

RESEARCH

Open Access



# DNA methylation and lipid metabolism: an EWAS of 226 metabolic measures

Monica del C. Gomez-Alonso<sup>1,2†</sup>, Anja Kretschmer<sup>1,2†</sup>, Rory Wilson<sup>1,2</sup>, Liliame Pfeiffer<sup>1,2</sup>, Ville Karhunen<sup>3,10</sup>, Ilkka Seppälä<sup>4</sup>, Weihua Zhang<sup>3,5</sup>, Kirstin Mittelstraß<sup>1,2</sup>, Simone Wahl<sup>1,2</sup>, Pamela R. Matias-Garcia<sup>1,2</sup>, Holger Prokisch<sup>6,7</sup>, Sacha Horn<sup>1,2</sup>, Thomas Meitinger<sup>6,7,8</sup>, Luis R. Serrano-Garcia<sup>1,2,9</sup>, Sylvain Sebert<sup>10</sup>, Olli Raitakari<sup>11,12,13</sup>, Marie Loh<sup>3,14</sup>, Wolfgang Rathmann<sup>15,16</sup>, Martina Müller-Nurasyid<sup>17,18,30</sup>, Christian Herder<sup>16,19,20</sup>, Michael Roden<sup>16,19,20</sup>, Mikko Hurme<sup>21</sup>, Marjo-Riitta Jarvelin<sup>3,10,22,23</sup>, Mika Ala-Korpela<sup>10,24,25,26</sup>, Jaspal S. Kooner<sup>5,27,28,29</sup>, Annette Peters<sup>2</sup>, Terho Lehtimäki<sup>4</sup>, John C. Chambers<sup>3,5,14,28,29</sup>, Christian Gieger<sup>1,2</sup>, Johannes Kettunen<sup>10,25,26</sup> and Melanie Waldenberger<sup>1,2,8\*</sup> 

## Abstract

**Background:** The discovery of robust and trans-ethnically replicated DNA methylation markers of metabolic phenotypes, has hinted at a potential role of epigenetic mechanisms in lipid metabolism. However, DNA methylation and the lipid compositions and lipid concentrations of lipoprotein sizes have been scarcely studied. Here, we present an epigenome-wide association study (EWAS) ( $N = 5414$  total) of mostly lipid-related metabolic measures, including a fine profiling of lipoproteins. As lipoproteins are the main players in the different stages of lipid metabolism, examination of epigenetic markers of detailed lipoprotein features might improve the diagnosis, prognosis, and treatment of metabolic disturbances.

**Results:** We conducted an EWAS of leukocyte DNA methylation and 226 metabolic measurements determined by nuclear magnetic resonance spectroscopy in the population-based KORA F4 study ( $N = 1662$ ) and replicated the results in the LOLIPOP, NFBC1966, and YFS cohorts ( $N = 3752$ ). Follow-up analyses in the discovery cohort included investigations into gene transcripts, metabolic-measure ratios for pathway analysis, and disease endpoints. We identified 161 associations ( $p$  value  $< 4.7 \times 10^{-10}$ ), covering 16 CpG sites at 11 loci and 57 metabolic measures. Identified metabolic measures were primarily medium and small lipoproteins, and fatty acids. For apolipoprotein B-containing lipoproteins, the associations mainly involved triglyceride composition and concentrations of cholesterol esters, triglycerides, free cholesterol, and phospholipids. All associations for HDL lipoproteins involved triglyceride measures only. Associated metabolic measure ratios, proxies of enzymatic activity, highlight amino acid, glucose, and lipid pathways as being potentially epigenetically implicated. Five CpG sites in four genes were associated with differential expression of transcripts in blood or adipose tissue. CpG sites in *ABCG1* and *PHGDH* showed associations with metabolic measures, gene transcription, and metabolic measure ratios and were additionally linked to obesity or previous myocardial infarction, extending previously reported observations.

\*Correspondence: waldenberger@helmholtz-muenchen.de

†Monica del C. Gomez-Alonso and Anja Kretschmer shared first authors

<sup>1</sup> Research Unit of Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München German Research Center for Environmental Health, Ingolstaedter Landstraße 1, 85764 Neuherberg, Germany

Full list of author information is available at the end of the article



© The Author(s) 2020. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusion:** Our study provides evidence of a link between DNA methylation and the lipid compositions and lipid concentrations of different lipoprotein size subclasses, thus offering in-depth insights into well-known associations of DNA methylation with total serum lipids. The results support detailed profiling of lipid metabolism to improve the molecular understanding of dyslipidemia and related disease mechanisms.

