OPEN ACCESS

University of BRISTOL

Peer reviewed version

Link to published version (if available):
10.1038/nature20784

Link to publication record in Explore Bristol Research
PDF-document

## University of Bristol - Explore Bristol Research
### General rights

**Extended Data**

Manuscript:   MS 2015-03-03573

DNA methylation and the adverse metabolic outcomes of adiposity.

Simone Wahl et al.

**Extended Data Figure 1**. Study design.

*Epigenome-wide association and replication testing* was performed in order to identify methylation sites associated with adiposity. In the discovery step, four large cohorts were included with Illumina 450k DNA methylation data available, which were preprocessed and quality controlled according to a harmonized protocol. Epigenome-wide association was performed in every single study with BMI as response variable and methylation β-value as independent variable, adjusting for covariates as described in the Online Methods. At a genome-wide significance level of $P<1\times10^{-7}$, 278 methylation sites from 207 regions were identified. In the replication step, 187 of these replicated in independent samples. *Genetic association and causality* analyses were used in order to investigate whether the identified methylation signals underlie the development of adiposity or are the consequence of adiposity. The findings were supported with the help of *longitudinal analyses.* The *cross-tissue analyses* represent a first step towards extending our observations in blood to metabolically relevant tissues. The *functional genomics and gene expression* analyses help to link the observed methylation associations to transcriptional outcomes, while the *gene-set enrichment analysis* provides a way to summarize the potentially affected metabolic pathways. Finally, we study the relationships of methylation to adiposity related *metabolic traits and type 2 diabetes* to address the clinical relevance of our findings.
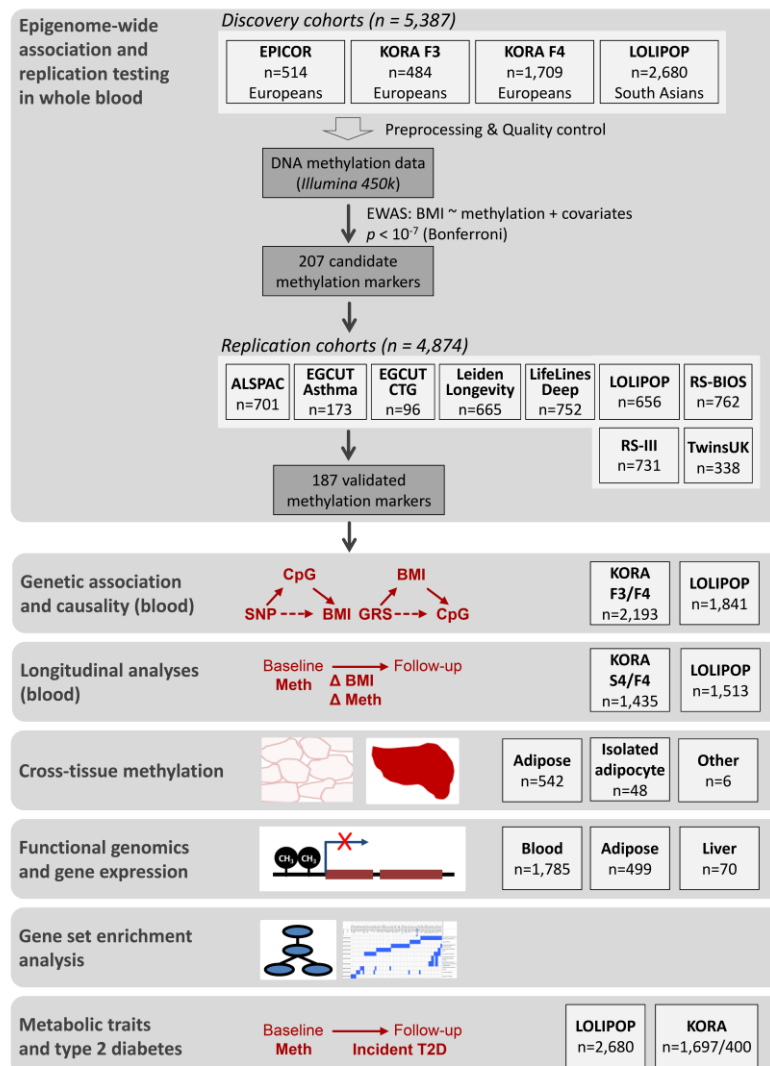
**Extended Data Figure 2.** Distribution of methylation values at the 187 sentinel CpG sites compared to the ~473K CpG sites assayed by the Illumina Infinium 450K Human Methylation array. The 187 identified methylation-BMI associations are strongly enriched for CpG sites with intermediate levels of methylation, consistent with the presence of epigenetic heterogeneity at these loci in blood (157/187 sites with 20-80% methylation, a 3.0-fold enrichment compared to microarray background, P=1.4x10[-22] Fisher's test).
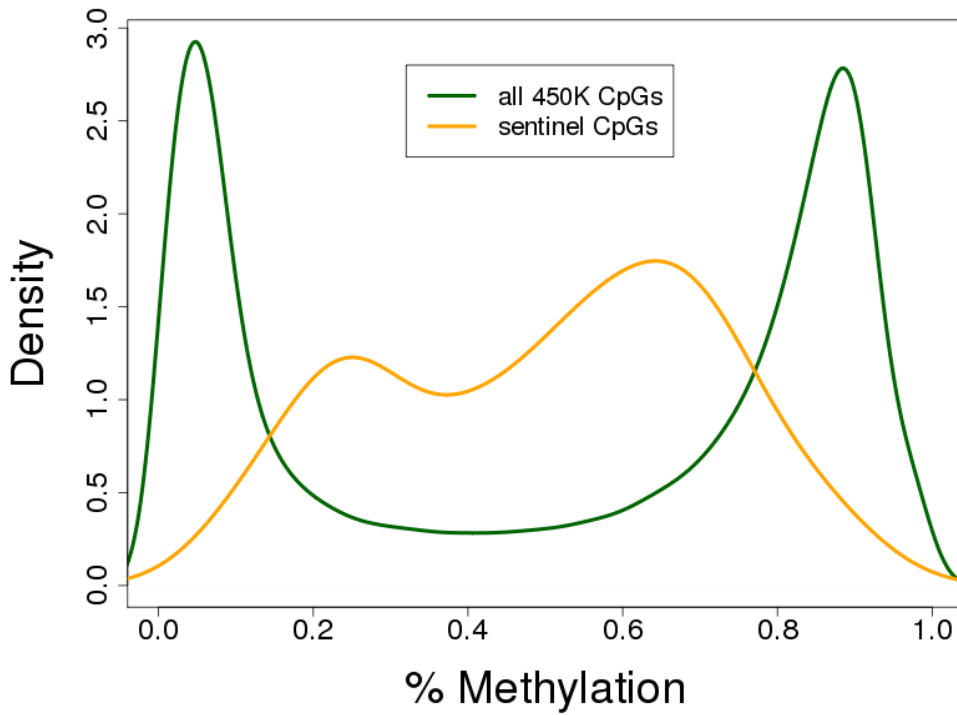
