A multi-ethnic epigenome-wide association study of leukocyte DNA methylation and blood lipids

Min-A Jhun1,2✉, Michael Mendelson3,4, Rory Wilson5,6, Rahul Gondalia7, Roby Joehanes8, Elias Salfati2, Xiaoping Zhao9, Kim Valeska Emilie Braun10, Anh Nguyet Do11, Åsa K. Hedman12, Tao Zhang13, Elena Carnero-Montoro14,15, Jincheng Shen16, Traci M. Bartz17, Jennifer A. Brody18, May E. Montasser19,20, Jeff R. O’Connell19,20, Chen Yao3, Rui Xia9, Eric Boerwinkle21, Megan Grove21, Weihua Guan22, Pfeiffer Liliane5,6, Paula Singmann5,6, Martina Müller-Nurasyid23,24,25, Thomas Meitinger25,26,27, Christian Gieger5,6, Annette Peters6,25, Wei Zhao1, Erin B. Ware1,28, Jennifer A. Smith1,28, Kloidan Dhana29, Joyce van Meurs30, Andre Uitterlinden30, Mohammad Arfan Ikram10, Mohsen Ghanbari10, Deugi Zhi31, Stefan Gustafsson12, Lars Lind12, Shengwu Li31, Dianjianyi Sun13,32, Tim D. Spector14, Yiider Ida Chen33, Coleen Damcott19,20, Alan R. Shuldiner19,20, Devin M. Absher34, Steve Horvath35,36, Philip S. Tsao16,27,38, Sharon Kardia1, Bruce M. Psaty39,40, Nona Sotoodehnia41, Jordana T. Bell14, Erik Ingelsson2,12,38,42, Wei Chen3, Abbas Dehghan10,43, Donna K. Arnett11, Melanie Waldenberger5,6, Lifang Hou44, Eric A. Whitsett7,45, Andrea Baccarelli16,47, Daniel Levy3,48, Myriam Fornage9, Marguerite R. Irvin11 & Themistocles L. Assimes2,37,38✉

Here we examine the association between DNA methylation in circulating leukocytes and blood lipids in a multi-ethnic sample of 16,265 subjects. We identify 148, 35, and 4 novel associations among Europeans, African Americans, and Hispanics, respectively, and an additional 186 novel associations through a trans-ethnic meta-analysis. We observe a high concordance in the direction of effects across racial/ethnic groups, a high correlation of effect sizes between high-density lipoprotein and triglycerides, a modest overlap of associations with epigenome-wide association studies of other cardio-metabolic traits, and a largely non-overlap with lipid loci identified to date through genome-wide association studies. Thirty CpGs reached significance in at least 2 racial/ethnic groups including 7 that showed association with the expression of an annotated gene. CpGs annotated to CPT1A showed evidence of being influenced by triglycerides levels. DNA methylation levels of circulating leukocytes show robust and consistent association with blood lipid levels across multiple racial/ethnic groups.
Abnormal blood lipid levels are important risk factors for various diseases including cardiovascular disease, diabetes, renal disease, Alzheimer’s disease, and cancers. Several large-scale genome-wide and exome-wide association studies of lipids have identified single nucleotide polymorphisms (SNPs) involved in lipid metabolism. However, the biological mechanisms behind abnormalities in lipid metabolism are not fully understood.

Complex traits are a manifestation of not only genetic but also environmental factors which in part express themselves through the epigenetic modification of DNA in all cell types. Epigenetic modifications can explain differences in a phenotype between monozygotic twins as well as changes in a phenotype within an individual over time. Epigenome-wide association studies (EWAS) provide an opportunity to document differences in epigenetic marks between individuals through the quantification of the degree of methylation at thousands of CpG sites across the genome. To date, such studies have identified methylation signatures associated with a number of cardiometabolic traits including cigarette smoking, hepatic fat, fasting insulin or HOMA-IR, incident type 2 diabetes, renal function, blood pressure and C-reactive protein. In addition, several EWAS identified CpGs significantly associated with blood lipid levels in populations of European ancestry. However, large-scale multi-ethnic studies to identify epigenetic determinants of blood lipid levels are lacking.

Here, we present a racial/ethnic groups specific meta-analysis of 15 EWAS of DNA from circulating white blood cells involving a total 16,265 participants from 3 racial/ethnic groups to identify CpGs with DNA methylation levels that are significantly associated with blood lipid levels. We identify 187 racial/ethnic specific novel CpG associations between 0% in the Amish to 44% in FHS.

Results

Study population. We analyzed 12 cohorts of Europeans (EA) involving 11,114 participants, 7 cohorts of African Americans (AA) involving 4,425 participants, and 2 cohorts of Hispanics (HISP) involving 699 participants (Table 1, Supplementary methods). The TwinsUK, WHI-BA23, and WHI-EMPC cohorts were composed of female participants only while NAS was composed of male participants only. The range of mean age, body mass index (BMI), high-density lipoprotein (HDL) levels, low-density lipoprotein (LDL) levels, and triglyceride (TG) levels was 42.7 to 76.0 year, 26.6 to 32.6 kg per m², 45.5 to 59.3 mg per dl, 74.1 to 168.5 mg per dl, respectively (Table 1). The percentage of study participants taking any lipid control medication at time of blood lipid measurement ranged between 0% in the Amish to 44% in FHS.

EWAS stratified by racial/ethnic group. We measured DNA methylation levels using the Illumina Infinium HumanMethylation 450 K Beadchip in peripheral blood leukocytes or whole blood, except in GOLDN where CD4+ T cells exclusively were used (Supplementary Data 1). We performed an EWAS on HDL, LDL, and TG using four linear mixed effects models in each cohort, stratified by racial/ethnic group and a random effects meta-analysis with genomic control (GC) and Bonferroni correction for the number of probes tested (P < 1.09 × 10⁻⁷) (Methods, Supplementary Table 1).