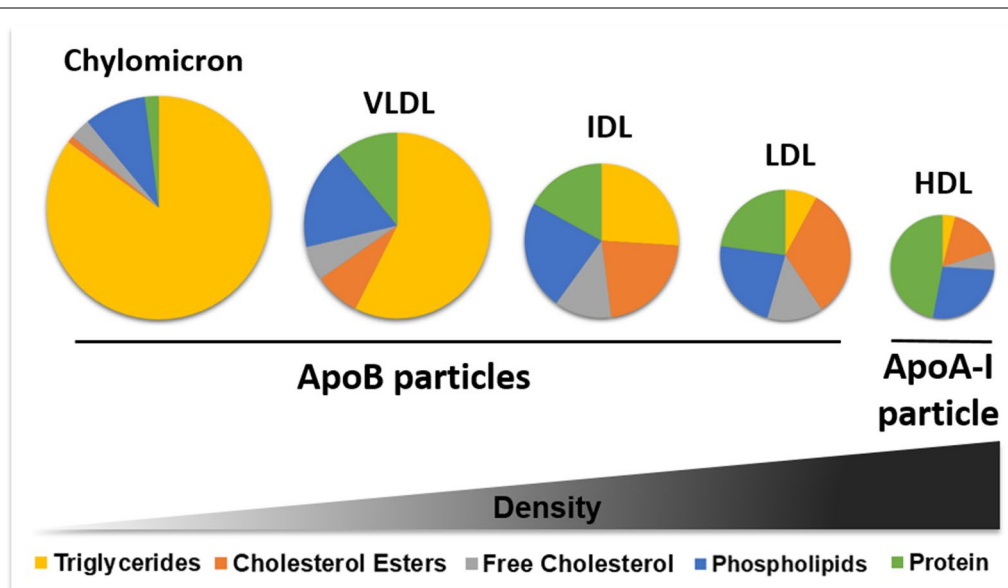
**Keywords:** CpG site, VLDL, LDL, HDL, Lipoprotein sizes, Lipoprotein composition, Fatty acids, Myocardial infarction, Obesity, NMR

## Background

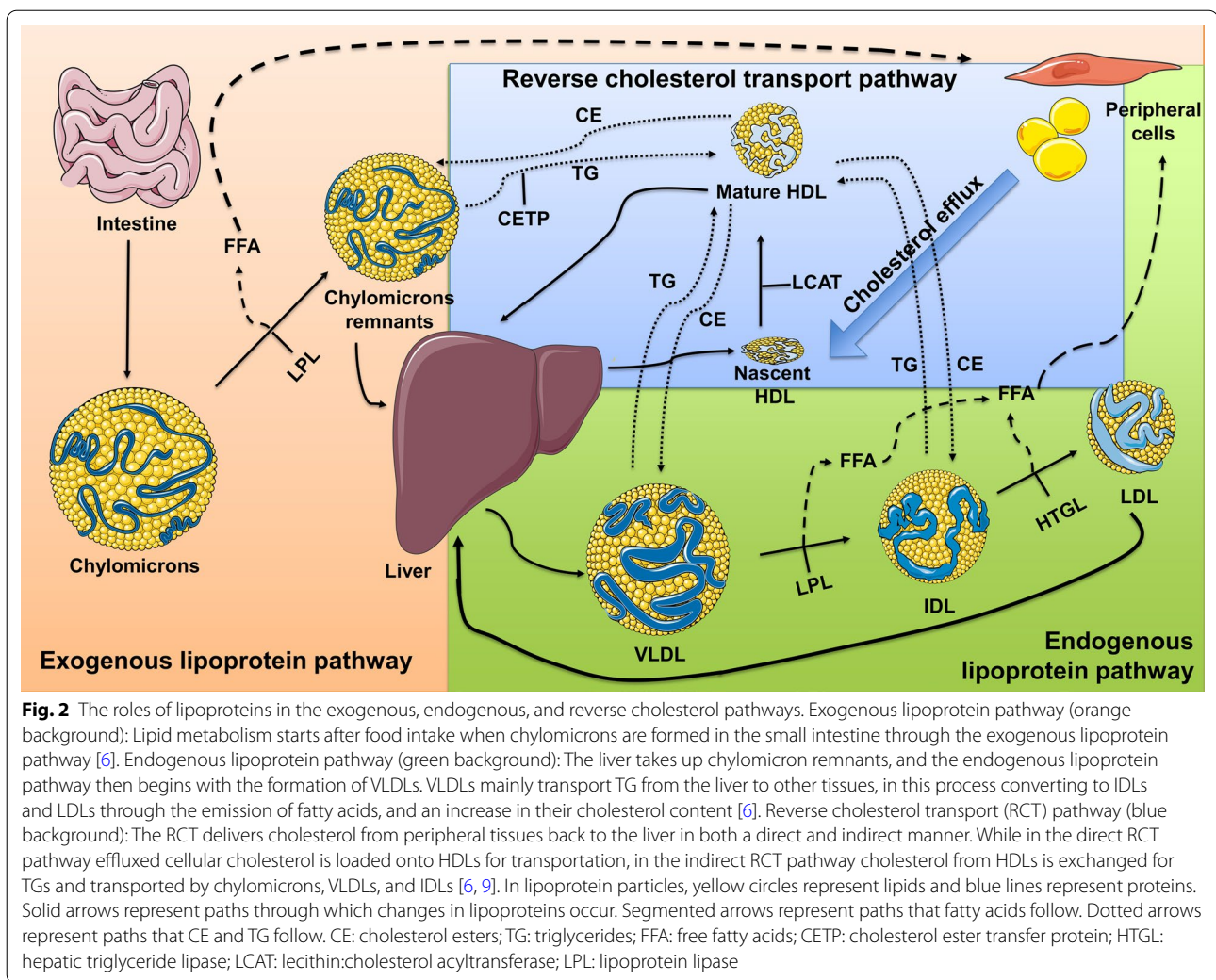
Dyslipidemia refers to abnormal levels of one or more lipids, such as plasma cholesterol, high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), and/or plasma triglycerides (TG) in blood, leading to complex cardiometabolic diseases such as atherosclerosis, type 2 diabetes (T2D), or myocardial infarction (MI) [1–5]. Due to their poor solubility in blood, lipids are transported in lipoprotein particles that can be categorized according to their size, density, and composition as shown in Fig. 1 [6–8]. Lipoproteins are main players of the exogenous, endogenous, and reverse cholesterol transport pathways, thus contributing to lipid metabolism as illustrated in Fig. 2 [6, 9]. The smallest lipid molecules contained in lipoprotein particles are saturated and unsaturated fatty acids. Unsaturated fatty acids consist of monounsaturated fatty acids (MUFA) and polyunsaturated fatty acids (PUFA). Omega-3 PUFAs (e.g., docosahexaenoic acid

