**A peripheral blood DNA methylation signature of hepatic fat reveals a potential causal pathway for non-alcoholic fatty liver disease**

**Authors:**

Jiantao Ma 1\*, Jana Nano 2,3\*, Jingzhong Ding 4\*, Yinan Zheng 5\*, Rachel Hennein 1, Chunyu Liu 1,6, Elizabeth K. Speliotes 7, Tianxiao Huan 1, Ci Song 1,8, Michael M. Mendelson 1,9, Roby Joehanes 1, Michelle T. Long 10, Liming Liang 11, 12, Jennifer A. Smith 13, Lindsay Reynolds 4, Mohsen Ghanbari 2,14, Taulant Muka 2, Joyce van Meurs 2,15, Louise J.M. Alferink 16, Oscar H. Franco 2, Abbas Dehghan 2,17, Scott Ratliff 13, Wei Zhao 13, Lawrence Bielak 13, Sharon LR Kardia 13, Patricia A Peyser 13, Hongyan Ning 18, Lisa B. VanWagner 18,19, Donald M Lloyd-Jones 18, John Jeffrey Carr 20, Philip Greenland 18, Alice H. Lichtenstein 21, Frank B. Hu 22, Yongmei Liu 4, Lifang Hou 5, Sarwa Darwish Murad 16, Daniel Levy 1

\* These authors contributed equally

**Affiliations**:

1. Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland and the Framingham Heart Study, Framingham, Massachusetts, USA

2. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands

3. Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Epidemiology, Neuherberg, Germany

4. Department of Epidemiology and Prevention, Wake Forest School of Medicine, North Carolina, USA

5. Center for Population Epigenetics, Robert H. Lurie Comprehensive Cancer Center and Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

6. Department of Biostatistics, Boston University, Boston, Massachusetts, USA

7. University of Michigan Medical School, Ann Arbor, Michigan, USA

8. Department of Medical Sciences, Molecular epidemiology, Uppsala University, Sweden

9. Department of Cardiology, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA

10. Section of Gastroenterology, Department of Medicine, Boston University School of

Medicine, Boston, Massachusetts, USA

11. Department of Epidemiology, Harvard T.H. Chan School of Public Health

12. Department of Biostatistics, Harvard T.H. Chan School of Public Health

13. Department of Epidemiology, School of Public Health, University of Michigan, MI, USA

14. Department of Genetics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

15. Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands

16. Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center, Rotterdam, the Netherlands

17. Department of Biostatistics and Epidemiology, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK

18. Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

19. Division of Gastroenterology & Hepatology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

20. Department of Radiology & Radiological Sciences, Vanderbilt University Medical Center, Nashville, Tennessee, USA

21. Cardiovascular Nutrition Laboratory, USDA Human Nutrition Research Center on Aging at Tufts University, Boston, Massachusetts, USA

22. Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

**Correspondence to:** Daniel Levy, MD, Framingham Heart Study, Framingham, MA, and the Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD. Email: levyd@nhlbi.nih.gov

**Word counts:** 6,176

**Competing interest statement:** All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

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

**Funding:** Detailed funding sources were included in acknowledgement in the Supplemental Method.

**Author contributions**: JM and DL designed the analysis; JM, JN, JD, YZ, TH, RJ, and LB analyzed the data; CL provided statistical support; and JM, JN, JD, YZ, and JAS wrote the manuscript; RH, EKS, CS, MMM, MTL, LR, MG, TM, JM, LJMA, OHF, AD, SR, WZ, SLRK, PAP, HN, LBW, DML, JJC, PG, AHL, FBH, YL, LH, SDM, and DL provided critical editorial comments; DL had primary responsibility for the final content. All authors read and approved the final manuscript.

**Availability of data and material:** The datasets analyzed in the present study are available at <https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000285.v3.p2> (Coronary Artery Risk Development in Young Adults); <https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v29.p10> (Framingham Heart Study); <https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001238.v1.p1> (Genetic Epidemiology Network of Arteriopathy); <https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000209.v13.p3> (Multi-Ethnic Study of Atherosclerosis); (Rotterdam Study)

**Abstract**

***Background and Aims***. Nonalcoholic fatty liver disease (NAFLD) is the most common liver disease. Methylation patterns of leukocyte DNA may reveal biomarkers and therapeutic targets to address the rising epidemic of NAFLD. We aimed to identify the peripheral blood DNA methylation signature of hepatic fat.

***Methods***. We conducted an epigenome-wide association study of hepatic fat in 3,400 European ancestry (EA) participants from four population-based cohort studies. Hepatic fat was measured using computed tomography or ultrasound imaging and DNA methylation was assessed at over 400,000 cytosine-guanine dinucleotides (CpGs) in whole blood or CD14+ monocytes using the Illumina BeadChip. Additionally, we implemented epigenome-wide association studies in 401 participants of Hispanic ancestry (HA) and 724 participants of African ancestry (AA).

***Results***. We identified 22 CpGs associated with hepatic fat in EA participants at a false discovery rate <0.05 (corresponding p=6.9×10-6) and replication at Bonferroni corrected p<8.6×10-4. Mendelian randomization analyses supported a causal contribution of hypomethylation of cg08309687 (*LINC00649*) on NAFLD (p=1.1×10-7). Hypomethylation at the same locus, cg08309687, was also putatively causal for increased fasting glucose (p=0.04). One of the 22 replicated CpGs in EA participants, cg19693031 (*TXNIP*), was associated with liver fat in HA participants (p=1.7×10-4). After correcting for multiple testing, no CpGs were significant in AA participants.

***Conclusion***. Our study demonstrates that a peripheral blood derived DNA methylation signature is robustly associated with hepatic fat accumulation. The potentially causal CpGs may represent attractive biomarkers and therapeutic targets for NAFLD. Future studies are warranted to explore underlying mechanisms and to examine DNA methylation signatures of NAFLD across racial/ethnic groups.