| Table 1 Descriptive of the participated cohorts (N = 16,265). |
|-------------------------------------------------------------|
| **Cohort** | **N** | **Age (year)** | **Female Percent** | **BMI (kg/m²)** | **Per cent on Lipid meds** | **HDL (mg/dl)** | **LDL (mg/dl)** | **TG (mg/dl)** |
|-------------|------|---------------|------------------|-----------------|---------------------------|----------------|----------------|---------------|
| **European (N = 11,114)** |      |               |                  |                 |                           |                |                |               |
| Amish       | 158  | 43.8 ± 13.2   | 54               | 27.8 ± 4.7      | 0             | 55.8 ± 14.9               | 139.5 ± 39.9   | 74.1 ± 39.4   |
| BHS         | 676  | 42.8 ± 4.5    | 55               | 29.9 ± 6.7      | 12            | 45.5 ± 13.2               | 126.7 ± 34.2   | 142.4 ± 92.9  |
| CHS         | 186  | 76.0 ± 5.0    | 55               | 27.1 ± 5.0      | 5             | 51.7 ± 13.5               | 117.1 ± 36.4   | 155.8 ± 82.0  |
| FHS*        | 2,648| 66.4 ± 8.9    | 54               | 28.3 ± 5.3      | 44            | 57.3 ± 18.0               | 104.6 ± 31.4   | 118.9 ± 69.8  |
| GOLDN*      | 714  | 48.5 ± 15.9   | 50               | 28.5 ± 5.5      | 0             | 46.1 ± 13.0               | 123.8 ± 31.5   | 137.4 ± 85.8  |
| KORA F4     | 1,651| 61.0 ± 8.9    | 51               | 28.1 ± 4.8      | 16            | 56.5 ± 14.7               | 140.3 ± 35.3   | 129.7 ± 77.8  |
| NAS         | 674  | 72.4 ± 6.8    | 50               | 28.0 ± 4.1      | 36            | 49.3 ± 12.9               | 120.1 ± 33.0   | 137.7 ± 84.0  |
| PIVUS       | 963  | 70.1 ± 0.15   | 50               | 27.0 ± 4.3      | 16            | 58.6 ± 16.4               | 128.9 ± 33.9   | 112.6 ± 51.4  |
| RS          | 724  | 59.9 ± 8.2    | 54               | 27.5 ± 4.8      | 26            | 54.3 ± 15.9               | 134.7 ± 37.6   | 131.5 ± 76.0  |
| TwinsUK*    | 708  | 58.1 ± 9.3    | 100              | 26.6 ± 4.8      | 14            | 51.0 ± 12.3               | 142.7 ± 37.6   | 154.0 ± 35.5  |
| WHI-BA23    | 940  | 68.3 ± 6.25   | 100              | 28.8 ± 5.9      | 10            | 59.2 ± 15.8               | 130.3 ± 34.1   | 158.6 ± 80.1  |
| WHI-EMPC    | 1,072| 64.7 ± 7.1    | 100              | 28.8 ± 5.8      | 14            | 51.0 ± 12.3               | 142.7 ± 37.6   | 154.0 ± 35.5  |
| African (N = 4,452) |  |               |                  |                 |                           |                |                |               |
| ARIC        | 1,877| 56.5 ± 5.9    | 64               | 30.0 ± 6.2      | 8             | 53.4 ± 17.0               | 134.7 ± 38.8   | 112.1 ± 57.3  |
| BHS         | 282  | 42.7 ± 4.7    | 61               | 32.6 ± 8.7      | 7             | 49.6 ± 14.5               | 120.8 ± 33.9   | 114.8 ± 76.9  |
| CHS         | 189  | 73.0 ± 5.5    | 67               | 28.8 ± 4.0      | 4             | 59.3 ± 16.5               | 124.2 ± 34.7   | 114.6 ± 58.6  |
| GENOA*      | 315  | 61.1 ± 7.6    | 73               | 31.0 ± 6.3      | 6             | 56.9 ± 17.4               | 125.2 ± 41.4   | 136.1 ± 63.6  |
| HyperGEN    | 604  | 48.4 ± 11.2   | 67               | 32.5 ± 8.2      | 5             | 53.6 ± 15.0               | 122.6 ± 38.0   | 107.8 ± 62.5  |
| WHI-BA23    | 652  | 62.8 ± 6.6    | 100              | 31.9 ± 6.8      | 17            | 54.8 ± 14.6               | 152.6 ± 43.8   | 122.6 ± 81.3  |
| WHI-EMPC    | 533  | 62.7 ± 6.9    | 100              | 31.5 ± 6.1      | 10            | 58.1 ± 14.7               | 135.9 ± 37.9   | 119.6 ± 54.3  |
| Hispanic (N = 699) |      |               |                  |                 |                           |                |                |               |
| WHI-BA23    | 389  | 62.2 ± 6.9    | 100              | 29.6 ± 5.4      | 16            | 50.6 ± 13.2               | 142.7 ± 37.6   | 168.5 ± 95.6  |
| WHI-EMPC    | 310  | 61.6 ± 6.2    | 100              | 29.6 ± 5.3      | 11            | 54.7 ± 12.9               | 128.8 ± 35.1   | 162.7 ± 77.1  |

*A family-based cohort; N Number of samples, SD Standard deviation, HDL high-density lipoprotein, LDL low-density lipoprotein, TG triglycerides.*
We identified 447, 25, and 496 CpGs for HDL, LDL, and TG, respectively, among EA using the basic set of covariates (model 1 adjusted for age, sex, smoking, lipid medication, four SNP PCs, estimated cell proportions, plate, row, and column of plate) (Methods, Supplementary Data 2). When we further adjusted for BMI (model 2 additionally adjusted for BMI), the numbers of significant CpGs decreased substantially for HDL (146) and TG (206) but increased modestly for LDL (30). Among AA, we identified 34, 7, and 76 CpGs in model 1 for HDL, LDL, and TG, respectively, and the numbers decreased to 9, 7, and 55 with a further adjustment with BMI (model 2). For HISP, we identified 2, 0, and 6 CpGs in model 1 for HDL, LDL, and TG, respectively, and the number decreased to 0 for HDL. Excluding participants taking any lipid lowering medication decreased the sample size by 18% and decreased power but the effect estimates remained similar (models 3 and 4 adjusted for the same set of covariates of models 1 and 2, respectively, with the exception of adjustment for the use of lipid medications, Methods). Among EA, we identified 74, 15, and 86 CpGs significantly associated (P < 1.09 × 10⁻⁷) with HDL, LDL, and TG, respectively, using this most conservative model 4 that excluded statin users and adjusted for BMI (Fig. 1, Supplementary Data 2, Supplementary Fig. 1). For AA, these numbers were 7, 5, and 43 and, for HISP, they were 2, 2, and 4 CpGs, respectively. Through trans-ethnic meta-analyses of the same model, we additionally identified 49, 24, and 119 significant (P < 1.09 × 10⁻⁷) CpG-lipid level associations for HDL, LDL, and TG, respectively, of which 46, 22, and 118 were novel when compared to both our racial/ethnic specific analyses and the literature.

Comparison of results across racial/ethnic groups. A majority of significant CpG-lipid level associations in European population did not reach statistical significance in other racial/ethnic groups. For CpGs found to be significant in at least one racial/ethnic group (model 4), we found a high rate of concordance (88 to 100%) in the direction of effect observed in Europeans versus that observed in African Americans and, separately, Hispanics for all three lipid fractions (Fig. 2). We found a high correlation between effect sizes (Pearson correlation coefficient = 0.69 to 0.93) for HDL and TG but not for LDL (−0.20 to 0.20). Lastly, regression slopes between any 2 sets of betas were close to 1 for HDL (0.75, 1.03) and TG (0.68 to 1.12) but not for LDL (−0.28 to 0.26) (Fig. 2). We calculated the correlations coefficients and estimated regression slopes for a higher number of CpGs (122 CpGs with P < 1.09 × 10⁻⁵) before and after natural log transformation to further explore whether the differences in correlations and regression slopes between LDL and HDL/TG could be a consequence of having much smaller number of significant CpGs and/or the use of a log transformed lipid measure. For these analyses, we found both higher correlation of betas (0.21 to 0.49) and regression slopes (0.34 to 0.47) for LDL although still not as high as those observed for HDL and TG (Supplementary Fig. 2).