**Extended Data Figure 3.** DNA methylation at the sentinel CpG sites in whole blood and in 4 isolated cell subsets (Monocytes, Neutrophils, CD4+, CD8+) from 60 individuals (30 obese cases, and 30 normal weight controls) by Illumina MethylationEPIC array, which quantifies 179 of the 187 sentinel markers. Results are shown as a heatmap, coded by methylation value (hypomethylation <0.2; intermediate methylation 0.2-0.8, hypermethylation >0.8). Results show the presence of intermediate methylation (and hence epigenetic heterogeneity) at the majority of loci, and in the majority of cell types, in both cases and controls.
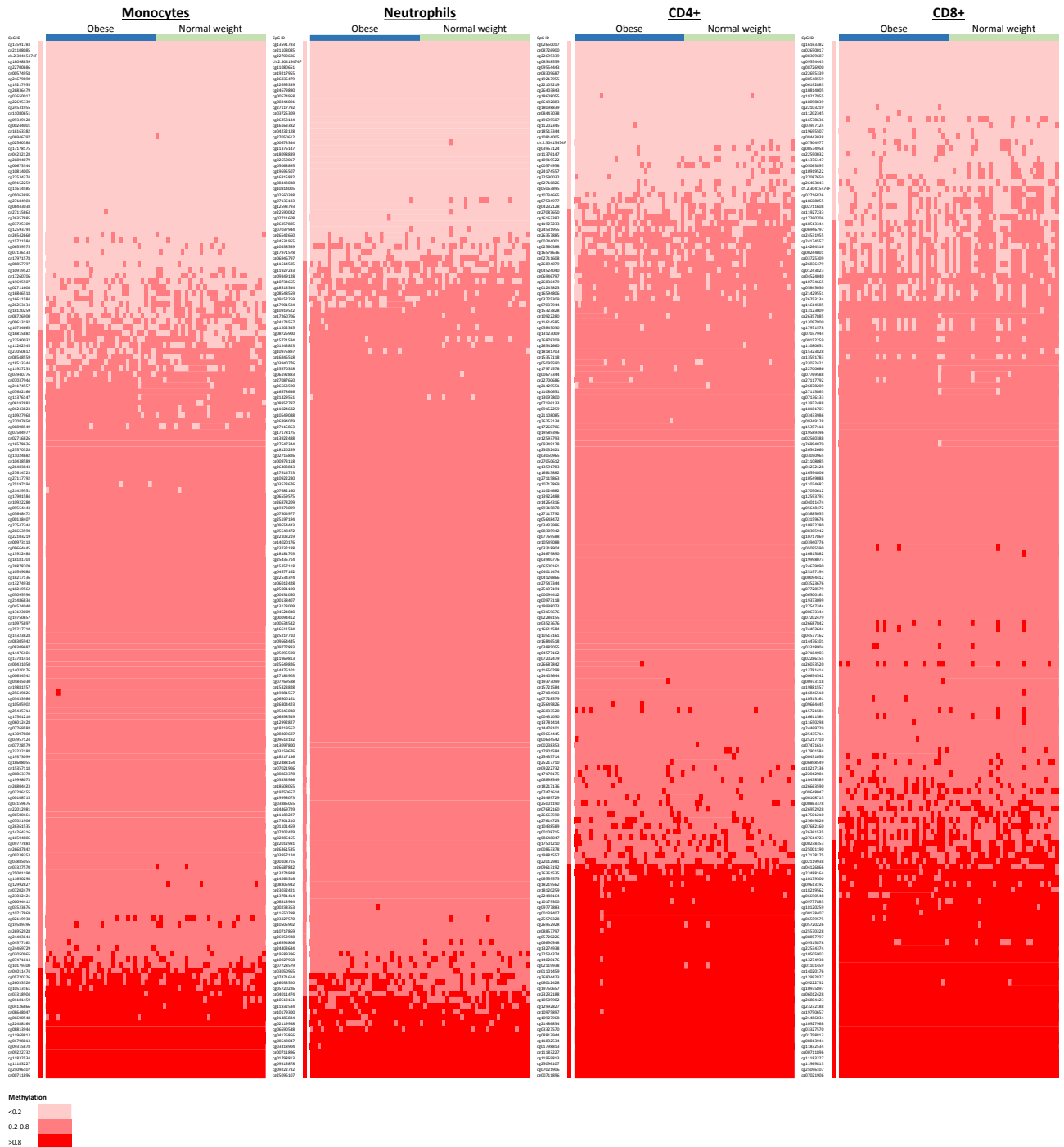
**Extended Data Figure 4.** Association of DNA methylation with obesity in the 4 cell subsets studied, based on quantification of methylation at 179 of the sentinel methylation markers amongst 30 obese cases and 30 normal weight controls. Results are presented as QQ plots of the observed association test statistics in each of the isolated cell subsets.
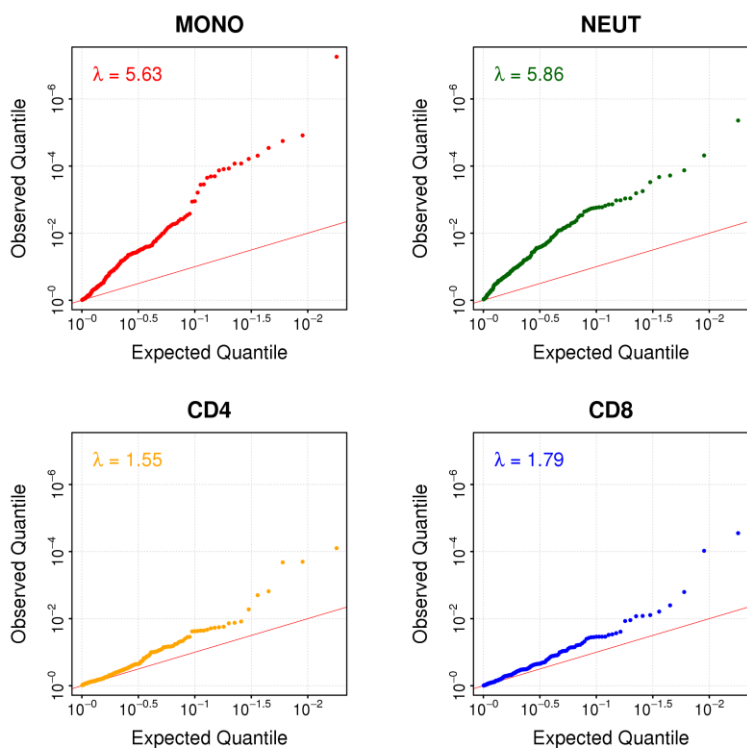