(DHA)) have been linked to prevention of metabolic disorders, whereas for omega-6 PUFAs (e.g., 18:2 linoleic acid (LA)) inconsistent results exist [10]. Omega-3 to omega-6 FAs ratios and branched chain amino acids, such as isoleucine, are associated with metabolic outcomes [11–13].

There is mounting evidence that epigenetic mechanisms play an important role in the regulation of metabolic phenotypes and in other complex diseases [14–25], thus representing a possible therapeutic target [26–28]. While DNA methylation studies have highlighted several robustly replicated methylation markers of cardiometabolic phenotypes [14–23], the full causal interplay is unknown. However, it has been proposed that most causal changes in methylation are a consequence rather than a cause of dyslipidemia and body mass index (BMI), therefore indicating that methylation may be more a biomarker of prevalent conditions rather than a predictor of incident conditions [20, 24, 25, 29–32].



**Fig. 1** Lipoprotein composition. Lipoproteins can be categorized according to their size, composition, and density. The lattermost classifies them into chylomicrons, very low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL), or high-density lipoproteins (HDL). The less dense the particles are, the larger they are in size and the more lipids they contain [6–8]. Chylomicrons, VLDLs, IDLs, and LDLs are known as apolipoprotein B-containing particles, due to their main structural protein, while HDLs contain apolipoprotein A-I. Lipid and protein compositions are represented in approximated percentages [7]



**Fig. 2** The roles of lipoproteins in the exogenous, endogenous, and reverse cholesterol pathways. Exogenous lipoprotein pathway (orange background): Lipid metabolism starts after food intake when chylomicrons are formed in the small intestine through the exogenous lipoprotein pathway [6]. Endogenous lipoprotein pathway (green background): The liver takes up chylomicron remnants, and the endogenous lipoprotein pathway then begins with the formation of VLDLs. VLDLs mainly transport TG from the liver to other tissues, in this process converting to IDLs and LDLs through the emission of fatty acids, and an increase in their cholesterol content [6]. Reverse cholesterol transport (RCT) pathway (blue background): The RCT delivers cholesterol from peripheral tissues back to the liver in both a direct and indirect manner. While in the direct RCT pathway effluxed cellular cholesterol is loaded onto HDLs for transportation, in the indirect RCT pathway cholesterol from HDLs is exchanged for TGs and transported by chylomicrons, VLDLs, and IDLs [6, 9]. In lipoprotein particles, yellow circles represent lipids and blue lines represent proteins. Solid arrows represent paths through which changes in lipoproteins occur. Segmented arrows represent paths that fatty acids follow. Dotted arrows represent paths that CE and TG follow. CE: cholesterol esters; TG: triglycerides; FFA: free fatty acids; CETP: cholesterol ester transfer protein; HTGL: hepatic triglyceride lipase; LCAT: lecithin:cholesterol acyltransferase; LPL: lipoprotein lipase