When comparing results across all 3 racial/ethnic groups, we identified 4, 1, and 26 CpGs associated with HDL, LDL, and TG, respectively, in more than one racial/ethnic group (Table 2). Of these, 1 CpG, cg06500161, in ABCG1 was associated with both HDL and TG in opposing directions. Consistent with our findings for significant CpG-lipid trait associations overall, we found high rate of concordance of the direction of the associations of the 30 CpGs across all 3 racial/ethnic group but variable effect sizes (Supplementary Figs. 3-5). The 4 CpGs in the ABCG1, carnitine palmitoyltransferase 1 A (CPT1A), sterol regulatory element binding transcription factor 1 (SREBF1), and thioredoxin interacting protein (TXNIP) genes identified in HISP for TG were also significant in the EA and AA populations. We found multiple CpGs to be significantly associated with TG levels located within or near the phosphoglycerate dehydrogenase (PHGDH), CPT1A, SREBF1, and ABCG1 genes. Twelve out of the 30 replicated CpGs significant in at least 2 racial/ethnic groups have not been previously reported²¹⁻²⁴ (Table 2).
Expression quantitative trait methylation analysis. The association between DNA methylation and gene expression was investigated in the Framingham Heart Study (N\textsubscript{EA} = 4,278 including 2,726 offspring cohort participants and 1,552 third generation cohort participants\textsuperscript{30}). The DNA methylation levels of 7 out of the 30 CpGs listed in Table 2 were negatively associated with the expression of their respective genes: phosphoglycerate dehydrogenase (PHGDH) \textsuperscript{4.44 × 10\textsuperscript{-5}}, poly (ADP-ribose) polymerase family member 9 (PARP9) \textsuperscript{4.22 × 10\textsuperscript{-5}}, myosin regulatory light chain interacting protein (MYLIP) \textsuperscript{5.43 × 10\textsuperscript{-5}}, and ABCG1 \textsuperscript{6.26 × 10\textsuperscript{-5}}. Three of these CpGs are in the 5' UTR region while the remaining four are in the gene body. Aside from one CpG in a CpG island, six were located in either the north or south shore regions (Supplementary Table 3).

Mendelian randomization approach. We explored the causal relationships between methylation and blood lipid levels for the 30 CpGs in EA (Table 2) using a bi-directional Mendelian Randomization (MR) study design\textsuperscript{31,32}. First, we used genetic risk scores (GRS) for HDL, LDL, and TG constructed from established susceptibility loci for these phenotypes as instruments to examine the relationship between blood lipids and methylation (Supplementary Table 4). We found the GRSs to be significantly associated with their respective lipid levels in the 4 cohorts (GOLDN (N\textsubscript{EA} = 713), KORA (N\textsubscript{EA} = 1,379), WHI-BA23 (N\textsubscript{EA} = 790), and WHI-EMPC (N\textsubscript{EA} = 494) participating in the MR follow-up analysis (HDL: \textit{P}\textsubscript{meta} = 1.86 × 10\textsuperscript{-17}, LDL: \textit{P}\textsubscript{meta} = 1.13 × 10\textsuperscript{-22}, TG: \textit{P}\textsubscript{meta} = 1.13 × 10\textsuperscript{-8}). Our Mendelian randomization analysis suggested that the DNA methylation levels of three CpGs, cg00574958 \textsuperscript{(4.72 × 10\textsuperscript{-4})}, cg09737197 \textsuperscript{(3.33 × 10\textsuperscript{-3})}, and cg17058475 \textsuperscript{(6.26 × 10\textsuperscript{-8})}, were in the 5' UTR region of the \texttt{CPT1A} were influenced by blood TG levels (Supplementary Table 5). Next, we investigated the effect of DNA methylation on blood lipid levels. A total of 7 out of the 30 CpGs had at least one significant mQTL with available GWAS results from the Global Lipids Genetics Consortium results (Supplementary Table 6). We implemented inverse-weighted MR method and MR-egger when >2 mQTLs were available for a given CpG (3 CpGs out of 7). None of the CpGs were significantly associated with lipid levels using MR-egger.

Discussion We report the first large-scale multi-ethnic epigenome-wide association study (EWAS) of blood lipids. Our population specific meta-analyses revealed 187 novel CpG-lipid trait associations and identified a high concordance of the direction of effects across racial/ethnic groups for all 3 lipid traits and a high correlation of tests = 5549, Bonferroni corrected \textit{P} = 9.01 × 10\textsuperscript{-6}, AA (number of tests = 8316, Bonferroni corrected \textit{P} = 6.01 × 10\textsuperscript{-6}), and HISP (number of tests = 4,713, Bonferroni corrected \textit{P} = 1.06 × 10\textsuperscript{-5}), respectively (Supplementary Data 3). We found 7 out of our 11 CpGs in EA to also be listed to have at least one mQTLs in the mQTL DB (http://www.mqtldb.org/)\textsuperscript{29} (Supplementary Data 4). For the 7 CpGs, 95% of the significant mQTL SNPs in mQTL DB were also significant in our EA population and had consistent direction of effect. Out of the 190 significant mQTLs, all SNPs-CpG pairs were found to be mQTLs in datasets spanning across the life course in the mQTL DB including birth, childhood, adolescence, pregnancy, and middle age.

Overlap with prior related GWAS and EWAS. We examined whether the CpGs identified in our lipid EWAS are located in or near the genes implicated in previous lipid GWAS\textsuperscript{27,28}. Among EA, a total of 4 out of 106 genes had both CpGs and SNPs significantly associated with a blood lipid: apolipoprotein A5 (APOA5), apolipoprotein B (APOB), myosin regulatory light chain interacting protein (MYLIP), and PARP9 (Fig. 3). Among AA, 4 out of 32 genes had been identified in both lipid EWAS and GWAS: PARP9, Chromosome 7 Open Reading Frame 50 (C7orf50), ATP Binding Cassette Subfamily A Member 1 (ABCA1), and MYLIP. We further identified 192 CpG-SNP pairs within 10Mbp at 19 loci of which 30 pairs at 10 loci were in 1Mbp (Supplementary Data 5).

Many of the CpGs identified in our lipid EWAS have been previously identified in EWAS of BMI\textsuperscript{14}, hepatic fat\textsuperscript{15}, fasting insulin or HOMA-IR\textsuperscript{16}, incident type 2 diabetes\textsuperscript{17}, estimated glomerular filtration rate (eGFR)\textsuperscript{18}, blood pressure\textsuperscript{19}, C-reactive protein\textsuperscript{20}, and/or cigarette smoking\textsuperscript{13} (Fig. 3, Supplementary Table 2). About one half of the genes (33/192 = 17% of CpGs) identified in our BMI adjusted lipid EWAS have also been implicated in previously published BMI EWAS\textsuperscript{14}. There were also twelve genes identified in more than two EWAS: ABCG1, CPT1A, 24-dehydrocholesterol reductase (DHCR24), PHGDH, phosphoethanolamine/phosphocholine phosphotransferase (PHOSPHO1), seryl-tRNA synthetase (SARS), SKI proto-oncogene (SKI), solute carrier family 1 member 5 (SLC1A5), solute carrier family 43 member 1 (SLC43A1), SLC7A11, SREBF1, and TXNIP. Among EA, 37 out of 106 genes (36% of 164 CpGs) have been identified in EWAS of other phenotypes. There were 14 out of 32 genes (14 out of 54 CpGs) and 2 out of 6 genes (4 of 8 CpGs) have been identified in EWAS of other phenotypes for AA and HISP, respectively.