**Extended Data Figure 5.** Comparison of effect sizes between isolated white cell subsets. Results are presented as the difference in methylation between obese cases and normal weight controls (Methylation in cases – methylation in controls, in absolute terms on % scale) in the respective isolated white cell subset (**y axis**) compared to the average case-control difference across all 4 cell subsets studied (**x axis**).
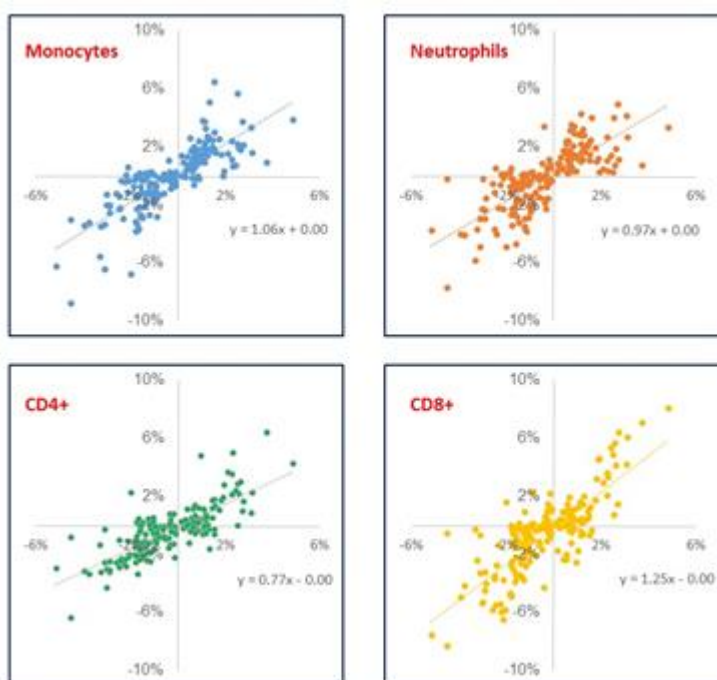
**Extended Data Figure 6**. Mean methylation levels at the 187 sentinel methylation markers associated with BMI, across 7 tissue types (blood: N=6; liver: N=5, muscle: N=6, omentum: N=6, pancreas: N=4, subcutaneous (SC) fat: N=6, spleen: N=3). The lower panel displays pairwise scatterplots (trendline in red), while the upper panel shows the Pearson correlation coefficient and P values.