**Introduction**

Nonalcoholic fatty liver disease (NAFLD) includes a spectrum of histologic features ranging from hepatic fat accumulation (steatosis) to inflammation and/or fibrosis (steatohepatitis) to end-stage cirrhosis; though steatosis is the most common phenotype [1](#_ENREF_1). The prevalence of NAFLD has increased substantially along with the increasing rates of obesity worldwide [2](#_ENREF_2). NAFLD is considered to be the hepatic manifestation of metabolic syndrome due to its strong correlation to type 2 diabetes and cardiovascular disease [3](#_ENREF_3) and it is currently the second leading contributor to hepatic failure necessitating transplantation [4](#_ENREF_4).

A prior study in three family-based cohorts estimated the heritability of steatosis to be 27%; however, common genetic variants from genome-wide association studies (GWAS) account for less than five percent of inter-individual variance in hepatic fat [5](#_ENREF_5). Epigenetics may explain part of the inter-individual variance of steatosis. DNA methylation is the most widely studied epigenetic phenomenon and several studies have demonstrated altered DNA methylation profiles in liver biopsy samples collected from individuals with NAFLD [6](#_ENREF_6), [7](#_ENREF_7) . One study showed that whole blood derived DNA hypermethylation at one cytosine-guanine dinucleotide (CpG; cg06690548) located at intron of gene *SLC7A11* may be associated with a lower risk of steatosis [8](#_ENREF_8). However, in general, these studies are limited by small sample sizes to discover DNA methylation sites associated with hepatic fat accumulation.

To fill this knowledge gap, we examined the ethnicity-specific epigenome-wide association between DNA methylation at over 400,000 CpGs and hepatic fat in European ancestry (EA), African ancestry (AA), and Hispanic ancestry (HA) participants from five population-based cohort studies with hepatic fat measurements derived from noninvasive imaging. For hepatic fat-associated CpGs, we further examined their relations to genetic variants, gene expression, and regulatory functions and potential causal relations to NAFLD and impaired glycemic traits.

**Methods**

*Study population*. The present study included multiethnic participants from five population-based cohorts including the Coronary Artery Risk Development in Young Adults (CARDIA) Study, the Framingham Heart Study (FHS), the Genetic Epidemiology Network of Arteriopathy (GENOA), the Multi-Ethnic Study of Atherosclerosis (MESA), and the Rotterdam Study (RS). We excluded participants with missing DNA methylation and hepatic fat measurements and those who reported that they consumed a high amount of alcohol, equivalent to ≥21 drinks/week in men or ≥14 drinks/week in women [1](#_ENREF_1). Depending on data availability, we excluded participants who had history of myocardial infarction and stroke, cancer (except for non-melanoma skin cancer), or bariatric surgery. We also excluded those who used medication (e.g., tamoxifen, steroids, or amiodarone) or have diseases (e.g., hepatitis C) that could cause secondary hepatic steatosis. Cohort-specific exclusion is detailed in the Supplemental Methods. Due to potential differences in DNA methylation patterns between different ethnicities [9](#_ENREF_9), we analyzed the association between DNA methylation and hepatic fat separately in EA (n=3,400), HA (n=401), and AA (n=724) participants. The protocol for each study was approved by the Institutional Review Board in each cohort. All participants provided written informed consent.

*Study design*. The study design flow chart is presented in Figure 1. We first conducted the epigenome-wide association studies of hepatic fat among EA participants, including both discovery and replication. We then examined differential DNA methylation in relation to hepatic fat in the HA and AA participants. We further examined the functional and regulatory annotations for the replicated CpGs and tested the potential causal associations of the identified CpGs with NAFLD and glycemic traits.

*Hepatic fat assessment*. Detailed description for hepatic fat assessment in each cohort is presented in the Supplemental Methods. The RS used ultrasound to estimate hepatic fat and diagnosed steatosis on a dichotomized scale. The other cohorts used computed tomography (CT) to quantify hepatic fat on a continuous scale by using either mean Hounsfield units of the liver image or the ratio of the Hounsfield units of the liver image to that of a control.

*DNA methylation profiling*. Methylation profiles were measured using DNA derived from all leucocytes in peripheral blood in the FHS, CARDIA, GENOA, and RS and from CD14+ monocytes in the MESA (Supplemental Table 1). In the FHS, GENOA, MESA, and RS cohorts, DNA methylation was assayed using the Infinium HumanMethylation 450 BeadChip, which contains over 450,000 CpG sites. In the CARDIA, DNA methylation was measured using the Infinium Methylation EPIC BeadChip, which contains the majority of 450 BeadChip CpGs. Details for DNA preparation, bisulfite conversion, methylation profiling, and quality control procedures in each cohort are described in the Supplemental Methods. Raw methylation signals were normalized using various schemes, primarily the DASEN option of the *WateRmelon* R package [10](#_ENREF_10). We analyzed either methylation signal M values (in the CARDIA study; calculated as logit transformation of β values) or β values (in all other cohorts; calculated as methylated signals divided by the sum of methylated and unmethylated signals). Non-autosomal probes were excluded from the present study. For quality control purposes, we excluded study samples if they had missing more than 1-5% of methylation probes, poor SNP matching, or outliers identified by multidimensional scaling techniques. We also excluded cross-hybridizing probes and previously identified single nucleotide polymorphism (SNP) probes as described in the Supplemental Methods.

*Epigenome-wide association study of hepatic fat*. We conducted the discovery epigenome-wide association study in FHS and interrogated the differentially methylated CpGs at false discovery rate (FDR) <0.05 in the replication cohorts (EA samples in CARDIA, MESA, and RS). Because hepatic fat was measured using different scales, we meta-analyzed the p-values in the replication cohorts using logit method based on the general fixed effect model using *metap* R package. We also extracted and reported the direction of the association in each cohort. Linear regression models or linear mixed models with consideration of family structures were conducted to examine directionality and calculate p-values in each cohort. The statistical significance in the replication analysis was determined using the Bonferroni corrected p-value threshold, defined as 0.05 divided by the number of significant CpGs in the discovery phase. We used sex- and age-adjusted models (model 1) in the discovery and replication analyses. Estimated leukocyte composition [11](#_ENREF_11) and technical variables were also adjusted for in a cohort-specific manner (Supplemental Methods). Included in the sensitivity analyses, we conducted global meta-analyses in all EA participants to examine the impact of potential confounders. We performed the same sex- and age-adjusted model (model 1). We additionally adjusted for lifestyle factors including smoking status, physical activity levels, and alcohol intake in model 2. We further adjusted for BMI in model 3. We also performed a discovery and replication analysis using model 3 in EA participants.