Methylation quantitative trait loci analysis. We searched for methylation quantitative trait loci (mQTL) influencing methylation levels of the 30 CpGs listed in Table 2. Five out of the 15 cohorts provided genetic data for this analysis including ARIC (N\textsubscript{AA} = 1,717), GOLDN (N\textsubscript{EA} = 713), KORA (N\textsubscript{EA} = 1,379), WHI-BA23 (N\textsubscript{EA} = 790), N\textsubscript{HISP} = 540, and N\textsubscript{HISP} = 324), and WHI-EMPC (N\textsubscript{EA} = 494, N\textsubscript{AA} = 424, and N\textsubscript{HISP} = 221). We restricted our analysis to SNPs located within 25 kilobases up- or downstream of the CpGs with a minor allele frequency (MAF) > 0.01 in each cohort and implemented a fixed effects meta-analysis within each of the three racial/ethnic groups. A total of 11, 18, and 5 CpGs had at least one significant mQTL in EA (number of
A majority of our significant CpG-lipid associations do not implicate genes previously identified through GWAS of lipids\(^7\)–\(^{10,27,28}\), but many of our associations overlap with those identified in EWAS to date of related cardiometabolic traits especially for TG and HDL\(^{14,19,33}\). Thirty CpG-lipid trait associations were significant in at least 2 racial/ethnic groups with ~1/3 of these being novel. In subgroup analyses, 19 significant CpGs also harbored mQTLs and 7 were inversely associated with levels of expression of the annotated gene. Lastly, our Mendelian randomization analyses suggested that DNA methylation levels at one locus appeared to be influenced by blood TG levels.

The numbers of statistically significant CpGs decreased dramatically for HDL and TG after adjusting for BMI. These findings suggest that BMI serves as either a strong confounder or a strong mediator of a large fraction of our CpG-lipid associations for these traits. Even after adjusting for BMI, we found ~1/3 of CpGs to be linked to BMI in other studies\(^{14}\). In addition to the potential of residual confounding, we hypothesize that many CpGs may be independently influenced by both BMI and blood lipid levels (akin to how diet and exercise have an independent effect on weight loss).

We found a relatively high overlap of findings from our study with previous EWAS findings for cardiometabolic traits and a
| Table 2 CpGs significantly associated with lipid levels in more than one racial/ethnic group. |
|----------------|----------------|----------------|----------------|----------------|
| CpG            | RACE          | GROUP          | BETA           | SE             |
| cg17901584*    | EA            | DHCR24         | 0.032          | 0.005          |
| cg24002003     | EA            | PHGDH          | 0.010          | 0.003          |
| cg02650017*    | EA            | PHGDH          | 0.006          | 0.001          |
| cg06500161*    | EA            | ABCGI          | −0.021         | 0.003          |
| cg03725309     | EA            | PARP9          | 0.0001         | 0.00002        |
| cg16246545     | EA            | PHGDH          | 0.011          | 0.002          |
| cg14476101*    | EA            | PHGDH          | 0.018          | 0.003          |
| cg19693031*    | EA            | TXNIP          | 0.019          | 0.002          |
| cg19266329     | EA            | Y              | 0.006          | 0.001          |
| cg19213703     | EA            | Y              | 0.005          | 0.001          |
| cg06690548*    | EA            | SLC7A11        | −0.012         | 0.0024         |
| cg26403843     | EA            | RNF145         | 0.011          | 0.002          |
| cg03717755*    | EA            | MYLIP          | 0.009          | 0.001          |
| cg18336453     | EA            | PTK7           | 0.004          | 0.0008         |
| cg07504977*    | EA            | SLC43A1        | 0.012          | 0.0016         |
| cg11376147*    | EA            | CPT1A          | 0.009          | 0.0012         |
| cg00574958*    | EA            | CPT1A          | 0.005          | 0.0008         |
| cg17058475*    | EA            | CPT1A          | 0.010          | 0.0013         |
| cg10919522     | EA            | C14orf43      | 0.005          | 0.001          |

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-23899-y | www.nature.com/naturecommunications
| CpG     | RACE | BETA  | SE     | P       | MAPINFO  | GENE  | GROUP  | CpG Island | Enhancer | DHS |
|---------|------|-------|--------|---------|----------|-------|--------|------------|----------|-----|
| cg15863539* | EA   | 0.003 | 0.0009 | 8.83 × 10$^{-17}$ | 17:17716950 | SREBF1 | Body  | S_Shore |
|         | AA   | 0.004 | 0.0006 | 1.93 × 10$^{-09}$ |          |       |        |            |          |     |
| cg20544516* | EA   | 0.006 | 0.0006 | 3.71 × 10$^{-25}$ | 17:17717183 | MIR33B;SREBF1 | Body;Body | S_Shore |
|         | AA   | 0.007 | 0.0008 | 3.18 × 10$^{-14}$ |          |       |        |            |          |     |
| cg11024682* | EA   | 0.010 | 0.0013 | 1.08 × 10$^{-50}$ | 17:17730094 | SREBF1 | Body  | S_Shelf |
|         | AA   | 0.016 | 0.0013 | 5.77 × 10$^{-28}$ |          |       |        |            |          |     |
|         | HA   | 0.019 | 0.0033 | 1.59 × 10$^{-08}$ |          |       |        |            |          |     |
| cg08857797* | EA   | 0.006 | 0.001 | 1.93 × 10$^{-12}$ | 17:40927699 | VPS25 | Body  |          |          |     |
|         | AA   | 0.008 | 0.0017 | 1.73 × 10$^{-10}$ |          |       |        |            |          |     |
| cg22304262 | EA   | −0.006 | 0.0012 | 6.04 × 10$^{-12}$ | 19:47287778 | SLC1AS | Body;5'UTR | N_Shelf |
|         | AA   | −0.008 | 0.0012 | 3.31 × 10$^{-08}$ |          |       |        |            |          |     |
| cg08309687 | EA   | −0.009 | 0.0017 | 4.49 × 10$^{-14}$ | 21:35320596 |       |        |            |          |     |
|         | AA   | −0.014 | 0.0024 | 3.00 × 10$^{-11}$ |          |       |        |            |          |     |
| cg0188189* | EA   | 0.005 | 0.0008 | 6.07 × 10$^{-21}$ | 21:43652704 | ABCG1 | Body  | N_Shelf |
|         | AA   | 0.007 | 0.0011 | 1.21 × 10$^{-17}$ |          |       |        |            |          |     |
| cg02370100* | EA   | 0.005 | 0.0008 | 1.45 × 10$^{-13}$ | 21:43655256 | ABCG1 | Body  | Island |
|         | AA   | 0.006 | 0.0014 | 1.49 × 10$^{-08}$ |          |       |        |            |          |     |
| cg07397296 | EA   | 0.007 | 0.0012 | 4.68 × 10$^{-16}$ | 21:43655316 | ABCG1 | Body  | Island | Y |
|         | AA   | 0.007 | 0.0012 | 3.98 × 10$^{-08}$ |          |       |        |            |          |     |
| cg06500161* | EA   | 0.016 | 0.0015 | 4.73 × 10$^{-11}$ | 21:43656587 | ABCG1 | Body  | S_Shore | Y |
|         | AA   | 0.017 | 0.0023 | 2.72 × 10$^{-39}$ |          |       |        |            |          |     |
|         | HA   | 0.021 | 0.0033 | 7.68 × 10$^{-10}$ |          |       |        |            |          |     |

*Identified in previous epigenome-wide association of lipids; SD: Standard error; CNR: Chromosome, DHS: DNase I hypersensitive site, HDL: high-density lipoprotein, LDL: low-density lipoprotein, TG: triglycerides, EA: European ancestry, AA: African ancestry, HA: Hispanic ancestry, N_EA = 11,114, N_AA = 4,452, N_HA = 699. The statistic is a likelihood ratio test (LRT) which asymptotically follows an equal mixture of 1 degree of freedom (df) $\chi^2$ distribution and 2 df $\chi^2$ distribution, CI can be calculated as (BETA−1.96*SE, BETA+1.96*SE). A $p$-value was considered significant in a specific racial/ethnic group if it was $<1.09 \times 10^{-7}$ (Bonferroni correction for the number of CpG probes tested).
relative paucity of overlap with lipid loci implicated through GWAS. We hypothesize that the influence of important upstream environmental determinants of the metabolic syndrome such as diet and exercise may be responsible for these patterns although substantial additional research is needed to prove this hypothesis including studies measuring changes in epigenetic profiles of multiple relevant cell types after dietary and physical activity interventions.