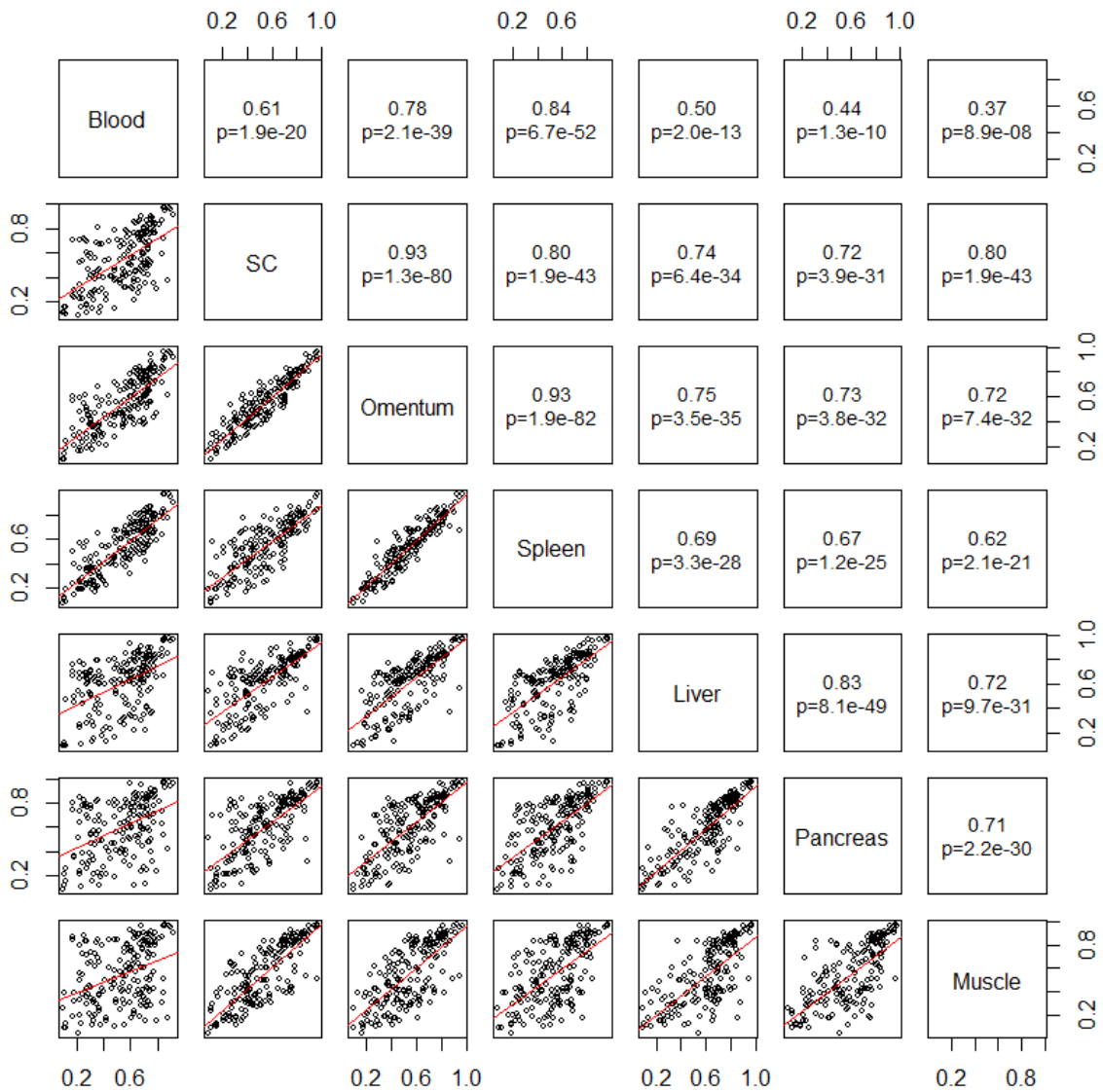
**Extended Data Figure 7.** Causality analysis in adipose tissue to investigate the potential relationships between BMI and DNA methylation. Left panel: Causality analysis in adipose tissue investigating whether DNA 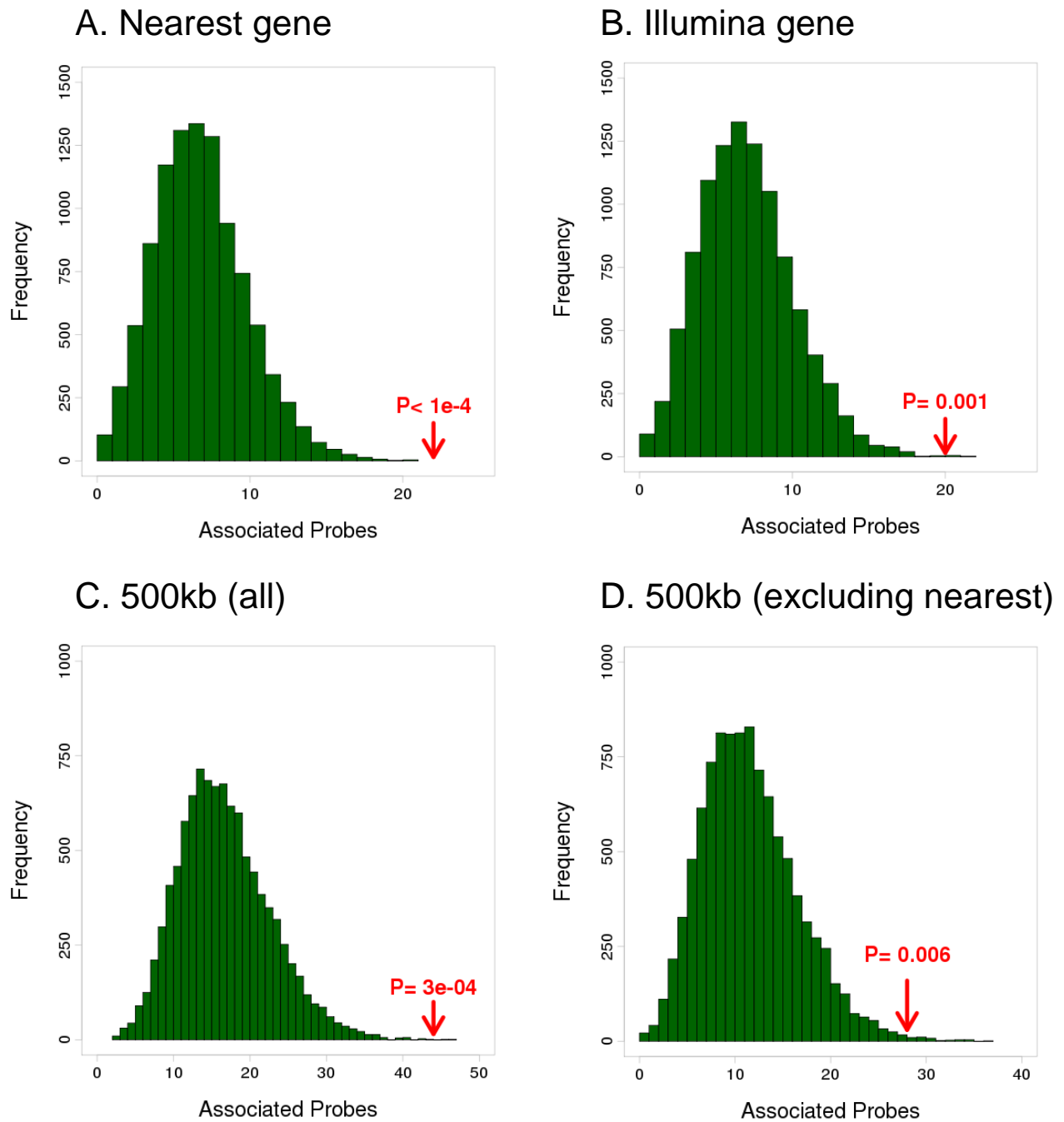methylation at sentinel CpG sites influences BMI. Units are change in BMI per copy of effect allele. For each sentinel CpG site we determined i. the effect of a previously identified cis-SNP on BMI predicted via methylation (x-axis), ii. the directly observed effect of SNP on BMI (y-axis). No CpG passed multiple testing correction for all three comparisons. Overall there