We conducted similar epigenome-wide association studies with adjustment for same covariates to identify hepatic fat related CpGs in MESA HA participants and AA participants in the CARDIA, GENOA, and MESA studies. Similar meta-analyses of p-values were performed for AA participants. We first tested whether the replicated CpGs in EA participants were also significant in separate analyses of HA and AA participants (at Bonferroni corrected p-value threshold). Additionally, we examined whether the significant CpGs in HA or AA participants (FDR <0.05) could be replicated in the global meta-analysis in EA participants (at Bonferroni corrected p-value threshold).

*Methylation Quantitative Trait Loci (meQTL)*. To determine meQTLs, defined as DNA sequences that affect methylation levels at CpG sites, we analyzed the association of SNPs and DNA methylation in 4,170 FHS participants. We obtained SNP data in the FHS using Affymetrix 550K Array and imputed with the 1,000 Genomes Project reference panel [12](#_ENREF_12). We first calculated the residuals for DNA methylation using linear regression models with adjustment for age, sex, and technical covariates. We then regressed the residuals on SNPs. We defined *cis*-meQTLs as SNPs associated with DNA methylation at nearby CpGs (±500 kilobases (kb) from CpG, MAF >0.01, imputation r2 >0.5, p-value <1×10-4).

*Expression* *Quantitative Trait Loci (eQTL)*. We conducted eQTL analysis in 5,256 participants in FHS as previously described [13](#_ENREF_13). We excluded eQTLs (SNPs) with MAF ≤0.01, imputation r2 ≤0.5, and p-value ≥1×10-4. We defined *cis*-eQTLs as SNPs residing within 500kb of a nearby gene.

*Gene expression association analysis.* In FHS, we profiled whole blood derived mRNA expression using the Affymetrix Human Exon 1.0 ST GeneChip platform, which contains more than 5.5 million probes for 17,873 genes [13](#_ENREF_13). We examined the associations between gene expression and DNA methylation and between gene expression and hepatic fat in FHS. To prioritize genes in these analyses, we selected Illumina-annotated genes. For CpGs without annotated genes, we identified a set of genes by overlapping *cis*-meQTLs (p-value threshold <5×10-7) with *cis*-eQTLs (p-value threshold <5×10-7). In addition, we examined the association between CpGs and nearby genes (±500kb of the CpG site).

The association between DNA methylation and gene expression was analyzed in 4,561 participants in the FHS as previously described [14](#_ENREF_14). Briefly, we first calculated the residuals for gene expression using linear regression models after adjusting for sex, age, technical covariates, and blood cell counts. We then calculated statistics by regressing residuals of gene expression on residuals of DNA methylation using linear mixed models to account for family structure. For genes significantly associated with CpGs (p-value <5×10-7), we further examined their association with hepatic fat using similar statistical procedures in 2,317 FHS participants. In the association analysis between selected genes and hepatic fat, we applied Bonferroni correction to account for multiple testing, i.e., 0.05 divided by the number of genes associated with CpGs. For genes that associated with both CpGs and hepatic fat, we conducted mediation tests to estimate the proportion of mediation by gene expression on the association of CpGs and hepatic fat. In this mediation analysis, we used sex- and age-adjusted linear mixed models as described above and used the quasi-Bayesian Monte Carlo method with 1000 simulations to calculate confidence intervals [15](#_ENREF_15).

*Mendelian randomization (MR) analysis*. We conducted MR analyses (Supplemental Figure 1) to test the potential causal association from the replicated CpGs to NAFLD. Due to the well documented association between NAFLD and type 2 diabetes [16](#_ENREF_16), we also examined whether the replicated CpGs were causally associated with glycemic traits including fasting glucose, fasting insulin, and type 2 diabetes. We performed MR analysis according to the procedure of the two-sample MR, which analyzed summary statistics from instrument-exposure and instrument-outcome association analyses [17](#_ENREF_17). We used independent *cis*-meQTLs, defined as pair-wised linkage disequilibrium (LD) r2 <0.1, as instrumental variables (IVs). Using *TwoSampleMR* R package, we performed the primary analysis using the inverse variance weighted (IVW) method and sensitivity analysis using the MR-Egger method. The effect sizes and standard errors for IVs-CpG were obtained in FHS as described above and the effect sizes and standard errors for IVs-NAFLD were obtained from the meta-analysis of GWASs in the GOLD consortium [5](#_ENREF_5). We used effect sizes and standard errors derived from previous GWASs conducted by the Meta-analyses of Glucose and Insulin-related Traits Consortium (MAGIC) [18](#_ENREF_18) for IVs-fasting glucose and IVs-fasting insulin and by the Diabetes Genetics Replication and Meta-analysis Consortium (DIAGRAM) for IVs-type 2 diabetes [19](#_ENREF_19).

*Functional and regulatory annotation*. We conducted hypergeometric tests with Bonferroni correction to examine genomic characteristics of the replicated CpGs using the Infinium HumanMethylation 450 BeadChip annotation files. We queried *cis*-meQTLs in the platform of Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA GWAS) [20](#_ENREF_20). Using this platform, we examined the overlap between *cis*-meQTLs with signals in the NHGRI-EBI Catalog of published GWAS [21](#_ENREF_21). We also studied genes using differentially expressed genes measured in human whole blood and liver samples in the Genotype-Tissue Expression (GTEx v6) database [22](#_ENREF_22) and visualized the genomic region for *cis*-meQTLs. To assess the relevance of the identified peripheral blood-derived CpGs in liver, we compared DNA methylation levels in whole blood with that in liver using data deposited in the Gene Expression Omnibus (GEO Series accession number GSE48472) [23](#_ENREF_23). Gene ontology (GO) biological process enrichment analysis was performed using the GO Consortium website (<http://www.geneontology.org/>; accessed on June 8, 2018).