We could not detect methylation QTls (mQTls) for many of our 30 CpGs-lipid trait associations replicating across 2 racial/ethnic groups. Among mQTls identified, a majority were not consistently associated with the CpG over the life course. Whether this observation reflects direct changes in methylation levels of CpGs associated with lipid levels that occur over a lifetime of accumulated environmental exposures, such as diet and exercise, remains unknown. Other environmental exposures or time-dependent events leading to subtle changes in white blood cell proportions may also be responsible for these observations.

We highlight 3 EWAS lipid loci uncovered through our meta-analysis among the plethora of both novel and known findings. First, we found a CpG to be associated with PARP9 among our few LDL findings. PARP9 is homologous to PARP10 and both are ADP-ribosyltransferases with 30% of their catalytic domains being identical. Both also have been previously identified as enzymes involved in high-density lipoprotein (HDL), low-density lipoprotein (LDL), or triglycerides (TG) levels after adjusting for BMI and excluding subjects on statins. Genes identified in more than one racial/ethnic population are in bold font. Genes identified in EWAS of related cardiometabolic traits are marked with special characters: BMI (#), hepatic fat (*), fasting insulin or HOMA-IR ($), incident type2 diabetes (&), eGFR (‘), blood pressure (%), C-reactive protein (CRP) (@), and cigarette smoking (+). Genes identified in previous GWAS are underlined.

developing type 2 diabetes while genetic variants mapped to this locus are linked to atherosclerosis. These constellations of findings suggest that ABCG1 may play a role in predisposing to or mediating the effects of the metabolic syndrome. Third, we found the methylation status of CpGs in CPT1A, a gene that initiates the oxidation of long-chain fatty acids, to be influenced by blood levels of TG through our MR analysis. The same CpG locus are linked to atherosclerosis. These constellations of findings suggest that CPT1A may mediate the downstream effects of the metabolic syndrome.

The two major strengths of our study are its size and ethnic diversity. These strengths allowed for improved power to detect novel CpG-blood lipid trait associations and to robustly explore the generalizability of the findings across multiple racial/ethnic groups. Our study has limitations in other respects some of which are common to all EWAS studies of blood. First, it remains unclear whether epigenetic changes in blood cells serve as a good surrogate of changes in the most relevant tissues controlling blood lipid levels. Establishing a reliable surrogate tissue, interindividual epigenetic differences must not only correlate between blood and the tissue of interest but also exposures must induce similar changes to both tissues. Such evidence either does not yet exist or is incomplete for most trait-exposure combinations. While circulating leucocytes are likely to exert at least partial direct control over blood lipid levels, other relevant tissues/cells that we did not examine include hepatocytes, adipocytes, and enterocytes. A second limitation common to all EWAS of circulating leucocytes includes the potential for findings to not be truly reflective of a chronic alteration of transcriptional regulation from environmental perturbations but rather residual.
confounding due to persistent cell subtype proportional heterogeneity despite the application of statistical deconvolution techniques [46,53,54]. Lastly, the cross-sectional design of our study makes it difficult to determine the directionality of our associations [54]. We attempted to provide additional evidence of directionality using Mendelian randomization techniques [55], but the power of our MR analyses was limited due to the lack of availability of genetic data for many cohorts limiting sample size combined with a lack of strong genetic instruments for many of CpGs examined [56]. In addition, the known shared genetic background of HDL, LDL, and TG introduces the possibility of biases due to pleiotropy in our MR analysis.

In conclusion, we identified 373 novel CpG-lipid traits associations through the largest multi-ethnic EWAS to date. We found a high concordance of the direction of effects for all 3 lipids traits across racial/ethnic groups and a high correlation of effects for HDL and TG with 30 CpGs—including 12 novel CpGs—reaching stringent statistical significance in at least 2 racial/ethnic groups. A large majority of implicated genes do not overlap with lipid loci identified to date through GWAS although many loci associated with HDL and TG in >2 racial/ethnic groups have been associated with related cardio-metabolic traits in previous EWAS. We provide some limited insights on mechanism of association through our mQTL, eQTM, and MR analyses but additional studies are needed before firm conclusions can be made on the causality and mechanisms behind a large majority of the associations we observed.

Methods

Study populations. A total of 15 cohorts (N = 16,265) from the epigenetics working group in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium participated in this study. These included the Old Order Amish (OOA), Atherosclerosis Risk in Communities (ARIC), Bogalusa Heart Study (BHS), Cardiovascular Health Study (CHS), Framingham Heart Study (FHS), Genetic Epidemiology Network of Arteriopathy (GENOA), Genetics of Lipid Lowering Drugs and Diet Network (GOULD), Hypertension Genetic Epidemiology Network (HyperGEN), Cooperative health research in the Region of Augsburg (KORA), Normative Aging Study (NAS), Prospective Investigation of Vascularity of Uppsala Elders Study (PIVUS), Rotterdam Study (RS), UK Adult Twins Registry (TwinsUK), Women’s Health Initiative (WHI) Hormone Replacement 23 (WHI-B23), and the Women’s Health Initiative Epigenetic Mechanisms of PM-Mediated CVD (WHI-EMPC) cohorts. Four cohorts, BHS, CHS, WHI-B23, and WHI-EMPC, examined more than one racial/ethnic group. The total number of cohorts in the European, African, and Hispanic study populations is 12 (N = 11,114), 9 (N = 4,452), and 2 (N = 699), respectively. The participating cohorts are described in the Supplementary materials. All study metadata were written informed consent from participants and were approved by local institutional review boards and ethics committees.

Lipid measurements. High-density-lipoprotein (HDL, mg per dl) and triglycerides (TG, mg per dl) were directly measured in blood samples taken from participants who did not fast for at least 8 h. We also excluded outliers as defined by >5 standard deviations from the mean of blood lipid in each cohort. To reduce skewness, HDL and triglycerides were natural log-transformed.

DNA methylation measurement, QC, and normalization. DNA methylation was produced by investigators from each cohort independently. Levels were measured from peripheral blood leukocytes isolated from whole blood in all studies except GOLDN where only CD4 + T cells were examined. The EZ DNA Methylation Gold Kit (Zymo Research, Orange CA) was used for bisulfite conversion. The Illumina® Infinium HumanMethylation450 BeadChip and the Illumina BeadXpress reader were used to perform the methylation assays. Either the SWAN²⁵ method in the minfi package, the Beta Mixture Quantile method (BMQ)⁶⁰, the DASEN method in the wateRmelon R package⁶⁰, or the GenomeStudio® Methylation Module was used for pre-processing and normalization of the data in each cohort (Supplementary Data 1). For each CpG site, a beta-value was calculated representing the percent methylation at that CpG site. We used an annotation file provided on the Illumina website to annotate CpGs to genes. CpGs were annotated to genes by Illumina using the following rules: those located within 1500 bp upstream of transcription start site (TSS)1500, TSS200, 5′UTR, 1′ exon, gene body, or 3′UTR of a gene were annotated to that gene. All other intergenic CpGs were not included as a gene. To reduce technical batch effects, phenotypic information were added as random effects in the association analyses. To reduce confounding from cellular heterogeneity⁶¹, we estimated cell proportions using Houseman’s method⁶² in each subject and used these proportions as covariates in the association analyses.