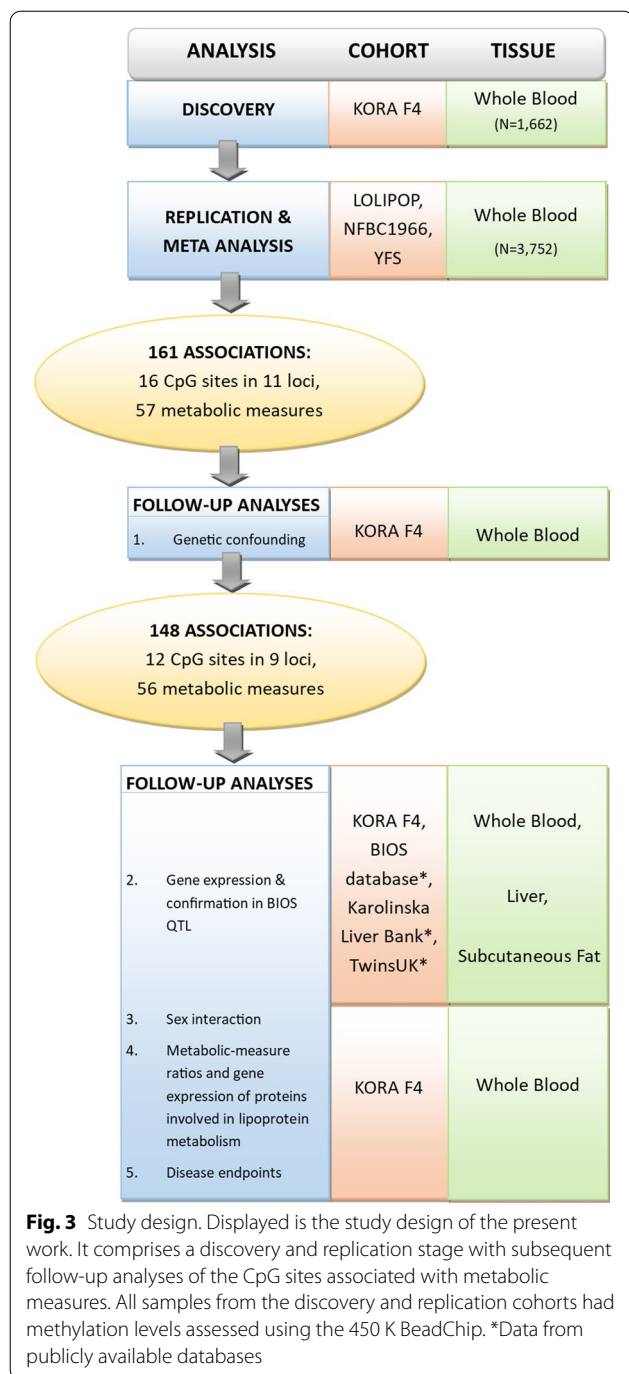
To this point, most of the epigenome-wide association studies (EWAS) on metabolic measures of lipids have used conventional clinical measures, which reflect total concentrations of lipids in serum. However, the lipid composition, lipid concentration, and particle size of lipoproteins can be associated with disease risk independently of total lipid concentrations [33]. Therefore, EWAS on detailed lipoprotein measures are warranted. A helpful tool in this regard is nuclear magnetic resonance (NMR), which allows fine profiling of lipoproteins at a large scale [34–36]. NMR has successfully identified several new markers of metabolic disease [35, 36]. Up to now, three studies, including our previous investigation, have evaluated associations between serum metabolic measures and DNA methylation—with a limited sample size or a limited number of measures—identifying several loci linked to disease response mechanisms or environmental insults [29, 37, 38]. To identify robust associations between DNA methylation and metabolic measurements,

we conducted an EWAS of 226 serum metabolic measures with subsequent replication in three independent cohorts ( $N=5414$  total). Metabolic measure-associated CpG sites were followed up for their associations with gene expression, relevant metabolic measure ratios, and disease endpoints in the discovery cohort only.

## Results

### Serum metabolic measures are associated with DNA methylation

The main goal of our study was to identify robust associations between DNA methylation and metabolic measures, thus identifying the most promising CpG sites for follow-up investigations. Therefore, associations of metabolic measures with CpG methylation were first assessed in the discovery cohort (KORA F4) and subsequently examined in three independent cohorts (LOLIPOP, NFBC1966, and YFS) (Fig. 3). Characteristics of all cohorts are shown in Table 1. Mean ages ranged



between 31.0 (NFBC1966) and 60.9 (KORA F4) years. At least in part due to their younger age at measurement, NFBC1966 and YFS participants were the healthiest, had no previous myocardial infarction, reported less hypertension, and less lipid-lowering drug intake.

The discovery stage ( $N=1662$ ) consisted of 226 epigenome-wide association studies: 226 metabolic measures (Additional file 1: Table S1) versus methylation

levels of 468151 CpG sites—a total of >100 M possible associations. The discovery EWAS of metabolic measures revealed 282 significant associations, including 274 robust associations as defined by our sensitivity analyses (Fig. 4; Additional file 2: Table S2). These 274 associations had a percentage of explained metabolite-level variance ranging from 1.2% (cg19693031 in *TXNIP* with isoleucine) to 12.4% (cg19610905 in *FADS2* with omega-3 to FA ratio), covering 24 CpG sites annotated to 12 genomic locations (Fig. 4; Additional file 2: Table S2): cg06500161 and cg27243685 in *ABCG1* (*ATP-binding cassette sub-family G member 1*); cg11024682, cg15863539, and cg20544516 in *SREBF1* (*sterol regulatory element binding transcription factor 1*); cg00574958 in *CPT1A* (*carnitine palmitoyltransferase 1A*); cg19693031 in *TXNIP* (*thioredoxin interacting protein*); cg17901584 in *DHCR24* [*24-dehydrocholesterol reductase*]; cg14476101 and cg16246545 in *PHGDH* (*D-3-phosphoglycerate dehydrogenase*); cg07626482 and cg02711608 in *SLC1A5*; cg06690548 in *SLC7A11* (*solute carrier family 7 member 11*); cg07689907 in *FADS1* (*fatty acid desaturase 1*); cg00603274, cg06781209, cg11250194, cg19610905, cg25324164, cg01400685, cg27386326 in *FADS2* (*fatty acid desaturase 2*); cg03440556 and cg24503796 in *SCD* (*stearoyl-CoA desaturase*); and cg07504977 in the promoter region of *LINC00263*, a long non-coding RNA (lncRNA).