was little relationship between the effects of SNPs on BMI predicted via methylation and the directly observed effect (R=-0.04 P=0.58). Right panel: Causality analysis in adipose tissue investigating whether DNA methylation at sentinel CpG sites is the consequence of BMI. Units are change in methylation per unit change in weighted genetic risk score (GRS). We identified SNPs reported to influence BMI in GWAS meta-analysis, and calculated a weighted GRS. 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 methylation (y-axis). No CpG passed multiple testing correction for all three comparisons. The overall correlation between observed and predicted effects (R=0.73; P=1.6 x 10$^{-32}$) replicates our findings in blood that methylation at the majority of CpG-sites is consequential to BMI.
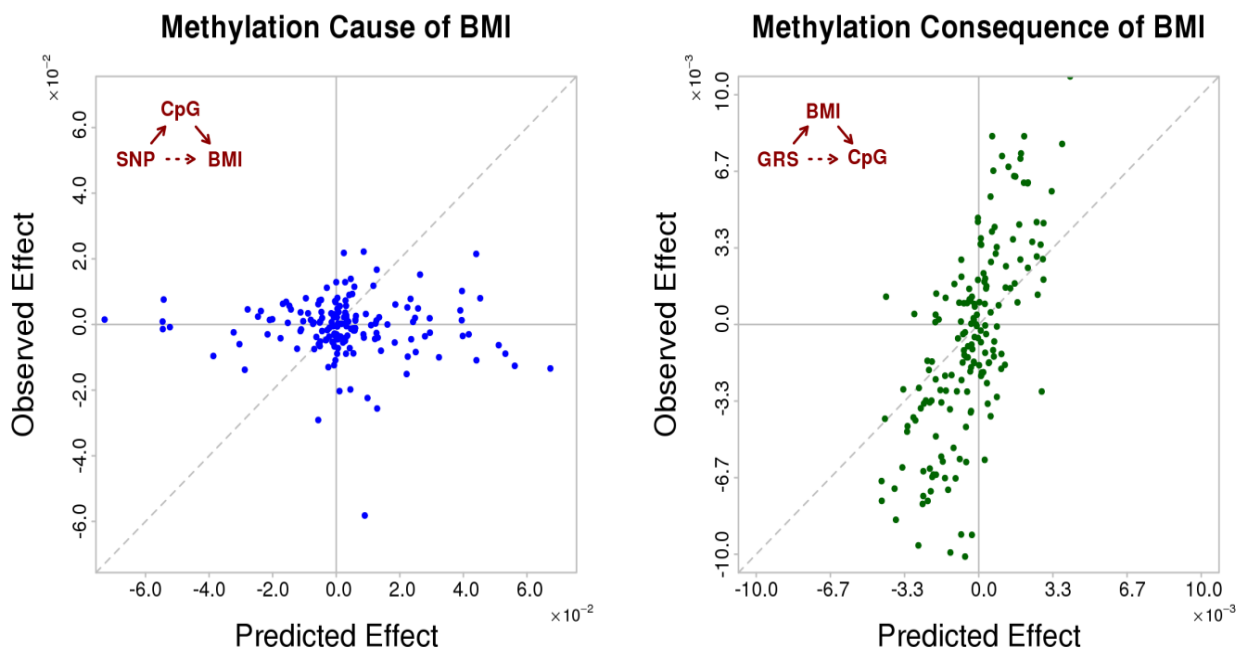
**Extended Data Figure 8.** The 187 sentinel CpGs are enriched for association with gene-expression in *cis* in blood. To derive an expectation under the null-hypothesis we generated 10,000 sets of matched CpGs (matched for mean methylation and for SD of methylation, see Online Methods), and tested their association with expression of A) the nearest gene, B) the gene allocated to the CpG by the Illumina annotation, C) all genes within a 500 kb distance and D) all genes within a 500 kb distance excluding the nearest gene. We observe significantly more expression-probes associated with the sentinel markers (red arrow) in blood compared to the 10,000 permuted sets (green bars).