Overlap with prior related genome-wide association studies. To identify the overlap between EWAS CpGs and genome-wide association studies (GWAS) SNPs of lipid traits, we performed an approach. (1) Identify a SNP in 4,452) that assembles and GWAS CpG pair located within 1 Mbp; (2) Identify a GWAS SNP and an EWAS CpG pair located within 10 Mbp; and (3) Identify a GWAS SNP and an EWAS CpG pair annotated to the same gene. SNPs were annotated to a gene if it is located within the transcript boundary of a protein-coding gene or a nearest gene if it is located outside of the transcript boundary. The CpGs were annotated to a gene if it is located within base pairs of the transcription start site, 5′-UTR, gene body, or 3′-UTR using the Illumina annotation file. Intergenic CpGs were not considered for this approach.

Methylation quantitative trait loci. We investigated the association between either imputed or genotyped SNPs located within 25 kb upstream or downstream of each CpG and DNA methylation levels to identify cis-acting methylation quantitative trait loci (mQTL). For imputed SNP data, we restricted the mQTL analysis to SNPs with a good quality imputation (IMPUTE info > 0.4 or MACH r2 > 0.3). Subjects taking lipid lowering medications were excluded from this analysis. Beta-values of DNA methylation levels were inverse-normal transformed and compared for each age, sex, smoking (current/former/never), BMI, at least 4 SNP PCs, cell proportions (WBC count and/or estimated WBC proportions (granulocytes as a reference), and technical covariates (plate, row, and column as random effects) (two-sided test). Family information was also included as a random effect if a cohort was a family-based study. We then regressed the residuals on each SNP interests of stratified by racial/ethnic group.

Expression quantitative trait methylation. We extracted associations between DNA methylation levels and gene expression in blood (expression quantitative trait methylation or eQTM) from a pre-existing investigation involving 4278 participants of the FHS (2,726 offspring cohort participants and 1552 third-generation cohort participants)⁶³. For gene expression, whole blood was collected in PAXgene tubes (PreAnalytiX, Hombrechtikon, Switzerland). Total RNA was extracted using the whole blood RNA System Kit (Qiagen, Venlo, Netherlands) and mRNA expression profiling was assessed using the Affymetrix
Human Exon 1.0 ST GeneChip platform (Affymetrix Inc, Santa Clara, CA), which contains more than 5.5 million probes targeting the expression of 17,873 genes. The RefSet multi-array Average (RMA) package64 as used to normalize the gene expression values and remove any technical or spurious background variation.

Linear mixed effects model was used to assess associations between residuals of DNA methylation levels and residuals of gene expression levels (two-sided test).

We first regressed out the fixed effects of age, sex, white blood cell type proportions as estimated using the Houseman method65 in the following covariates (sample storage time, RNA integrity number), and the first component of a principal component analysis and the random effects of amplification batch in the gene expression levels.

Next, we regressed out the fixed effects of age, sex, Houseman’s white blood cell type proportions, and the first two PCs of a principal component analysis and the random effect of batch (plate) from the DNA methylation levels. We then applied an additional adjustment for 25 surrogate variables to both the gene expression and DNA methylation data before the methylation-gene expression association analysis. Surrogate variables were calculated using the gene expression data (residualized for PC1, and technical variables as fixed effects and amplification batch as random effect) and the DNA methylation data (residualized for PC1 and PC2 as fixed effects and batch as random effect) adjusting for age, sex, and Houseman’s blood cell type prediction.

Mendelian randomization approach. We applied a bi-directional Mendelian randomization approach to shed light on the causal nature of associations we identified between blood lipid levels and DNA methylation levels. We applied a bi-directional Mendelian randomization analysis in each study among participants not taking lipid lowering medication followed by a meta-analysis of analogous results in each study. Each lipid measure was regressed on the corresponding GRS adjusting for age, sex, smoking, BMI, at least 4 SNP PCs, and a family effect if a cohort was a family-based study. The DNA methylation level of each CpG was also regressed on the corresponding GRS adjusting for age, sex, smoking, BMI, at least 4 SNP PCs, cell proportions, technical batch effects, and a family effect if a cohort was family-based study. The study-specific estimates and standard errors of the GRS-lipid associations were used as input for a meta-analysis to obtain overall estimates for GRS-lipid and GRS-CpG associations. The overall estimates were then used to assess the causal effect of each lipid measure to a CpG with a Mendelian randomization approach.

Secondly, we implemented Mendelian randomization65,66 method to study the effect of DNA methylation on blood lipid levels. The identified significant mQTL SNPs were used as an instrument for the CpGs of interest. Study participants taking any lipid lowering medication were excluded in the Mendelian randomization analysis. Beta-values of DNA methylation levels were inverse-normal transformed and regressed on age, sex, smoking (current, former, or never), BMI, at least 4 SNP PCs, cell proportions (WBC count and/or estimated WBC proportions as a reference), and technical covariates (plate, row, and column as random effects). Family information was also included as a random effect if a cohort is a family-based study. Then the residuals were regressed on each SNP located 25 kb up and downstream of the CpG site. The mQTL analyses were performed separately for each racial/ethnic group. Each SNPs were also searched to identify associations with lipids levels in the previous GWAS results from Global Lipids Genetics Consortium (GLGC, http://cog.sph.umich.edu/willer/public/ lipids2013)2. The estimates and standard errors of the SNP-CPG (obtained from our samples) and SNP-lipid associations (obtained from the GLGC GWAS results) were used as input for inverse-weighted MR and MR-egger methods to assess the causal effect of each CpG to a lipid measure.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The methylation QTL association results as well as the full summary statistics for the meta-analysis of the epigenome wide association study performed within each and across all racial/ethnic groups combined for all four models and all three lipid traits are available at [https://doi.org/10.5061/dryad.q5dts98](https://doi.org/10.5061/dryad.q5dts98). All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding authors upon reasonable request. A reporting summary for this Article is available as a Supplementary Information file.

Code availability

The following freely available software was used for analyses: R (V.3.4.4 and later, https://www.r-project.org/) including minff, watermelon, and qmap packages; Python 2.7.5 (https://www.python.org/downloads/); METASOFT (V.2.8.1, http://genetics.cs.ucla.edu/ meta_temo/doc/).