**Results**

*Cohort characteristics of EA participants*

The clinical characteristics of study participants are shown in Supplemental Table 1. The discovery analysis included 1,496 EA participants (mean age 59 years; 48% women) in FHS and the replication analysis included 1,904 EA participants (mean age 63 years; 52% women) from three cohort studies: RS, MESA, and CARDIA. In the additional analyses, HA participants were from MESA (n=401; mean age 68 years; 50% women) and AA participants from MESA, CARDIA, and GENOA cohorts (n=724; mean age 60 years; 61% women).

*Epigenome-wide association study of hepatic fat in European ancestry participants*

Of the 400,129 CpGs analyzed in age- and sex-adjusted models, 58 CpGs were significantly associated with hepatic fat in the discovery cohort (FHS) at FDR <0.05 (corresponding p-value = 6.9×10-6; Supplemental Table 2; Manhattan plot is displayed in Supplemental Figure 2; QQ plot with lambda is displayed in Supplemental Figure 3). The t-statistics for the 58 CpGs were correlated between the FHS and each of the replication cohorts (CARDIA, MESA, and RS EA samples; Supplemental Figure 4). In CARDIA, hepatic fat was measured at the year 20 examination. The t-statistics calculated between hepatic fat and DNA methylation measured at the year 20 examination were highly correlated with those calculated using DNA methylation measured at the year 15 examination (Supplemental Figure 5). Twenty-four (41%) CpGs replicated (Bonferroni corrected p-value <0.05/58=8.6×10-4) in the meta-analysis of the replication cohorts (n=1,904; Table 1). We removed two CpGs from the replication analysis because they were highly correlated (|*r*| ≥0.7) and close to other CpGs with lower p-values in the association analysis (Supplemental Table 3). The two CpGs were cg16246545 (51 bases upstream of cg14476101 on chromosome 1; annotated to *PHGDH*; *r* = 0.89) and cg03068497 (76 bases downstream of cg21429551 on chromosome 7; annotated to *GARS*; *r* = 0.90). Pairwise correlations among the remaining sentinel CpGs located in the same chromosome were low to moderate (|*r*| ranging from 0.09 to 0.54; Supplemental Table 3).

*Sensitivity analysis*

In the global meta-analysis of all EA participants, compared with the sex- and age-adjusted models, additional adjustment for lifestyle factors including smoking status, physical activity levels, and alcohol consumption did not materially change the association between DNA methylation levels and hepatic fat (Figure 2). Further adjusting for BMI, which is correlated with liver fat (Spearman *r* = 0.45 in FHS), reduced the strength of the associations (Figure 2); however, all 22 CpGs remained nominally associated with hepatic fat (p-value <0.05; Table 1). After adjusting for sex, age, lifestyle factors, and BMI, two CpGs, cg06690548 (*SLC7A11*) and cg19693031 (*TXNIP*), remained significant in FHS at FDR <0.05 (corresponding p-value = 1.2×10-7; Supplemental Table 4) and in the replication samples (Bonferroni corrected p-value <0.025). Both CpGs were among the replicated CpGs in the sex- and age-adjusted analysis. Leave-one-cohort-out analysis in EA participants showed p-values in the global meta-analysis with exclusion of one cohort were highly correlated with those in all samples, *r* ranging from 0.83 to 0.87 (Supplemental Figure 6).

*DNA methylation profiles in HA and AA participants*

For the 22 CpGs that replicated in the EA participants, one CpG (cg19693031; annotated to *TXNIP*) remained significant in the sex- and age-adjusted model in HA participants after Bonferroni correction for multiple testing (p-value <2.3×10-3; Supplemental Table 5). Additionally, of the 22 CpGs that replicated in the EA participants, four CpGs were nominally significant in HA participants (p-value <0.05; Supplemental Table 5) and three CpGs were nominally significant in the meta-analysis of AA participants (p <0.05; Supplemental Table 6). No CpG was detected at FDR <0.05 in HA participants. We discovered 26 CpGs at FDR <0.05 in the meta-analysis of AA participants (Supplemental Table 7), of which, two CpGs were nominally significant (p-value = 0.02 and 0.04, respectively) in the global meta-analysis of EA participants.

*Functional and regulatory annotation of hepatic fat-associated CpGs*

Compared to all analyzed CpGs on the microarray, the 22 hepatic fat-associated CpGs in EA participants were more likely to reside in the south shore (0 – 2kb downstream of CpG island; p-value = 5.5×10-4) or south shelf (2 – 4kb downstream of CpG islands; p-value = 4.1×10-4), in DNase I hypersensitivity sites (p-value = 1.7×10-3), in reprogramming-specific differentially methylated regions (p-value = 3.8×10-4), and in gene body regions (p-value = 2.4×10-4).

The mean DNA methylation levels of the 22 replicated CpGs in EA participants measured in whole blood were moderately correlated with those measured in liver tissue [23](#_ENREF_23) (Supplemental Figure 7; Pearson *r* = 0.59; p-value = 0.004). This suggests that whole blood derived DNA methylation markers may be useful proxies for the corresponding DNA methylation patterns in the liver.

The 22 hepatic fat-associated CpGs were annotated to 18 unique genes (Table 1; gene function described in Supplemental Table 8). A heatmap for average expression of the 18 genes in 53 specific tissue types included in GTEx [22](#_ENREF_22) is provided in Supplemental Figure 8. Several genes were moderately to highly expressed in liver and adipose tissues. Based on liver-specific differentially expressed genes in GTEx, the 18 annotated genes were enriched with genes that are up-regulated in the liver, including *DHCR24*, *SLC43A1*, *CPT1A*, *SREBF1*, *SC4MOL*, and *SLC9A3R1* (Bonferroni corrected p-value =0.005; Supplemental Table 9). Gene ontology (GO) pathway analysis for these annotated genes showed enrichment for 18 biological processes (Fisher’s exact with FDR corrected p <0.05; Supplemental Table 10). The most significant enriched pathway was positive regulation of the cholesterol biosynthetic process (GO:0045542; >100-fold enrichment; FDR adjusted p-value = 0.02), which included two known lipid-metabolism related genes, *ABCG1* and *SREBF1*[24](#_ENREF_24), [25](#_ENREF_25).

