1, H3K4me3 and H3K27Ac, marks of open chromatin) and DNaseI Hypersensitivity Sites in 127 different cell types in the Roadmap and ENCODE (Release 9, UCSC) datasets. We mapped each probe on the Illumina 450k array background to the annotation categories and recorded overlap at each probe as a binary variable. To determine whether enrichment occurred more often than expected by chance, we generated 10,000 sets of 187 CpGs, each matched with the BMI sentinel CpGs for methylation mean (±2%) and standard deviation (±0.2%), but otherwise selected at random. For each epigenetic mark, we then calculated the number of overlapping sites amongst the 187 replicating markers (observed) and 10,000 permuted sets of 187 markers (expected). We calculated the fold enrichment as observed/mean(expected) and obtained an empirical P value from the distribution of expected.

Gene expression studies

Transcriptome-wide measurements of gene expression in blood along with measurements of DNA methylation from the same blood sample were available for participants of both the KORA F4 (N=703) and LOLIPOP (N=1,082, 907 Indian Asians, 175 Europeans) studies (**Supplementary Table 15**). KORA samples were analysed with the Illumina HumanHT-12 v3 BeadChip array. Blood sample collection and RNA isolation and preparation have been described in detail.[41,42] Gene expression data were quantile normalized and log2 transformed using the R package *lumi*, version 2.8.0, from Bioconductor in R, version 2.14.2. In LOLIPOP, gene expression analysis was performed with the Illumina HumanHT-12 v4 BeadChip array according to manufacturer's protocol. Background correction (using negative controls), quantile normalisation and log2 transformation was performed using the R-package *limma* (function *neqc*).

To examine associations of DNA methylation with gene expression we carried out linear regression with log2 transformed gene expression as the response variable and methylation βvalues as independent variable. In KORA, the model was adjusted for the

discovery covariates and technical covariates related to the expression measurement (RNA integrity number, RNA amplification plate, sample storage time). In LOLIPOP, the model was adjusted for age, sex, methylation control probe PCs and technical covariates related to the expression measurement (RNA integrity number, RNA extraction batch, RNA conversion batch, scanning batch, array and array position). Results were analysed in KORA, LOLIPOP Indian Asians and LOLIPOP Europeans separately, then combined by inverse-variance meta-analysis using METAL (version 2011-03.25). Statistical significance was inferred at $P<9.0x10^{-6}$ (i.e. $P<0.05$ after Bonferroni correction for 5,551 CpG-expression pairs).

To assess whether the 187 sentinel CpGs were enriched for association with gene expression, we used the same testing concept as described above based on constructing a null distribution from 10,000 randomly selected matched sets of 187 CpGs. For each permuted set we determined the number of significantly associated expression probes in *cis* ($P<9.0x10^{-6}$) as described above and compare the resulting distribution with the observed number of gene expression associations for the 187 sentinel CpG sites to calculate an empirical P value.

Finally, we examined the association between DNA methylation and gene expression in TwinsUK adipose tissue samples (N=499) for the 44 methylation-expression pairs that were significant in blood. Expression values were adjusted for age and chip using a linear model. The association of methylation and expression was then determined in linear mixed-effects models with adjusted expression as response and methylation as the independent variable, adjusting for age, chip, bisulfite conversion level and bisulfite conversion efficiency, with zygosity and family as random effects. After QC filtering of methylation and expression data, results were available for 36 methylation-expression pairs.

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.[43] 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.

Candidate genes and gene-set enrichment analyses

The standard Illumina annotation does not identify a gene for all CpG sites on the 450K microarray. We therefore identified candidate genes based on the following criteria: i. Proximity: gene nearest to the CpG site (N=187 genes) and ii. Gene expression: all local genes (up to ±500 kb) with expression associated with the marker at P<0.05 after Bonferroni correction for 5,551 tests (N=38 genes). This resulted in a list of 210 unique genes (**Supplementary Table 20**).