A meta-analysis ( $N=3752$ ) confirmed 161 of the 274 robust associations (58.8%), covering 16 of the 24 CpG sites found in the discovery step, annotated to 11 of the 12 genomic locations initially found, and 57 unique metabolic measures (Figs. 3 and 4; Additional file 2: Table S2). Across replication cohorts we observed consistent directions of effects, but effect sizes tended to be smaller in NFBC1966 and YFS (Additional file 3: Table S3; Additional file 4: Supplemental Results and Methods: Comparison across cohorts; Additional file 5: Table S4). For all 274 associations, the effect direction of the meta-analysis was concordant with the discovery direction, even those that were not successfully replicated, and 240 (87.6%) had a meta-analysis nominal  $p < 0.05$ . For eight CpG sites, corresponding to 4 genomic locations, associations did not replicate (cg15863539 in *SREBF1*; cg16246545 in *PHGDH*; cg07626482 and cg02711608 in *SLC1A5* (*Solute Carrier Family 1 Member 5*); cg00603274, cg06781209, cg19610905, and cg01400685 in *FADS2*) (Fig. 4; Additional file 2: Table S2). No further CpG sites were associated with additional loci coding for enzymes or proteins directly involved in lipoprotein metabolism. Follow-up analyses involved only the 161 replicated associations, except for the correlation analyses of associated metabolic measures.



**Table 1 Characteristics of the study participants**

	KORA F4 (N = 1662)	LOLIPOP (N = 2805)	YFS (N = 176)	NFBC1966 (N = 771)
Age (years)	61.0 (8.9)	51.4 (10.1)	44.2 (3.3)	31.0 (0.3)
Sex = male	49.0%	68.0%	37.5%	43.7%
BMI (kg/m <sup>2</sup> ) <sup>a</sup>	28.1 (4.8)	27.7 (4.4)	25.9 (4.5)	24.4 (3.7)
Current smokers	14.4%	8.9%	16.5%	26.2%
Ex-smokers	41.5%	8.7%	25.6%	19.2%
Never smokers	44.1%	82.4%	58.0%	53.3%
Alcohol consumption (g/day)	15.5 (20.3)	6.38 (14.2)	0.716 (0.9) <sup>b</sup>	8.7 (14.3)
C-reactive protein (mg/l)	2.5 (5.2)	4.2 (7.2)	1.4 (2.4)	1.85 (3.3)
HDL-C (mmol/l)	1.7 (0.4)	1.3 (0.3)	1.7 (0.4)	1.7 (0.5)
LDL-C (mmol/l)	2.1 (0.6)	1.9 (0.6)	2.1 (0.6)	2.0 (0.7)
Total cholesterol (mmol/l)	5.7 (1.0)	4.9 (1.1)	5.3 (1.0)	5.5 (1.4)
Total TG <sup>c</sup> (mmol/l)	1.6 (0.8)	1.5 (0.7)	1.2 (0.6)	1.2 (0.7)
VLDL-TG (mmol/l)	1.1 (0.8)	0.9 (0.6)	0.7 (0.6)	0.7 (0.5)
Total FA (mmol/l)	14.5 (2.4)	12.5 (3.0)	12.9 (2.3)	13.4 (3.5)
Hypertension <sup>e</sup>	45.5%	38.0%	19.3%	16.7%
Hospitalized myocardial infarction <sup>e</sup>	3.6%	3.9%	0.0%	0.0%
Intake of lipid-lowering drugs	16.2%	NA <sup>d</sup>	4.0%	0.0%

Displayed are characteristics of the discovery (KORA F4) and replication studies. Continuous and categorical characteristics are given as mean (standard deviation) for continuous variables and proportions for categorical variables. For every characteristic the *p* value for a difference between studies is < 0.001, given by a one-way ANOVA for continuous variables, and the Chi-square test for categorical variables

<sup>a</sup> BMI: body mass index

<sup>b</sup> Drinks per day (directly proportional to g/day)