*GWAS analysis*

We were able to identify 3,737 *cis*-meQTL (i.e. SNPs associated with CpGs) variants for 18 of the 22 hepatic fat-associated CpGs (Table 1) in FHS. By overlapping *cis*-meQTL variants with GWAS results in the NHGRI-EBI GWAS Catalog [21](#_ENREF_21), we found that *cis*-meQTLs or strong proxies of *cis*-meQTLs (LD R2 >0.8) for nine CpGs were associated with 26 unique traits in GWAS (Supplemental Table 11). For example, rs637868 for cg14476101 (*PHDGH*) was associated in GWAS with alanine aminotransferase (ALT) levels [26](#_ENREF_26) and rs2834288 for cg08309687 (LINC00649) was associated in GWAS with abundance of gut microbiota [27](#_ENREF_27).

*Three-way association and mediation analysis of CpGs, gene expression, and liver fat*

In FHS, seven of the 22 CpGs replicated in EA participants were associated with whole blood derived expression of six annotated genes (cg19693031 with *TXNIP*, cg17901584 with *DHCR24*, cg14476101 with *PHGDH*, cg06690548 with *SLC7A11*, cg00574958 with *CPT1A*, cg06500161 with *ABCG1*, and cg27243685 with *ABCG1*; Supplemental Table 12) at a p-value threshold of <5×10-7. In addition, cg17501210 (annotated to *RPS6KA2*) was associated with expression of one non-annotated *cis*-gene, *RNASET2* (transcription start site residing 301kb downstream from cg17501210; p-value = 9.4×10-11; Supplemental Table 12). Among these seven genes, expression levels of *ABCG1* and *CPT1A* were significantly associated with liver fat in FHS (p-value = 1.2×10-30 and 2.0×10-17, respectively). Overall, hypermethylation of cg06500161 and cg27243685 were associated with decreased gene expression of *ABCG1* and increased hepatic fat, whereas, hypomethylation of cg00574958 was associated with increased gene expression of *CPT1A* and increased hepatic fat (Supplemental Figure 9). In addition, expression of *ABCG1* and *CPT1A* mediated the association between corresponding CpGs and hepatic fat by ~20% and 10%, respectively (Supplemental Figure 9).

Two CpGs, cg08309687 (LINC00649) and cg18120259 (LOC100132354), reside in intergenic non-protein coding regions and had no annotated genes in the Illumina annotation database. We overlapped their *cis*-meQTLs (p-value <5×10-7) with whole-blood derived *cis*-eQTLs (p-value <5×10-7) and identified eight *cis*-genes for cg08309687 and one *cis*-gene for cg18120259 (regional plots in Supplemental Figure 10A and 10B). None of the CpG-gene pairs was significantly associated at the predefined threshold (p-value <5×10-7); however, five of these pairs were nominally significant (p-value <0.05; Supplemental Table 12). Among the five genes, whole blood derived expression of *TMEM50B* was directly associated with DNA methylation level of cg08309687 (p-value = 6.3×10-4) and inversely associated with hepatic fat (p-value = 4.1×10-5) in FHS (Figure 3). Expression of *TMEM50B* mediated 8% (95% CI: 2-16%; p-value <2.2×10-16) of the association between cg08309687 and hepatic fat. We lacked expression data for LINC00649 in our study sample; however, in GTEx whole blood samples [22](#_ENREF_22), expression level of LINC00649 was associated with expression of *TMEM50B* (Spearman *r* = 0.76, p-value <2.2×10-16).

*MR analysis for a potential causal role of DNA methylation on NAFLD*

We conducted MR analyses to identify CpGs that may be causal for NAFLD using *cis*-meQTLs as IVs. The IVW analysis (Supplemental Table 13) showed that hypomethylation at cg08309687 (LINC00649) was significantly associated with NAFLD (Figure 4; p-value = 1.1x10-7). In addition, hypomethylation at cg14476101 (*PHDGH*) was nominally associated with NAFLD (Supplemental Figure 11; p-value = 0.02). Neither CpG was significant in the sensitivity analysis using the MR-Egger method (p-value = 0.25 and 0.91, respectively). No horizontal pleiotropy effect was detected (p-value = 0.11 and 0.27).

*MR analyses for CpGs in relation to glycemic traits*

As depicted in Supplemental Table 14, although no CpG sites showed significant causal association with glycemic traits at the Bonferroni-corrected p-value threshold, hypermethylation of cg27243685 (*ABCG1*) and hypomethylation of cg08309687 (LINC00649) were nominally associated with fasting glucose concentrations (p-value = 0.01 and 0.04, respectively). Additionally, hypermethylation of cg21429551 (*GARS*) and cg14020176 (*SLC9A3R1*) and hypomethylation of cg02711608 (*SLC1A5*) were nominally associated with increased fasting insulin concentrations (p-value =0.02, 0.049, 0.02, respectively). Lastly, hypomethylation of cg14020176 (*SLC9A3R1*) was nominally associated with increased risk of type 2 diabetes (p-value =0.01).

**Discussion**

We found that differential methylation of peripheral blood derived DNA at 22 CpG sites was associated with hepatic fat in 3,400 EA participants in four population-based cohort studies. These CpGs reside at several loci regulating key biological processes relevant to the development of steatosis. Findings from the MR analyses are consistent with a potential causal role of differential DNA methylation in relation to fat accumulation and implicate steatosis as a potential causal factor for altered glucose metabolism. For example, hypomethylation at cg08309687 was associated with increased hepatic fat and higher fasting glucose concentrations. Taken together, our study demonstrates a unique peripheral blood derived DNA methylation signature for hepatic fat accumulation and provides insights into how environmental exposures may cause NAFLD and its downstream consequences mediated by DNA methylation.