Gene annotations were downloaded from ensembl (grch37.ensembl.org) using R package *biomaRt*, version 2.18.0, from Bioconductor, and overlapped with the cg positions as annotated in the Illumina annotation using the R package *GenomicRanges*, version 1.14.4, from Bioconductor. We downloaded curated pathway information (c2.all.v5.0.symbols.gmt) from the GSEA MSigDB platform (http://www.broadinstitute.org/gsea/msigdb), resulting in 1,135 pathways, to investigate enrichment of the set of candidate genes against curated pathway sets (BIOCARTA, KEGG, REACTOME).[39-41] An enrichment P value was calculated empirically based on permutation testing, using the Benjamini-Hochberg (false-discovery-rate) procedure. As a sensitivity analysis the gene-set enrichment analysis was repeated using the genes annotated by Illumina, and using more permissive proximity criteria (**Supplementary Table 22**).

Clinical implications

DNA methylation and metabolic traits

We investigated the association between the 187 sentinel methylation markers and metabolic disturbances associated with adiposity amongst participants of the KORA (N=1,697) and LOLIPOP (N=2,462) studies with available measurements of the following BMI-related clinical traits: LDL cholesterol, HDL cholesterol, total cholesterol, fasting triglycerides, fasting glucose, fasting insulin, HbA1c, systolic and diastolic blood pressure, C-reactive protein, weight, height and waist-hip ratio. Linear models were used with trait as response and methylation as independent variable, adjusting for the discovery covariates. Results from KORA and LOLIPOP studies were analysed separately, then combined by inverse variance meta-analysis using METAL, version 2011-03-25. Associations were

considered significant at $P<2.1x10^{-5}$ (corresponding to P<0.05 after Bonferroni correction for 187 x 13 tests).

To investigate potential causal relationships between the methylation markers and BMI-related clinical traits, we performed causality analyses as described above for the primary phenotype (BMI). For each clinical trait, GWAS datasets of the most comprehensive meta-analyses published to date with access to genome-wide association results were retrieved (**Supplementary Table 24**), to provide SNPs influencing trait. SNPs associated with multiple traits were assigned to the most strongly associated trait (lowest P value). Clinical traits were transformed as described in the respective GWAS. Genetic risk scores were calculated as described above for BMI, after removal of SNPs with direct genomic effects (SNPs that remain associated with the sentinel CpG after adjustment for the trait). Regression analyses were carried out in the KORA F4 and LOLIPOP cohorts separately and results were combined by inverse variance meta-analysis using METAL, version 2011-03-25.

Association with incident T2D

We tested the association of DNA methylation at the 187 identified CpG sites with incident T2D amongst participants of the LOLIPOP study. All participants (N=2,664) were free from T2D at the time of measurement of DNA methylation; incident T2D (N=1,074) was defined as either new physician diagnosis, or HbA1c≥6.5%. Associations with T2D were evaluated by logistic regression adjusted for the discovery covariates. We initially tested the association in single marker tests, then in a fully saturated model comprising all 187 markers to identify independent effects.

To combine information across loci, we calculated a weighted methylation risk score (MRS) as the sum of the standardised methylation values at each marker that reached nominal significance (P<0.05) in the fully saturated multivariate model, weighted by marker-specific effect size. We then tested the association of the MRS with incident T2D using logistic regression, before and after adjustment for traditional T2D risk factors (BMI, WHR, glucose, HbA1c).

Replication testing of the association of MRS with T2D was carried out in a nested case-control study within the KORA S3/S4 comprising 200 subjects with newly diagnosed T2D and 200 control matched for age (±2 years), sex, cohort and observation time until diagnosis of diabetes. Data were analysed using conditional logistic regression using the function *clogit* of the R package *survival*, version 2.37.4.

Software

Unless stated otherwise, all calculations were performed using R, version 3.0.1. For all meta-analyses, METAL, version 2011-03-25, was used. Custom R code is available upon request.