Received: 11 July 2019; Accepted: 25 May 2021; Published online: 28 June 2021

References

1. Lozano, J. V. et al. Serum lipid profiles and their relationship to cardiovascular disease in the elderly: the PREV-ICTUS study. Curr. Med Res Opin. 24, 659–670 (2008).
2. Petitti, D. B. et al. Serum lipids and glucose control: the SEARCH for Diabetes in Youth study. Arch. Pediatr. Adolesc. Med. 161, 159–165 (2007).
3. Trevisan, R., Dodesini, A. R. & Lepore, G. Lipids and renal disease. J. Am. Soc. Nephrol. 17, S145–S147 (2006).
4. Presecki, P. et al. Serum lipid levels in patients with Alzheimer’s disease. Coll. Antropol. 35, 115–120 (2011).
5. Chandler, P. D. et al. Lipid biomarkers and long-term risk of cancer in the Women’s Health Study. Am. J. Clin. Nutr. 103, 1397–1407 (2016).
6. His, M. et al. Prospective associations between serum biomarkers of lipid metabolism and overall, breast and prostate cancer risk. Eur. J. Epidemiol. 29, 119–132 (2014).
7. Global Lipids Genetics, C. et al. Discovery and refinement of loci associated with lipid levels. Nat. Genet. 45, 1274–1283 (2013).
8. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci with lipid levels. Nature 466, 707–713 (2010).
9. Liu, D. J. et al. Exome-wide association study of plasma lipids in >300,000 individuals. Nat. Genet. 49, 1758–1766 (2017).
10. Gao, C. et al. Exome sequencing identifies genetic variants associated with circulating lipid levels in Mexican Americans: The Insulin Resistance Atherosclerosis Family Study (IRASFS). Sci. Rep. 8, 5603 (2018).
11. Fraga, M. F. et al. Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl Acad. Sci. USA 102, 10694–10699 (2005).
12. Bjorncs, H. T. et al. Intra-individual change over time in DNA methylation with familial clustering. JAMA 299, 2877–2883 (2008).
13. Joehanes, R. et al. Epigenetic signatures of cigarette smoking. Circ. Cardiovasc Genet. 9, 436–447 (2016).
14. Wahl, S. et al. Epigenome-wide association study of body mass index, and the adipocyte transcriptome as outcomes of adiposity. Nature 541, 81–86 (2017).
15. Ma, J. et al. A peripheral blood DNA methylation signature of hepatic fat reveals a potential causal pathway for non-alcoholic fatty liver disease. Diabetes (2019).
16. 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 (2014).
17. Chambers, J. C. et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. Lancet Diabetes Endocrinol. 3, 526–534 (2015).
18. Chu, A. Y. et al. Epigenome-wide association studies identify DNA methylation associated with kidney function. Nat. Commun. 8, 1286 (2017).
19. Richard, M. A. et al. DNA methylation analysis identifies loci for blood pressure regulation. Am. J. Hum. Genet. 101, 888–902 (2017).
20. Ligthart, S. et al. DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. Genome Biol. 17, 255 (2016).
21. Irvin, M. R. et al. Epigenome-wide association study of fasting blood lipids in the Genetics of Lipid-Lowering Drugs and Diet Network study. Circulation 130, 565–572 (2014).
22. Dekkers, K. F. et al. Blood lipids influence DNA methylation in circulating cells. Genome Biol. 17, 138 (2016).
23. Sayols-Baixeras, S. et al. Identification and validation of seven new loci showing differential DNA methylation related to serum lipid profile: an epigenome-wide approach. The REGICOR study. Hum. Mol. Genet. 25, 4536–4565 (2016).
24. Braun, K. V. E. et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. Clin. Epigenetics 9, 15 (2017).
25. Braun, K. V. et al. The role of DNA methylation in dyslipidemia: a systematic review. Prog. Lipid Res. 64, 178–191 (2016).
26. Han, B. & Eskin, E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am. J. Hum. Genet. 88, 308–3098 (2011).
27. Klarin, D. et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat. Genet. 50, 1514–1523 (2018).
28. Hoffmann, T. J. et al. A large electronic-health-record-based genome-wide study of serum lipids. Nat. Genet. 50, 401–413 (2018).
29. Gaunt, T. R. et al. Systematic identification of genetic influences on methylation across the human life course. Genome Biol. 17, 61 (2016).

30. Johnson, R. C. et al. Systematic genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. Genome Biol. 18, 16 (2017).

31. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum. Mol. Genet. 23, R89–R98 (2014).

32. Retlon, C. L. & Davey Smith, G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J. Epidemiol. 41, 161–176 (2012).

33. Bhat, K. S. et al. Association of DNA methylation at loci with non-obese and obese patients. PLoS Genet. 28 (2018).

34. Mittelstrass, K. & Waldenberger, M. DNA methylation in human metabolic and related diseases. Curr. Opin. Lipid. 29, 116–124 (2018).

35. Crujeiras, A. B. et al. DNA methylation map in circulating leukocytes mirrors the preparative ultracentrifuge. J. Intern. Med. 270, 729 (2014).

36. Huang, Y. T. et al. Epigenome-wide association studies. Epigenetics 6, 264 (2003).

37. Yang, C. S. et al. Ubiquitin modification by the E3 Ligase/ADP-Ribosyltransferase Dtx3L/Parp9. Mol. Cell. 66, 503–516 e5 (2017).

38. Verheugd, P. et al. Regulation of DNA methylation across the human life course. Nature Genet. 48 (2016).

39. Barajas-Olmos, F. et al. Altered DNA methylation in liver and adipose tissues from non-obese and obese patients. PLoS Genet. 8, e1002522 (2012).

40. Inouye, M. et al. An immune response network associated with blood lipid metabolism and related diseases. Circ. Cardiovasc Genet. 11, 1197–1207 (2018).

41. Peng, P. et al. A preliminary study of the relationship between promoter methylation of the ABCG1, GALNT2 and HMGCR genes and coronary heart disease. PLoS ON 9, e1002625 (2014).

42. Aslibekyan, S. et al. CPT1A methylation is associated with plasma adiponectin. Nutr. Metab. Cardiovasc. Dis. 27, 225–233 (2017).

43. Das, M. et al. Association of DNA Methylation at CPT1A Locus with Metabolic Syndrome in the Genotypes of Lipid Lowering Drugs and Diet Network (GOLDN) Study. PLoS ON 11, e0145789 (2016).

44. Barsnes, Ó. et al. Correspondence of DNA methylation between blood and adipose tissues with King tuberculosis. Schizophr. Bull. 41, 302 (2015).

45. Pfeiffer, L. et al. DNA methylation of lipid-related genes affects blood lipid levels. Circ. Cardiovasc Genet. 8, 334–342 (2015).

46. Michels, K. B. et al. Recommendations for the design and analysis of epigenome-wide association studies. Epigenetics 6, 4 (2011).

47. Huang, Y. T. et al. Epigenome-wide association studies. Hum. Mol. Genet. 283 (2016).

48. Walton, E. et al. Correspondence of DNA methylation between blood and adipose tissues with King tuberculosis. Schizophr. Bull. 41, 302 (2015).

49. Crujeiras, A. B. et al. DNA methylation map in circulating leukocytes mirrors the preparative ultracentrifuge. J. Intern. Med. 270, 729 (2014).

50. Barajas-Olmos, F. et al. Altered DNA methylation in liver and adipose tissues from non-obese and obese patients. PLoS Genet. 8, e1002522 (2012).

51. Inouye, M. et al. An immune response network associated with blood lipid metabolism and related diseases. Circ. Cardiovasc Genet. 11, 1197–1207 (2018).

52. Barajas-Olmos, F. et al. Altered DNA methylation in liver and adipose tissues derived from individuals with obesity and type 2 diabetes. BMC Med. Genet. 19, 28 (2018).

53. Michaëlis, D., Meitinger, T. & Candiano, G. epigenome-wide association studies. Nature Rev. Genet. 19, 185–189 (2015).

54. Nomura, H. & Sato, K. Altered DNA methylation in liver and adipose tissues derived from individuals with obesity and type 2 diabetes. Diabetes Metab. J. 44, 231–236 (2010).