<sup>c</sup> TG: triglycerides

<sup>d</sup> NA: variable not available

<sup>e</sup> Self-reported history

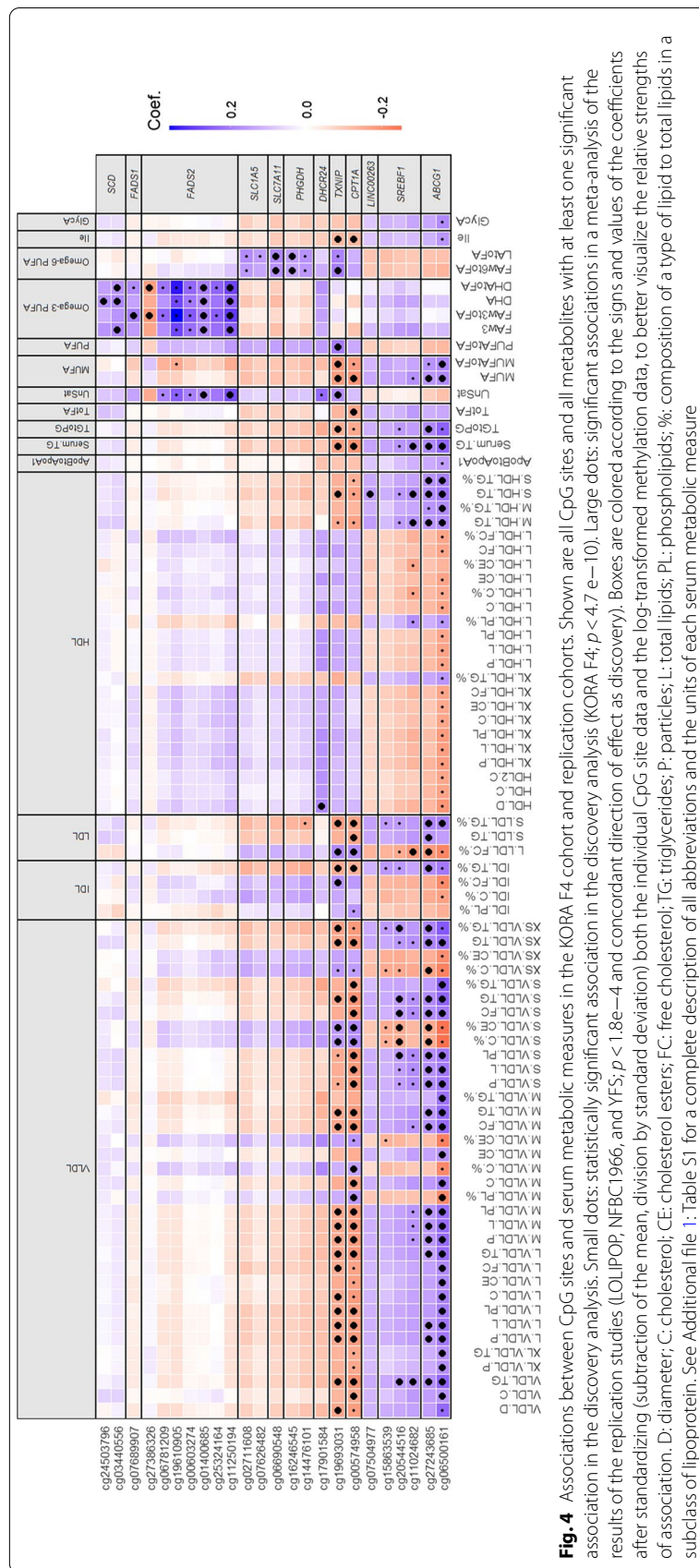
### Strong correlations between associated metabolic measures cluster DNA methylation in three groups

Strong correlations between associated metabolic measures found in the discovery cohort were observed for measures that showed the same directions of effect in associations with DNA methylation (Additional file 6: Figure S1). For instance, lipid compositions of larger VLDLs and TG measures in smaller HDLs showed strong positive correlations. We identified two groups of CpG sites showing most of their associations with lipoproteins (Fig. 4; Additional file 6: Figure S1). The first group consisted of CpG sites in *ABCG1*, *SREBF1*, and *LINC00263*. Methylation at these CpG sites showed most of their positive associations with lipid concentrations in ApoB lipoproteins, TG composition and concentration of HDLs, total serum TG, and MUFA measures. The second group consisted of CpG sites in *CPT1A* and *TXNIP*, showing most of their negative associations with lipid concentrations in ApoB lipoproteins, TG concentration of HDLs, total serum TG, MUFA measures, and isoleucine. A third group included CpG sites in *PHGDH*, *SLC7A11*, *FADS1/2*, and *SCD*. Methylation at these CpG sites only showed positive associations with PUFA measures and the degree of saturation of fatty acids.

Principal component analysis performed in the discovery cohort also points to these clusters of CpG sites. In EWAS of the top 8 principal components (PCs) of the metabolic measures data (see Additional File 4: Supplemental Results and Methods: Replication and meta-analysis; Additional file 7: Figure S2; Additional file 8: Figure S3), 6 CpG sites were found to be associated with PC1 (PC1 explaining 33% of the variance of the metabolite data) and 6 CpG sites were found to be associated with PC7 (2% explained variance) at a Bonferroni-corrected significance threshold of  $p = 0.05 / (8 \times 468151) = 1.34e-8$  (Additional file 9: Table S5). No CpG sites were associated with both PCs. All sites were found in the discovery EWAS to be significant with at least one metabolite measure, and 9 of the 12 were found in the replicated results. The CpG sites associated with PC1 were annotated to the genes *TXNIP*, *SREBF1* (2 sites), *ABCG1* (2 sites), and *CPT1A*. The CpG sites associated with PC7 were mapped to *FADS2* (5 sites) and *SCD*.