Availability of data

Summary statistics from the epigenome-wide association study can be accessed from the European Genome-Phenome Archive (accession number: *on publication*). KORA methylation data are available upon request through the application tool KORA.PASST (http:/epi.helmholtz-muenchen.de); LOLIPOP data are available from the Gene Expression Omnibus (Ref: GSE55763); EPICOR data are deposited in the HuGeF repository (http://www.hugef-torino.org) and are available on request.

References

1 Wang, Y. C., McPherson, K., Marsh, T., Gortmaker, S. L. & Brown, M. Health and economic burden of the projected obesity trends in the USA and the UK. *Lancet* **378**, 815-825, doi:10.1016/S0140-6736(11)60814-3 (2011).

2 Ng, M. *et al.* Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 766-781, doi:10.1016/S0140-6736(14)60460-8 (2014).

3 Guh, D. P. *et al.* The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC public health* **9**, 88, doi:10.1186/1471-2458-9-88 (2009).

4 Fall, T. *et al.* The role of adiposity in cardiometabolic traits: a Mendelian randomization analysis. *PLoS medicine* **10**, e1001474, doi:10.1371/journal.pmed.1001474 (2013).

5 Stefan, N., Haring, H. U., Hu, F. B. & Schulze, M. B. Metabolically healthy obesity: epidemiology, mechanisms, and clinical implications. *The lancet. Diabetes & endocrinology* **1**, 152-162, doi:10.1016/S2213-8587(13)70062-7 (2013).

6 Dick, K. J. *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* **383**, 1990-1998, doi:10.1016/S0140-6736(13)62674-4 (2014).

7 Feinberg, A. P. *et al.* Personalized epigenomic signatures that are stable over time and covary with body mass index. *Science translational medicine* **2**, 49ra67, doi:10.1126/scitranslmed.3001262 (2010).

8 Milagro, F. I. *et al.* A dual epigenomic approach for the search of obesity biomarkers: DNA methylation in relation to diet-induced weight loss. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **25**, 1378-1389, doi:10.1096/fj.10-170365 (2011).

9 Xu, X. *et al.* A genome-wide methylation study on obesity: differential variability and differential methylation. *Epigenetics : official journal of the DNA Methylation Society* **8**, 522-533, doi:10.4161/epi.24506 (2013).

10 Demerath, E. W. *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Human molecular genetics* **24**, 4464-4479, doi:10.1093/hmg/ddv161 (2015).

11 Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nature biotechnology* **28**, 1057-1068, doi:10.1038/nbt.1685 (2010).

12 Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nature reviews. Genetics* **12**, 529-541, doi:10.1038/nrg3000 (2011).

13 Danaei, G. *et al.* National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet* **378**, 31-40, doi:10.1016/S0140-6736(11)60679-X (2011).

14 Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288-295, doi:10.1016/j.ygeno.2011.07.007 (2011).

15 Lehne, B. *et al.* A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biology* **16**, 37 (2015).

16 Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC bioinformatics* **13**, 86, doi:10.1186/1471-2105-13-86 (2012).