Previous studies using array-based methods have examined epigenome-wide DNA methylation patterns in the liver samples of individuals with biopsy-proven NAFLD [6](#_ENREF_6), [7](#_ENREF_7). While these studies showed a strong contrast between DNA methylation profiles of individuals with nonalcoholic steatohepatitis (NASH) compared with controls (i.e., individuals without steatosis or steatohepatitis), differences in DNA methylation patterns of individuals with steatosis (fatty liver alone) versus controls were less obvious [6](#_ENREF_6), [7](#_ENREF_7). In contrast to prior studies using liver biopsies, we used non-invasive imaging to assess hepatic fat, which is a safe and cost-effective screening tool in population-based studies. We therefore had a much larger sample size and more statistical power to detect epigenetic signals associated with elevated hepatic fat.

Two recent large epigenome-wide association studies identified CpGs associated with BMI [14](#_ENREF_14), [28](#_ENREF_28). The majority (17 of the 22) of the replicated CpGs in the present study were also observed in at least one of the two large-scale BMI-DNA methylation studies. Several CpGs associated with both liver fat and BMI are annotated to key genes involved in lipid metabolism pathways, namely *SREBF1*, *CPT1A*, *ABCG1*, and *DHCR24* [24](#_ENREF_24), [25](#_ENREF_25). These genes are in a top gene network associated with general adiposity identified in a previous analysis in MESA – a participating cohort in the present study [29](#_ENREF_29). Therefore, the overlap between liver fat associated CpGs with those linked to BMI supports a key role of adiposity and adiposity-related pathways in the pathogenesis of steatosis.

Furthermore, two hepatic fat-associated CpGs (cg14476101 and cg21429551) are annotated to *PHGDH* and *SARS*, which are key regulators involved in L-serine metabolism and have been previously related to the severity of steatosis [30](#_ENREF_30). *PHGDH* (*phosphoglycerate dehydrogenase*) encodes a key enzyme for serine synthesis. One study using genomic-scale metabolic models showed that patients with NASH had low expression levels of *PHGDH* and serine deficiency [30](#_ENREF_30). Consistent with those findings, the present study observed that hypermethylation of blood-derived DNA at cg14476101 was associated with downregulated blood-derived gene expression of *PHGDH* and increased hepatic fat accumulation. Additionally, a similar association between cg14476101 and expression levels of *PHGDH* was observed in a study comparing liver biopsy samples obtained from suspected NAFLD patients with normal control samples [28](#_ENREF_28).

Our MR analyses implicated cg08309687 as a potential causal factor for NAFLD. This CpG is located in a long intergenic non-coding region (LINC00649) that may play a role in transcription regulation relevant to steatosis. We could not examine if LINC00649 mediated the observed association between cg08309687 and hepatic fat because we lacked expression data for LINC00649 in our study cohorts. However, we observed that *cis*-meQTL variants for cg08309687 coincide with *cis*-eQTLs of several nearby genes. We also established three-way associations of DNA methylation levels at cg08309687, hepatic fat, and gene expression levels for a nearby gene, *TMEM50B* (Figure 3). Together with data from GTEx, our analysis indicates DNA methylation at cg08309687 may affect LINC00649 and subsequently alter expression of nearby genes (e.g., *TMEM50B*) and impact hepatic fat accumulation. In addition, our data suggest that cg08309687 may be involved in the regulation of the gut microbiome, which has been postulated to play a critical role in the development of NAFLD [31](#_ENREF_31). Future studies are needed to explore pathways underlying the observed association between DNA hypomethylation at cg08309687 and NAFLD.

Using models with additional adjustment for BMI, we showed that two CpGs, cg06690548 (*SLC7A11*) and cg19693031 (*TXNIP*), are independently associated with hepatic fat in the EA population. Additionally, cg19693031 replicated in HA participants, and the association was also independent of BMI (p-value = 0.03). Investigators from the RS (one of the contributing cohorts in the present study) previously knocked down *SLC7A11* in human hepatocytes (HepaRG cells) and observed up-regulation of multiple essential lipid metabolism genes, suggesting that *SLC7A11* may play an important role in hepatic lipid metabolism [8](#_ENREF_8). Although cg19693031 (*TXNIP*) has not been previously linked to obesity-related traits, several studies have shown that hypomethylation at cg19693031 (*TXNIP*) was associated with increased risk of type 2 diabetes [32](#_ENREF_32). As NAFLD is strongly associated with type 2 diabetes [16](#_ENREF_16), these observations are in accordance with our finding that DNA hypomethylation at this locus is associated with increased hepatic fat.

Information regarding other CpGs highlighted in our study is limited. The DNA methylation site cg05119988 (*SC4MOL*) has been associated with blood low-density lipoprotein cholesterol concentrations [33](#_ENREF_33). The annotated gene for this CpG, *SC4MOL,* is among a key gene network associated with BMI [29](#_ENREF_29). A GWAS study showed that a genetic variant annotated to *SC4MOL* was associated with fasting insulin and glucose concentrations in Africans [34](#_ENREF_34). Experimental studies also support the role of *SC4MOL* in lipid metabolism [35](#_ENREF_35), [36](#_ENREF_36). Nevertheless, additional studies are needed to examine the functions of the replicated CpGs in the present study.

To our knowledge, this is the first epigenome-wide association study investigating peripheral blood derived DNA methylation signatures of hepatic fat in the general population. However, some limitations deserve mention. The present analyses included participants from multiple ancestries. Differences in DNA methylation patterns associated with hepatic fat were observed between ethnic groups. Such variability is consistent with other observations that DNA methylation levels differ by race and/or ethnicity [9](#_ENREF_9). However, the sample size in the present study may not be sufficient for robust comparisons among the AA and HA participants. The MR analyses in the present study may be limited for a few reasons. First, the effect sizes and standard errors for the IVs were estimated using relatively small scale GWASs. Second, effect sizes and standard errors for instrument-exposure and instrument-outcome associations were estimated in partially overlapped study samples, which may lead to instrument bias [37](#_ENREF_37). Further, limitations exist due to differences in data collection methodologies among cohorts. However, we observed that DNA methylation signals were correlated among the cohorts of EA participants. As demonstrated in the CARDIA study, DNA methylation signals were relatively stable over time (Supplemental Figure 5). In addition, CpGs measured in monocytes were correlated well with those measured using whole blood samples (i.e. all leucocytes). This finding is consistent with observations from one prior BMI-DNA methylation study in which the observed associations between DNA methylation and BMI were shared across different cell subsets [28](#_ENREF_28).