55. Barajas-Olmos, F. et al. Altered DNA methylation in liver and adipose tissues derived from individuals with obesity and type 2 diabetes. BMC Med. Genet. 19, 28 (2018).

56. Inouye, M. et al. An immune response network associated with blood lipid levels. PLoS Genet. 6, e1001113 (2010).

57. Nath, A. P. et al. An interaction map of circulating metabolites, immune gene networks, and their genetic regulation. Genome Biol. 18, 146 (2017).

58. Houseman, E. A. et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinform. 13, 86 (2012).

59. Birney, E., Smith, G. D. & Greally, J. M. Epigenome-wide association studies and the interpretation of disease-omics. PLoS Genet. 12, e1006105 (2016).

60. Bhat, K. S. et al. Association of DNA methylation at loci with non-obese and obese patients. PLoS Genet. 28 (2018).

61. Barajas-Olmos, F. et al. Altered DNA methylation in liver and adipose tissues derived from individuals with obesity and type 2 diabetes. BMC Med. Genet. 19, 28 (2018).

62. Chen, Y. A. et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8, 203–209 (2013).

63. Wesling, B., Roeder, K. & Wasserman, L. Mendelian correction, a new approach to genetic-based association studies. Theor. Popul. Biol. 60, 155–166 (2001).

64. Retlon, C. L. & Davey Smith, G. Two-step epigenetic Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J. Epidemiol. 44, 512–525 (2015).

65. Greenland, S. An introduction to instrumental variables for epidemiologists. Int J. Epidemiol. 29, 722–729 (2000).

Acknowledgements

We thank our Amish community and research volunteers for their long-standing partnership in research, and acknowledge the dedication of our Amish liaisons, field workers etc., without which none of this work would have been possible. We thank the staff and participants of the ARIC study for their important contributions. We would also like to thank the families that participated in the GENOA study. The authors are grateful to the Rotterdam Study participants, the staff involved with the Rotterdam Study and the participating general practitioners and pharmacists. The lipid management and the Illumina 450 K (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. We thank Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marjin Verkerk, and Lisette Stoek PhD for their help in creating the methylation database. We thank Ms. Mila Jhamai, Ms. Sarah Higgins, Marjolein Peters, Ms. Marjin Verkerk and Jeroen Jovou, MS for the help in creating the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of the Illumina 450 K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. We thank Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marjin Verkerk, and Lisette Stoek PhD for their help in creating the methylation database. We thank Ms. Mila Jhamai, Ms. Sarah Higgins, Marjolein Peters, Ms. Marjin Verkerk and Jeroen Jovou, MS for the help in creating the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of the Illumina 450 K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. We thank Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marjin Verkerk, and Lisette Stoek PhD for their help in creating the methylation database.
Foundation (grant no. 2013.0126), Swedish Heart-Lung Foundation (grant no. 20140422), and Swedish Diabetes Foundation (grant no. 2013-024). The Rotterdam Study EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIO5) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). The generation and management of RNA-expression array data for the Rotterdam Study was executed and funded by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The Women's Health Initiative data, WHI-BA23 and WHI-EMPC, were generated through an NHLBI Broad Agency Announcement contract (HHSN268201300006C) and a National Institute of Environmental Health Sciences grant (ROI-E5028836). The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100000C, HHSN268201100000C, HHSN268201100003C, HHSN268201100004C, and HHSN272101100004C.

Author contributions
M.A.J., M.M., M.R.I., and T.L.A. contributed to study design. M.A.J., R.W., R.G., R.J., E.S., X.Z., K.V.E.B. AND, A.K.H., T.Z., E.C.M., I.S., T.M.B., J.A.B., M.E.M., J.R.O., C.Y., P.S., W.Z., E.R.W., S.G. contributed to cohort-specific data analyses. M.A.J. contributed to meta-analyses of EWAS and mQTL, and Mendelian randomization analyses. M.A.J. and T.L.A. contributed to interpretation of the results and writing of manuscript. M.M., R.W., R.G., K.V.E.B., T.M.B., B.M.P., E.I., M.F., J.T.B., L.L., M.W. contributed to critical review of manuscript. E.R., M.G., W.G., L.P., M.M.N., T.M., C.G., A.P., E.B.W., J.A.S., K.D., J.v.M., A.U., M.A.I., M.G., D.Z., M.L., S.L., T.D.S., Y.I.C., C.D., K.R.S., D.M.A., S.H., P.S.T., S.K., B.M.P., N.S., J.T.B., E.I., W.C., A.D., D.A., M.W., L.H., E.A.W., A.B., D.L., M.F., M.R.I., and T.L.A. contributed to cohort design and management, and data collection.

Competing interests
Alan R Shuldiner is an employee of Regeneron Pharmaceuticals, Inc. Bruce M Pasy serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Kim Valeska Emilie Braun works in Erasmus/AG; a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.), Metagenics Inc. and AXA.

All other authors declare no competing interests. The funders had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review or approval of the manuscript.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-23899-y.

Correspondence and requests for materials should be addressed to M.-A.J. or T.L.A.

Peer review information Nature Communications thanks Jenny van Dongen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA. 2Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. 3Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA. 4Department of Cardiology, Boston Children's Hospital, Boston, MA, USA. 5Research Unit Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany. 6Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany. 7Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA. 8Hebrew SeniorLife, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. 9The Brown Foundation Institute of Molecular Medicine, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA. 10Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands. 11Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. 12Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. 13Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA. 14Department of Twin Research and Genetic Epidemiology, School of Life Course Sciences, King’s College London, London, UK. 15GENYO, Center for Genomics and Oncological Research Pfizer/University of Granada/Andalusian Regional Government, Granada, Spain. 16Department of Population Health Sciences, University of Utah, Salt Lake City, UT, USA. 17Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA. 18Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. 19Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA. 20Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. 21School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA. 22Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA. 23Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. 24IBE, Faculty of Medicine, LMU Munich, Munich, Germany. 25Institute of Medical Biostatistics, Epidemiology and Informatics (IMBE), University Medical Center, Johannes Gutenberg University, Mainz, Germany. 26Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Munich, Germany. 27Institute of Human Genetics, Technical University Munich, Munich, Germany. 28Survey Research Center, Institute for Social Research, Ann Arbor, MI, USA. 29Department of Internal Medicine, Rush University Medical Center, Chicago, IL, USA. 30Department of Biostatistics, Erasmus MC University Medical Center, Rotterdam, The Netherlands. 31School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, USA. 32Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China. 33Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute, and Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, USA. 34HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA. 35Department of Human Genomics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. 36Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, USA. 37VA Palo Alto Healthcare System, Palo Alto, CA, USA.
Alto, CA, USA. 38Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA. 39Cardiovascular Health Research Unit, Departments of Epidemiology, Medicine, and Health Services, University of Washington, Seattle, WA, USA. 40Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. 41Cardiovascular Health Research Unit, Division of Cardiology, University of Washington, Seattle, WA, USA.
42Stanford Diabetes Research Center, Stanford University, Stanford, CA, USA. 43Department of Biostatistics and Epidemiology, MRC Centre for Environment and Health, School of Public Health, Imperial College, London, UK. 44Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. 45Department of Medicine, University of North Carolina, Chapel Hill, NC, USA. 46Department of Environmental Health Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, USA. 47Department of Environmental Health Sciences, Columbia University Mailman School of Public Health, New York, NY, USA. 48Framingham Heart Study, Framingham, MA, USA.

email: mina.jhun82@gmail.com; tassimes@stanford.edu