17 Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theoretical population biology* **60**, 155-166, doi:10.1006/tpbi.2001.1542 (2001).
18 Relton, C. L. & Davey Smith, G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. *International journal of epidemiology* **41**, 161-176, doi:10.1093/ije/dyr233 (2012).
19 Bochud, M. & Rousson, V. Usefulness of Mendelian randomization in observational epidemiology. *International journal of environmental research and public health* **7**, 711-728, doi:10.3390/ijerph7030711 (2010).
20 Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology* **32**, 1-22 (2003).
21 Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature genetics* **42**, 937-948, doi:10.1038/ng.686 (2010).
22 Ren, D. *et al.* Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. *The Journal of clinical investigation* **117**, 397-406, doi:10.1172/JCI29417 (2007).
23 Johansson, L. E. *et al.* Differential gene expression in adipose tissue from obese human subjects during weight loss and weight maintenance. *The American journal of clinical nutrition* **96**, 196-207, doi:10.3945/ajcn.111.020578 (2012).
24 Aron-Wisnewsky, J. *et al.* Effect of bariatric surgery-induced weight loss on SR-BI-, ABCG1-, and ABCA1-mediated cellular cholesterol efflux in obese women. *The Journal of clinical endocrinology and metabolism* **96**, 1151-1159, doi:10.1210/jc.2010-2378 (2011).
25 Slieker, R. C. *et al.* Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics & chromatin* **6**, 26, doi:10.1186/1756-8935-6-26 (2013).
26 Ram, O. *et al.* Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells. *Cell* **147**, 1628-1639, doi:10.1016/j.cell.2011.09.057 (2011).
27 Liu, Y. *et al.* Methylomics of gene expression in human monocytes. *Human molecular genetics* **22**, 5065-5074, doi:10.1093/hmg/ddt356 (2013).
28 Karin, M. & Ben-Neriah, Y. Phosphorylation meets ubiquitination: the control of NF-[kappa]B activity. *Annual review of immunology* **18**, 621-663, doi:10.1146/annurev.immunol.18.1.621 (2000).
29 Pfeifferm, L. *et al.* DNA Methylation of Lipid-Related Genes Affects Blood Lipid Levels. *Circulation. Cardiovascular genetics*, doi:10.1161/CIRCGENETICS.114.000804 (2015).
30 Hidalgo, B. *et al.* Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. *Diabetes* **63**, 801-807, doi:10.2337/db13-1100 (2014).
31 Tarling, E. J. Expanding roles of ABCG1 and sterol transport. *Current opinion in lipidology* **24**, 138-146, doi:10.1097/MOL.0b013e32835da122 (2013).
32 Sturek, J. M. *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* **120**, 2575-2589, doi:10.1172/JCI41280 (2010).
33 Kruit, J. K. *et al.* Loss of both ABCA1 and ABCG1 results in increased disturbances in islet sterol homeostasis, inflammation, and impaired beta-cell function. *Diabetes* **61**, 659-664, doi:10.2337/db11-1341 (2012).

34 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).
35 Burgess, S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *International journal of epidemiology* **43**, 922-929, doi:10.1093/ije/dyu005 (2014).
36 Freeman, G., Cowling, B. J. & Schooling, C. M. Power and sample size calculations for Mendelian randomization studies using one genetic instrument. *International journal of epidemiology* **42**, 1157-1163, doi:10.1093/ije/dyt110 (2013).
37 Spalding, K. L. *et al.* Dynamics of fat cell turnover in humans. *Nature* **453**, 783-787, doi:10.1038/nature06902 (2008).
38 Stahl, A., Evans, J. G., Pattel, S., Hirsch, D. & Lodish, H. F. Insulin causes fatty acid transport protein translocation and enhanced fatty acid uptake in adipocytes. *Developmental cell* **2**, 477-488 (2002).
39 Carswell, K. A., Lee, M. J. & Fried, S. K. Culture of isolated human adipocytes and isolated adipose tissue. *Methods in molecular biology* **806**, 203-214, doi:10.1007/978-1-61779-367-7_14 (2012).
40 Multhaup, M. L. *et al.* Mouse-human experimental epigenetic analysis unmasks dietary targets and genetic liability for diabetic phenotypes. *Cell metabolism* **21**, 138-149, doi:10.1016/j.cmet.2014.12.014 (2015).
41 Schurmann, C. *et al.* Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. *PloS one* **7**, e50938, doi:10.1371/journal.pone.0050938 (2012).
42 Doring, A. *et al.* SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nature genetics* **40**, 430-436, doi:10.1038/ng.107 (2008).
43 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).

Figure 1.