In conclusion, the present study demonstrates a unique DNA methylation pattern related to hepatic fat in EA participants. The unique epigenetic signature is potentially useful for predicting NAFLD. Although the mechanisms are not fully understood, our study showed that DNA methylation at several CpG sites may play causal roles in hepatic fat accumulation. These findings may be useful to design better strategies for the diagnosis of NAFLD as well as aid in its prevention and treatment. Future studies with larger and more ethnically diverse sample sizes are needed to validate our findings and to explore the biological role of DNA methylation in the development and progression of NAFLD.

**References**

1. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018;67:328-357.

2. Younossi Z, Anstee QM, Marietti M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2018;15:11-20.

3. Targher G, Byrne CD, Lonardo A, et al. Non-alcoholic fatty liver disease and risk of incident cardiovascular disease: A meta-analysis. J Hepatol 2016;65:589-600.

4. Wong RJ, Aguilar M, Cheung R, et al. Nonalcoholic steatohepatitis is the second leading etiology of liver disease among adults awaiting liver transplantation in the United States. Gastroenterology 2015;148:547-55.

5. Speliotes EK, Yerges-Armstrong LM, Wu J, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet 2011;7:e1001324.

6. de Mello VD, Matte A, Perfilyev A, et al. Human liver epigenetic alterations in non-alcoholic steatohepatitis are related to insulin action. Epigenetics 2017;12:287-295.

7. Murphy SK, Yang H, Moylan CA, et al. Relationship between methylome and transcriptome in patients with nonalcoholic fatty liver disease. Gastroenterology 2013;145:1076-87.

8. Nano J, Ghanbari M, Wang W, et al. Epigenome-Wide Association Study Identifies Methylation Sites Associated With Liver Enzymes and Hepatic Steatosis. Gastroenterology 2017;153:1096-1106 e2.

9. Galanter JM, Gignoux CR, Oh SS, et al. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. Elife 2017;6.

10. Pidsley R, CC YW, Volta M, et al. A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics 2013;14:293.

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

12. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. Nature 2015;526:68-74.

13. Joehanes R, Zhang X, Huan T, et al. Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. Genome Biol 2017;18:16.

14. Mendelson MM, Marioni RE, Joehanes R, et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. PLoS Med 2017;14:e1002215.

15. Dustin T, Teppei Y, Kentaro H, et al. Mediation: R Package for Causal Mediation Analysis. Journal of Statistical Software 2014;59:1-38.

16. Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. Nat Rev Gastroenterol Hepatol 2013;10:330-44.

17. Hartwig FP, Davies NM, Hemani G, et al. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. Int J Epidemiol 2016;45:1717-1726.

18. Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet 2010;42:105-16.

19. Scott RA, Scott LJ, Magi R, et al. An Expanded Genome-Wide Association Study of Type 2 Diabetes in Europeans. Diabetes 2017;66:2888-2902.

20. Watanabe K, Taskesen E, van Bochoven A, et al. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8:1826.

21. Galperin MY, Fernandez-Suarez XM, Rigden DJ. The 24th annual Nucleic Acids Research database issue: a look back and upcoming changes. Nucleic Acids Res 2017;45:D1-D11.

22. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580-5.

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

24. Strable MS, Ntambi JM. Genetic control of de novo lipogenesis: role in diet-induced obesity. Crit Rev Biochem Mol Biol 2010;45:199-214.

25. Frisdal E, Le Lay S, Hooton H, et al. Adipocyte ATP-binding cassette G1 promotes triglyceride storage, fat mass growth, and human obesity. Diabetes 2015;64:840-55.

26. Liu Y, Fernandez CA, Smith C, et al. Genome-Wide Study Links PNPLA3 Variant With Elevated Hepatic Transaminase After Acute Lymphoblastic Leukemia Therapy. Clin Pharmacol Ther 2017.

27. Bonder MJ, Kurilshikov A, Tigchelaar EF, et al. The effect of host genetics on the gut microbiome. Nat Genet 2016;48:1407-1412.

28. Wahl S, Drong A, Lehne B, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 2017;541:81-86.

29. Ding J, Reynolds LM, Zeller T, et al. Alterations of a Cellular Cholesterol Metabolism Network Are a Molecular Feature of Obesity-Related Type 2 Diabetes and Cardiovascular Disease. Diabetes 2015;64:3464-74.

30. Mardinoglu A, Agren R, Kampf C, et al. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. Nat Commun 2014;5:3083.

31. Leung C, Rivera L, Furness JB, et al. The role of the gut microbiota in NAFLD. Nat Rev Gastroenterol Hepatol 2016;13:412-25.

32. Soriano-Tarraga C, Jimenez-Conde J, Giralt-Steinhauer E, et al. Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. Hum Mol Genet 2016;25:609-19.

33. Dekkers KF, van Iterson M, Slieker RC, et al. Blood lipids influence DNA methylation in circulating cells. Genome Biol 2016;17:138.

34. Chen G, Bentley A, Adeyemo A, et al. Genome-wide association study identifies novel loci association with fasting insulin and insulin resistance in African Americans. Hum Mol Genet 2012;21:4530-6.

35. Lu Y, Dolle ME, Imholz S, et al. Multiple genetic variants along candidate pathways influence plasma high-density lipoprotein cholesterol concentrations. J Lipid Res 2008;49:2582-9.

36. Rodriguez C, Raposo B, Martinez-Gonzalez J, et al. Modulation of ERG25 expression by LDL in vascular cells. Cardiovasc Res 2003;58:178-85.