### DNA methylation is associated with fatty acids and the lipid concentrations and compositions in lipoprotein subclasses

Of the 161 replicated associations, 159 were related to lipid metabolism and 121 involved lipoprotein subclasses,



**Table 2 Results of gene expression analysis of CpG sites associated with metabolic measures**

CpG site	CHR	Probe ID	CpG annotated gene	Transcript annotated gene	Distance	Beta value	P	P Bonferroni	N	Tissue
cg16246545	1	ILMN_1704537	PHGDH	PHGDH	0	-2.77	1.44e-24	6.91e-22	634	Whole blood
cg24503796	10	ILMN_1689329	SCD	SCD	0	-1.45	2.39e-08	1.15e-05	633	Whole blood
cg03440556	10	ILMN_1689329	SCD	SCD	0	-0.57	7.43e-07	3.56e-4	609	Whole blood
cg06500161	21	ILMN_2329927	ABCG1	ABCG1	0	-2.79	2.67e-07	1.28e-4	635	Whole blood
cg20544516	17	ILMN_1663035	SREBF1	SREBF1	1542	-0.38	8.83e-10	4.59e-07	626	Subcutaneous fat
cg20544516	17	ILMN_1745806	SREBF1	PEMT	308200	0.27	1.64e-05	8.53e-3	626	Subcutaneous fat
cg20544516	17	ILMN_1811933	SREBF1	SHMT1	514197	-0.37	2.85e-09	1.48e-6	626	Subcutaneous fat
cg20544516	17	ILMN_2328986	SREBF1	SREBF1	1208	-0.28	1.26e-05	6.57e-3	626	Subcutaneous fat

Statistically significant associations between metabolic measure-associated CpG sites and expression of *cis*-transcripts in whole blood (KORA F4, Bonferroni-adjusted significance threshold  $p < 0.05/480 = 1.0e-4$ ) and subcutaneous fat (TwinsUK study,  $p < 0.05/521 = 9.6e-5$ ). "Probe ID" and "Transcript\_annotated\_gene" are the transcript ID and annotated gene from the Illumina annotation files; CHR: chromosome location of the CpG site; distance: distance between the CpG site and the transcript based on positions given in the annotation files; beta value: beta-coefficient; P: *p* value; P Bonferroni: Bonferroni-corrected *p* value; N: number of observations in the model. All coefficients are change in log<sub>2</sub>-transformed expression intensity per unit increase in methylation (beta value on 0–1 scale), except for subcutaneous fat, which is correlation assessed using the R-package rmcrr

most of these being subclasses of ApoB lipoproteins (Fig. 4; Additional file 2: Table S2). When considering sizes of ApoB lipoproteins, extra-large VLDLs showed only two associations, while medium and small VLDLs were the most associated subclasses, with almost 60 associations altogether. All subclasses of ApoB lipoproteins, except extra-large and large VLDLs, had associations involving TG composition, with some subclasses additionally having associations with respect to cholesterol, cholesterol ester, and phospholipid composition. In terms of lipid concentrations of ApoB lipoproteins, almost all types of lipids showed associations in large, medium, and small VLDLs. In extra-small VLDLs and small LDLs, the concentration of TG was the only concentration associated with methylation. All associations for HDL lipoproteins, except for one featuring the diameter of HDLs, involved either TG composition or concentration of TG in medium and small HDLs. Associations with MUFAs and serum TG showed the same CpG sites and directions of effects as the associations pertaining to the ApoB lipoproteins. The CpG sites associated with PUFAs did not show associations with lipoproteins.

#### CpG sites associated with metabolic measures have been linked to metabolic traits

To investigate common environmental and lifestyle-dependent drivers of the CpG site-metabolic measure associations, we performed searches in the EWAS Atlas and the MRC-IEU EWAS Catalog (Additional file 10: Table S6) [39, 40]. All CpG sites were found with at least one association in these databases, except for cg07689907 of *FADS1*. cg00574958 in *CPT1A* was found in the most unique publications (33 total), most

associations cited being lipid-related and metabolic-related traits such as kidney disease and gamma-glutamyl transferase. The CpG sites of *SCD* have been less cited, appearing in only seven publications total, no outcomes being obviously related to fatty acids.

#### Genetic effects on associations between DNA methylation and serum metabolic measures

We next performed follow-up analyses to test spuriousness caused by genetic confounding in the replicated associations, i.e., whether the replicated associations between DNA methylation and metabolic measures (at  $p = 4.7 e-10$ ) are driven by genetic variants in *cis* of the CpG sites. Thirty of the 161 CpG-metabolite associations became non-significant due to the influence of *cis*-SNPs (*cis*-methQTLs) (Additional file 11: Table S7). For 17 of the 30 pairs, the results did not reflect strong evidence of genetic effects on the associations, as *p* values and coefficients changed only very little when adjusting for SNPs. However, for 13 of the 30 pairs (with CpG sites all located in the *FADS* cluster), the addition of a single SNP radically decreased the magnitude of the estimated effect size (and its *p* value) of the CpG site, likely indicating the association is being confounded by genetic effects (Additional file 11: Table S7). Because each association involving CpG sites within the *FADS* region showed strong evidence of being confounded by genetic factors, we eliminated these CpG sites from further analysis. This left a total of 12 associated CpG sites in 9 loci, 56 unique metabolic measures, and 148 total CpG-metabolic measure pairs carried forward to all further follow-up analyses (Fig. 3).

### DNA methylation associated with serum metabolic measures is linked to transcriptional differences in blood, adipose tissue, and liver

We then explored associations between CpG sites showing replicated associations, and gene transcripts within 1 Mb in blood, adipose tissue, and liver (Table 2; Additional file 12: Table S8).