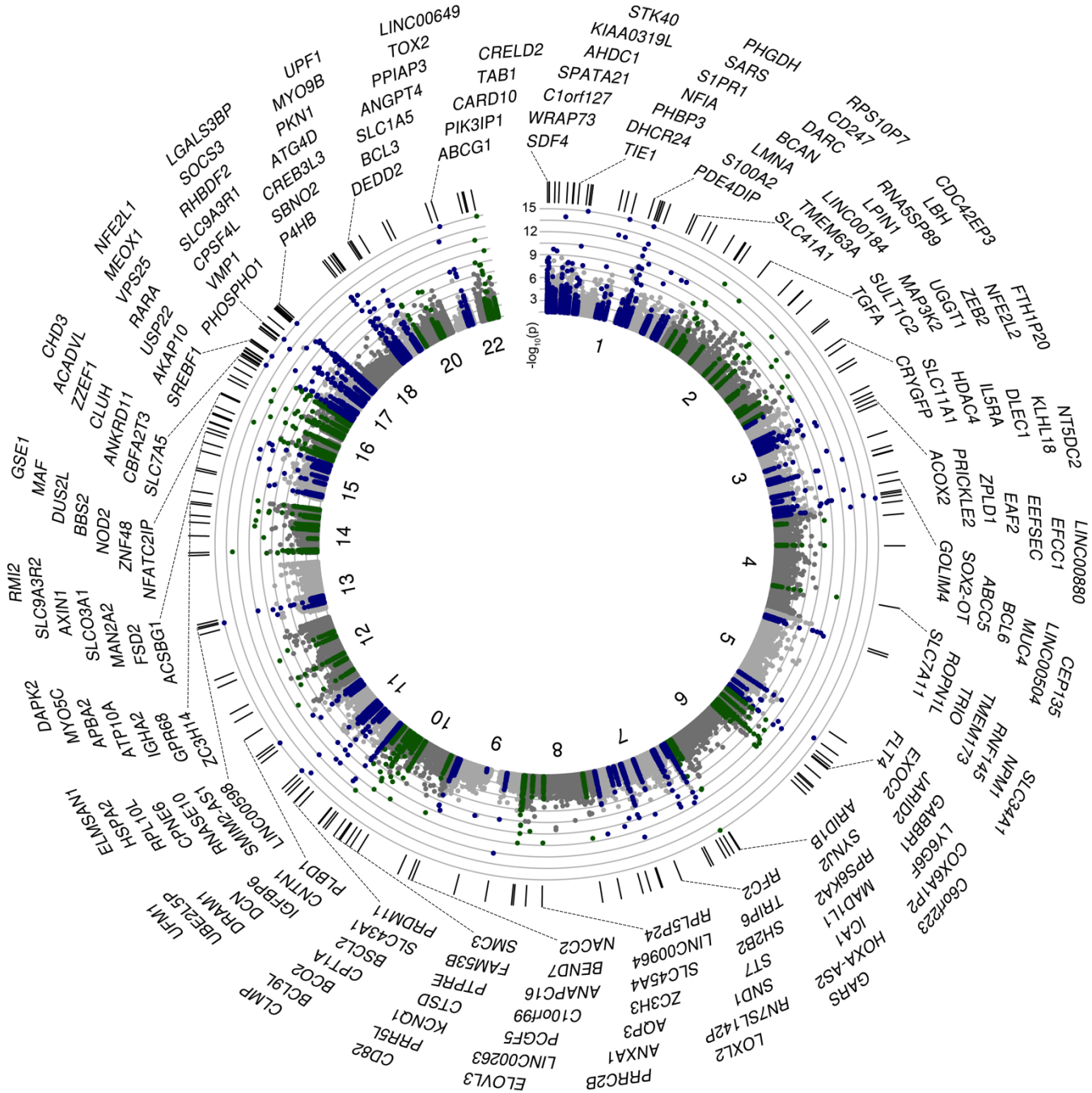
-log10(p)
15
12
9
6
3
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
20
22
STK40
KIAA0319L
AHDC1
SPATA21
C1orf127
WRAP73
SDF4
TIE1
PHGDH
SARS
S1PR1
NFIA
PHBP3
DHCR24
RPS10P7
CD247
DARC
BCAN
LMNA
S100A2
PDE4DIP
CDC42EP3
LBH
RNA5SP89
LPIN1
LINC00184
TMEM63A
SLC41A1
FTH1P20
NFE2L2
ZEB2
UGGT1
MAP3K2
SULT1C2
TGFA
NT5DC2
KLHL18
DLEC1
IL5RA
HDAC4
SLC11A1
CRYGFP
LINC00880
EFCC1
EEFSEC
EAF2
ZPLD1
PRICKLE2
ACOX2
CEP135
LINC00504
MUC4
BCL6
ABCC5
SOX2-OT
GOLM4
RNF145
TMEM173
TRIO
ROPN1L
SLC7A11
SLC34A1
NAWL
FLT4
EXOC2
JARID2
GABBR1
LY6G6F
COX6A1P2
C6orf223
ARID1B
SYNJ2
RPS6KA2
MAD1L1
ICA1
HOXA-AS2
GARS
LOXL2
RN7SL142P
SND1
STT
SH2B2
TRIP6