37. Hartwig FP, Davies NM. Why internal weights should be avoided (not only) in MR-Egger regression. Int J Epidemiol 2016;45:1676-1678.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1. Significant CpGs in the epigenome-wide association study for hepatic fat in European ancestry participants | | | | | | | | | | |
| CpG | CHR | Position | Gene | Has *cis*-meQTLs | Primary discovery-replication analysis | | | Global meta-analysis | | |
| Discovery  p-value | Replication p-value | Direction | p-value 1 | p-value 2 | p-value 3 |
| cg09469355 | 1 | 2161886 | SKI | Yes | 2.93E-08 | 4.07E-04 | -,-,-,- | 6.66E-09 | 7.60E-08 | 5.14E-05 |
| cg17901584 | 1 | 55353706 | DHCR24 | Yes | 4.76E-08 | 2.10E-04 | -,-,-,- | 5.28E-09 | 1.48E-08 | 2.94E-03 |
| cg03725309 | 1 | 109757585 | SARS | Yes | 1.37E-06 | 1.29E-06 | -,-,-,- | 6.21E-10 | 4.60E-10 | 1.02E-04 |
| cg14476101 | 1 | 120255992 | PHGDH | Yes | 7.54E-08 | 1.10E-05 | -,+,-,- | 6.67E-10 | 2.10E-09 | 9.84E-05 |
| cg19693031 | 1 | 145441552 | TXNIP | No | 1.33E-12 | 3.29E-04 | -,+,-,- | 1.96E-11 | 4.32E-11 | 1.66E-07 |
| cg06690548 | 4 | 139162808 | SLC7A11 | No | 1.78E-15 | 5.27E-06 | -,-,-,- | 6.77E-14 | 2.30E-14 | 9.25E-09 |
| cg05119988 | 4 | 166251189 | SC4MOL | Yes | 6.91E-06 | 7.75E-05 | -,-,-,- | 5.50E-08 | 2.06E-07 | 8.51E-04 |
| cg03957124 | 6 | 37016869 |  | No | 9.08E-08 | 9.80E-05 | -,-,-,- | 4.25E-09 | 1.31E-08 | 9.10E-03 |
| cg18120259 | 6 | 43894639 |  | Yes | 3.35E-06 | 3.06E-04 | -,+,-,- | 1.12E-07 | 1.22E-06 | 7.93E-03 |
| cg17501210 | 6 | 166970252 | RPS6KA2 | Yes | 1.30E-07 | 3.38E-07 | -,-,-,- | 5.53E-11 | 4.23E-10 | 1.76E-03 |
| cg21429551 | 7 | 30635762 | GARS | Yes | 7.29E-09 | 1.85E-04 | -,+,-,- | 1.53E-09 | 2.94E-09 | 1.42E-04 |
| cg11376147 | 11 | 57261198 | SLC43A1 | Yes | 4.88E-06 | 8.76E-05 | -,+,-,- | 4.87E-08 | 7.11E-08 | 2.70E-03 |
| cg00574958 | 11 | 68607622 | CPT1A | No | 3.27E-10 | 9.93E-06 | -,-,-,- | 3.05E-11 | 6.28E-11 | 4.46E-03 |
| cg26894079 | 11 | 122954435 | ASAM | Yes | 5.58E-06 | 8.57E-04 | -,-,-,- | 3.94E-07 | 2.81E-06 | 3.83E-03 |
| cg11024682 | 17 | 17730094 | SREBF1 | Yes | 6.58E-07 | 2.68E-07 | +,+,+,+ | 1.10E-10 | 1.67E-10 | 1.08E-03 |
| cg14020176 | 17 | 72764985 | SLC9A3R1 | Yes | 2.36E-06 | 7.97E-04 | +,+,+,+ | 2.05E-07 | 2.73E-07 | 3.36E-05 |
| cg19016694 | 17 | 80821826 | TBCD | Yes | 9.49E-09 | 2.42E-05 | -,-,-,- | 3.74E-10 | 4.92E-10 | 7.63E-04 |
| cg15860624 | 19 | 3811194 | ZFR2 | Yes | 3.46E-07 | 5.12E-05 | +,+,+,+ | 5.73E-09 | 4.74E-09 | 1.76E-06 |
| cg02711608 | 19 | 47287964 | SLC1A5 | Yes | 2.76E-09 | 1.21E-04 | -,+,-,- | 6.23E-10 | 1.06E-09 | 2.44E-03 |
| cg08309687 | 21 | 35320596 |  | Yes | 2.48E-07 | 3.57E-06 | -,-,-,- | 5.37E-10 | 5.29E-10 | 4.65E-05 |
| cg27243685 | 21 | 43642366 | ABCG1 | Yes | 1.15E-13 | 1.14E-05 | +,+,+,+ | 6.77E-13 | 2.08E-12 | 1.20E-06 |
| cg06500161 | 21 | 43656587 | ABCG1 | Yes | 1.95E-16 | 3.35E-09 | +,+,+,+ | 3.45E-16 | 7.33E-16 | 2.45E-09 |
| The discovery-replication analysis used sex- and age-adjusted models. The global meta-analysis was conducted in all European ancestry participants. 1. Model adjusted for sex and age; 2. Model adjusted for sex, age and lifestyle covariates including smoking, physical activity, and alcohol; 3. Model adjusted for sex, age, lifestyle covariates, and BMI. ‘+’ sign represents hypermethylation is associated with increased hepatic fat and ‘-’ sign represents hypomethylation is associated with increased hepatic fat. CHR: chromosome; cis-meQTL: cis-methylation quantitative trait loci | | | | | | | | | | |

Figure 1. Study design flow chart.

Figure 2. Comparisons of sequential adjustment models in European ancestry participants. Y-axis values are observed -log10(p-values) in the global meta-analysis of all EA participants. X-axis are sorted by CpGs (symbols at same vertical position are same CpG). Blue cross represents sex- and age-adjusted model. Green square represents model with additional adjustment for lifestyle factors (smoking, physical activity, and alcohol intake). Red circle represents the fully adjusted model including sex, age, smoking, physical activity, alcohol intake, and BMI.

Figure 3. Three-way association of whole-blood derived DNA methylation level at cg08309687, whole-blood derived gene expression for *TMEM50B*, and CT-derived liver-phantom ratio (hepatic fat) in the Framingham Heart Study.

Figure 4. Potential causal association of cg08309687 to non-alcoholic fatty liver disease using Mendelian randomization (MR) analysis. Plot depicts IVW (solid line) and MR Egger estimate (dashed line). No horizontal pleiotropy effect was detected (p-value= 0.11). SNP=single nucleotide polymorphism; IVW=inverse variance weighted.