A. Nearest gene — P< 1e-4

B. Illumina gene — P= 0.001

C. 500kb (all) — P= 3e-04

D. 500kb (excluding nearest) — P= 0.006

**Extended Data Figure 9**. Summary statistics for the causality analyses investigating the relationship between DNA methylation in blood and metabolic disturbances.

**Panel A.** DNA methylation in blood as a potential determinant of the metabolic disturbances associated with adiposity (causal analysis). For each of the sentinel CpG sites we identified the cis-SNP (1Mb) most closely associated with DNA methylation levels. For each of the SNPs we then determined i. the effect of SNP on phenotype predicted via methylation, ii. the directly observed effect of SNP on phenotype. Results are presented as the $R^2$ between phenotype specific observed and predicted effects across the 187 CpG sites, calculated using linear regression.

**Panel B.** DNA methylation in blood as a potential consequence of the metabolic disturbances associated with adiposity (consequential analysis). We identified the SNPs reported to influence each phenotypic trait (using the most recent GWAS meta-analysis, Supplementary Table 24), and calculated phenotype specific weighted genetic risk scores (GRS). For each of the CpG sites, and each of the phenotypes, we then determined i. the effect of GRS on methylation predicted via phenotype, with ii. the directly observed effect of GRS on methylation. Results are presented as the $R^2$ between phenotype specific observed and predicted effects across the 187 CpG sites, calculated using linear regression. P values are shown for correlations between observed and predicted effects that reach $P<0.05$.
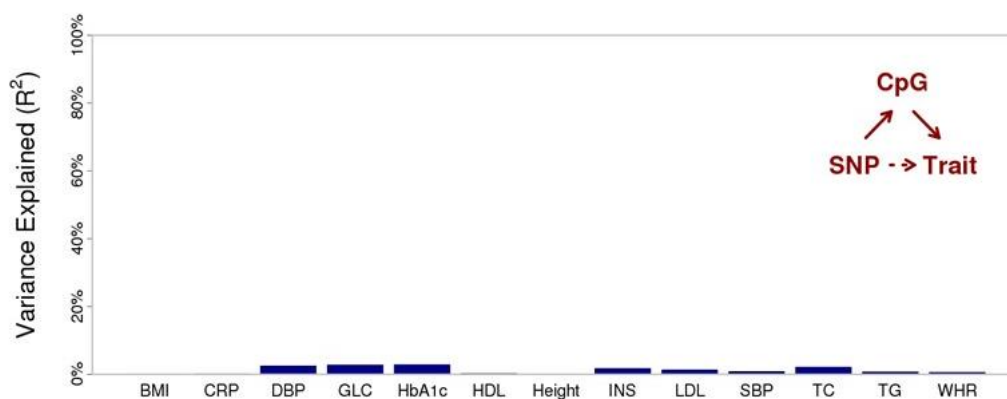
**Extended Data Figure10.** Association of established and emergent biomarkers with T2D. Results are presented as risk of T2D associated with the specified biomarkers in three models: i. Model 1 – adjusted for age and sex; ii. Model 2 – as for Model 1, but additionally for body mass index and impaired fasting glucose; iii. Model 3 – as for Model 2, but additionally for central obesity and insulin concentrations. CRP: C-reactive protein; MRS: methylation risk score. Results for quantitative traits (amino acids, CRP, insulin, MRS) are presented as risk of T2D in Q4 compared to Q1.

| Model 1 | P-value |
|---|---|
| Overweight | 4.6e-11 |
| Centrally obese | 3.3e-13 |
| CRP | 1.8e-10 |
| Phenylalanine | 7.9e-11 |
| Tyrosine | 7.2e-12 |
| Isoleucine | 1.7e-09 |
| Obese | 4.5e-29 |
| Impaired Fasting Glucose | 5.2e-38 |
| MRS | 3.4e-31 |
| Insulin | 2.1e-35 |

| Model 2 | P-value |
|---|---|
| Central obesity | 2.8e-04 |
| CRP | 6.9e-03 |
| Tyrosine | 7.6e-04 |
| Phenylalanine | 9.4e-05 |
| Isoleucine | 1.5e-04 |
| Insulin | 2.6e-12 |
| MRS | 1.8e-16 |

| Model 3 | P-value |
|---|---|
| CRP | 9.3e-02 |
| Tyrosine | 3.3e-02 |
| Isoleucine | 5.2e-03 |
| Phenylalanine | 1.7e-04 |
| MRS | 2.0e-12 |

Relative Risk (95% CI)