RFC2
RPL5P24
LINC00964
SLC45A4
ZC3H3
AOP3
ANXA1
PRRC2B
ELOVL3
LINC00263
PCGF5
C10orf99
ANAPC16
BEND7
NAC2
SMC3
FAM53B
PTPRE
CTSD
KCNQ1
PRR5L
CD82
CLMP
BCL9L
BCO2
CPT1A
BSCL2
SLC43A1
PRDM11
UFM1
UBE2L5P
DRAM1
DCN
IGFBP6
CNTN1
PLBD1
LINC00598
SMIM2-AS1
RNASET10
CPNE9
RPL10L
HSPA2
ELMSAN1
ZC3H14
GPR68
IGHA2
ATP10A
APBA2
MYO5C
DAPK2
ACSBG1
FSD2
MAN2A2
SLCO3A1
AXIN1
SLC9A3R2
RMI2
NFATC2IP
ZNF48
NOD2
BBS2
DUS2L
MAF
GSE1
SLC7A5
CBFA2T3
ANKRD11
CLUH
ZZEF1
ACADVL
CHD3
SREBF1
AKAP10
USP22
RARA
VPS25
MEOX1
NFE2L1
PHOSPHO1
VMP1
CPSF4L
SLC9A3R1
RHBDF2
SOCS3
LGALS3BP
P4HB
SBNO2
CREB3L3
ATG4D
PKN1
MYO9B
UPF1
DEDD2
BCL3
SLC1A5
ANGPT4
PPIAP3
TOX2
LINC00649
ABCG1
PIK3IP1
CARD10
TAB1
CRELD2

Figure 2.

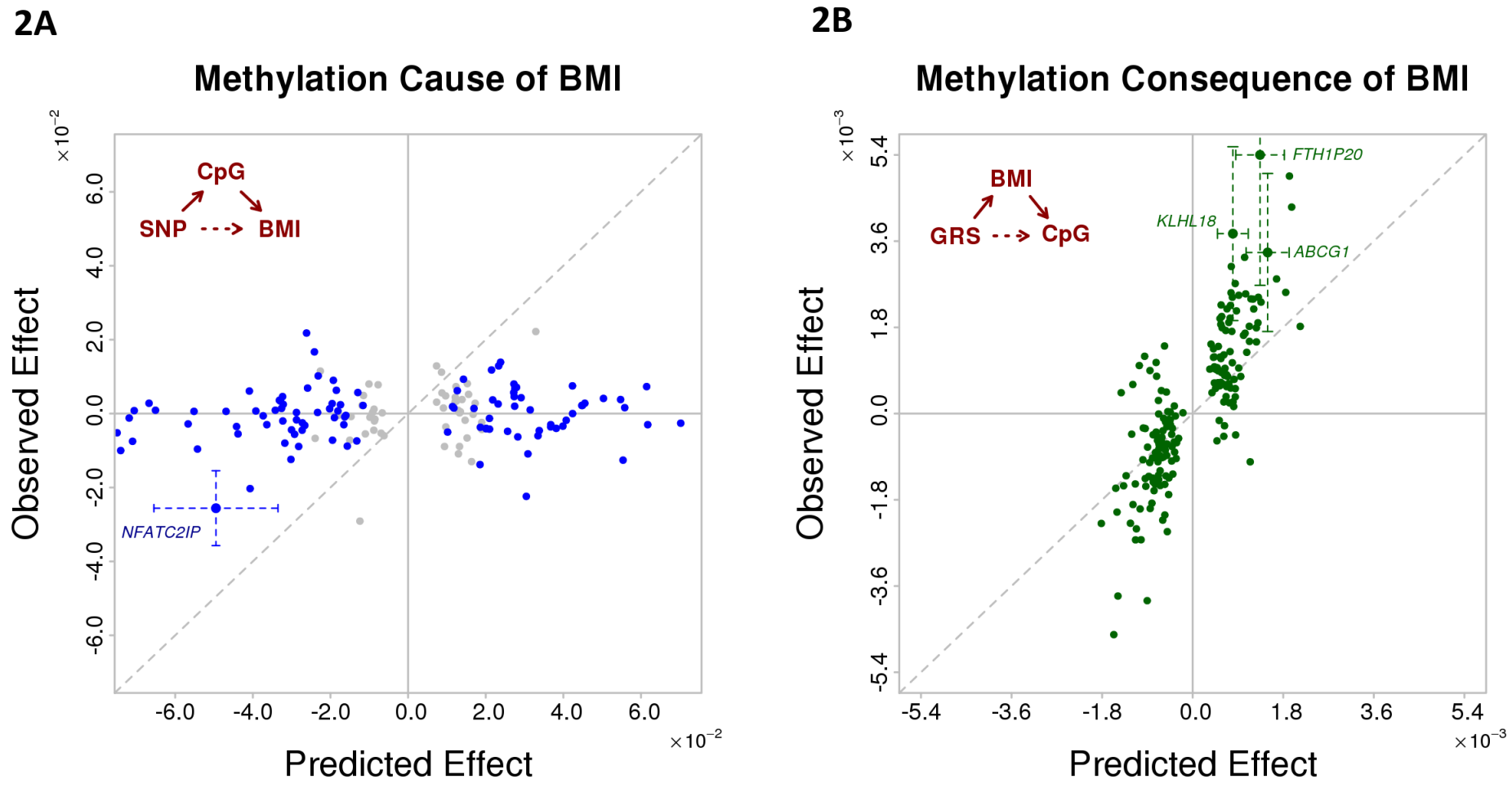
2A
Methylation Cause of BMI
CpG
SNP ---> BMI
NFATC2IP
Observed Effect
Predicted Effect
×10⁻²
2B
Methylation Consequence of BMI
BMI
GRS ---> CpG
FTH1P20
KLHL18
ABCG1
Observed Effect
Predicted Effect
×10⁻³

Figure 3.

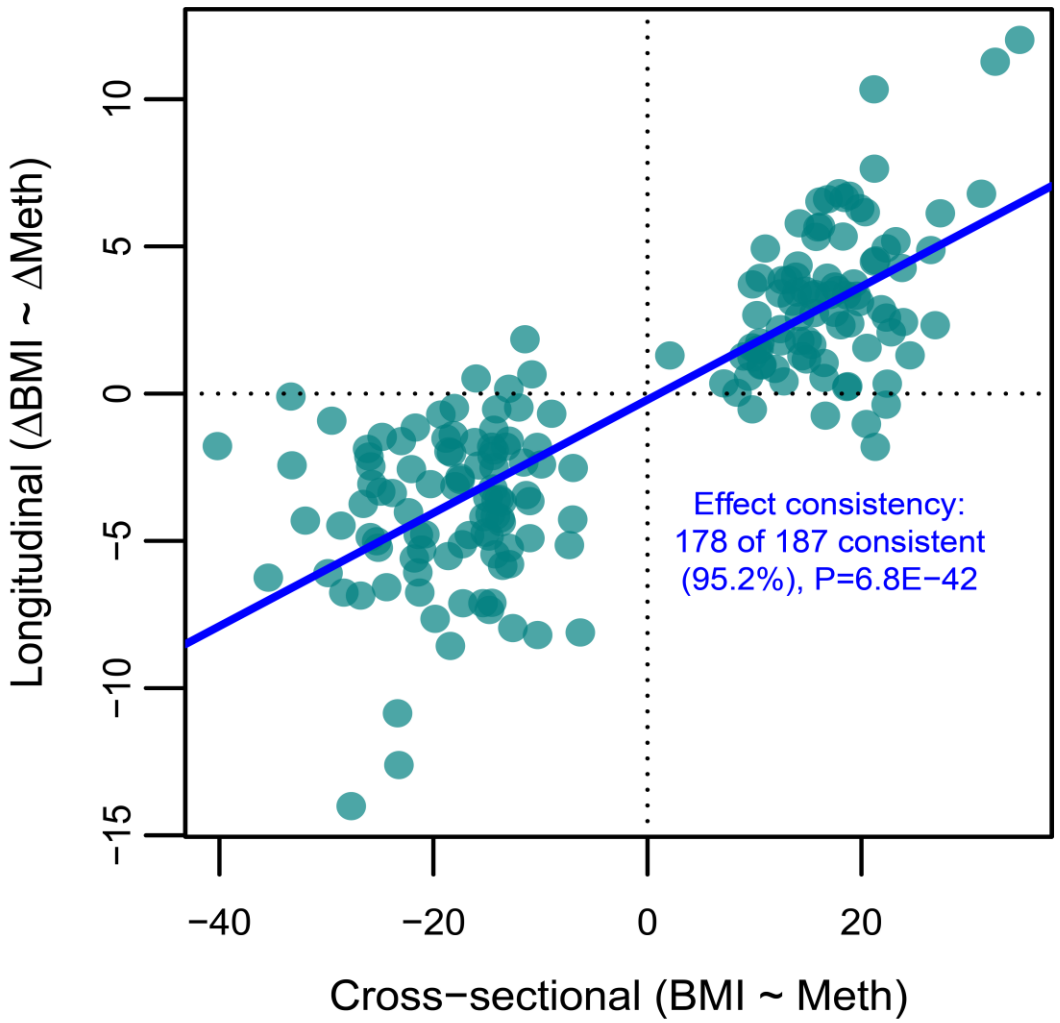
Longitudinal (ΔBMI ~ ΔMeth)
10
5
0
−5
−10
−15
−40
−20
0
20
Cross−sectional (BMI ~ Meth)
Effect consistency:
178 of 187 consistent
(95.2%), P=6.8E−42

Figure 4

	Controls/Cases	P	P-trend
Normal			
Q1	144/29	NA	3.85E-7
Q2	141/32	4.97E-1	
Q3	130/43	2.61E-2	
Q4	106/69	1.89E-7	
Overweight			
Q1	129/27	9.5E-1	5.66E-19
Q2	185/78	7.7E-4	
Q3	169/115	9.0E-8	
Q4	301/321	4.0E-16	
Obese			
Q1	27/17	2.5E-3	4.19E-7
Q2	50/28	5.2E-4	
Q3	59/61	5.1E-10	
Q4	149/255	7.9E-22