In the discovery cohort, we investigated a total of 480 CpG site-transcript pairs in whole blood. cg06500161 in *ABCG1* and cg16246545 in *PHGDH* were negatively associated (Bonferroni  $p < 0.05/480 = 1.0e-4$ ) with their corresponding gene transcript, results replicating those found in the BIOS QTL database [41] (Table 2; Additional file 12: Table S8). The association involving cg16246545 in *PHGDH* was slightly mitigated by adjustment for a *cis*-SNP, but remained strongly significant ( $p = 3.7e-14$ , results not shown). cg24503796 and cg03440556 in *SCD* were also negatively associated with *SCD* expression. Adjusting for SNPs in *cis* of the CpG sites had little or no effect on the results with regard to the expression results, except where noted.

We then performed an analysis using publicly available data for liver tissue (Karolinska Liver Bank cohort) for a total of 271 CpG site-expression probe pairs. In a limited sample of 92 individuals, no pairs were significant at a Bonferroni-corrected threshold ( $p < 0.05/271 = 1.8e-4$ ), but three sites had negative associations with transcripts at  $FDR < 0.05$ : cg17901584 in *DHCR24* with a transcript of *TTC4*; and cg24503796 and cg03440556 in *SCD* with transcripts annotated to *CHUK* and *COX15*, respectively (Additional file 12: Table S8).

Again using publicly available data, we performed a similar analysis for subcutaneous fat (TwinsUK study) for a total of 521 pairs. Only cg20544516 in *SREBF1* showed significant associations at a Bonferroni-corrected threshold ( $p < 0.05/521 = 9.6e-5$ ), with one transcript annotated to *PEMT*, one to *SHMT1*, and two annotated to *SREBF1*. The associations of cg20544516 with the two transcripts of *SREBF1* were replicated in the BIOS database. There were four additional significant associations at  $FDR < 0.05$ : cg24503796 in *SCD* with one transcript of *NDUFB8* and one of *PAX2*; cg20544516 in *SREBF1* with a transcript of *LOC201164*; and cg19693031 in *TXNIP* with a transcript of *DARS2* (Table 2; Additional file 12: Table S8).

### Limited to no evidence for sex specificity in DNA methylation-metabolic measure associations

In the discovery cohort, a CpG-by-sex interaction analysis of the replicated pairs revealed no Bonferroni-corrected significant ( $p < 0.05/148 = 3.4e-4$ ) differences between men and women for the associations (Additional file 13: Table S9). However, nominally significant

( $p < 0.05$ ) associations were found for 15 pairs, 14 of which involved small or very small VLDLs and CpG sites in either *SREBF1* or *CPT1A*.

### DNA methylation is linked to a variety of metabolic pathways

As a further approach to link enzymatic activity of selected metabolic pathways with CpG sites showing replicated associations, we examined 60 ratios of metabolic measures closely related to enzymatic substrates or products, or ratios linked to metabolic diseases in the discovery cohort (Additional file 4: Supplemental Results and Methods: Associations with metabolic ratios from additional pathways, Additional file 14: Table S10). A total of 189 associations were obtained for calculated ratios of metabolic measures (Bonferroni-corrected significance threshold of  $p < 0.05/(60 * 12) \approx 6.9e-5$ ). All 12 CpG sites were found to be associated with at least one ratio reflecting enzyme activity or linked to metabolic diseases, such as 25 associations relating to glucose metabolism and 38 involving branched-chain amino acids. However, most of the associations were obtained with lipid ratios and, among all assessed CpG sites, the lowest  $p$  values were largely for associations with ratios related to lipid metabolism. Therefore, we then tested associations of the 12 CpG sites with 30 transcripts of proteins directly involved in lipoprotein metabolism such as enzymes, transfer proteins, and lipid transporters, most of these located in *trans* to the CpG sites. Additional transcripts located within a  $\pm 500$  bp region were also included. We observed a total of four pairs showing associations (Bonferroni-corrected significance threshold of  $p < 0.05/(62 * 12) \approx 6.7e-5$ ) (Additional file 15: Table S11).

### DNA methylation associated with serum metabolic measures is linked to lipid-related clinical phenotypes

We further investigated the relevance of the replicated associations for lipid-related clinical phenotypes in the discovery cohort. Building on our previous publications [15, 25], we tested whether CpG sites associated with metabolic measures were related to type 2 diabetes (T2D), obesity, myocardial infarction (MI), or hypertension (Table 3; Additional file 16: Table S12). To explore the effects of lipid-lowering drugs on the relationships, we ran two models: one without this covariate (model 1) and one with (model 2). Statistical significance was based on a Bonferroni-corrected threshold of  $p < 0.05/(4 * 12) \approx 1.0e-3$ .

cg19693031 in *TXNIP* showed a strong association with T2D (*TXNIP*: odds ratio (OR) = 0.56 for a 1 standard deviation increase in methylation, 95% confidence interval (CI) = 0.47–0.69; model 1; Table 3), whereas seven CpG sites across six loci were significantly associated



**Table 3 Results of the investigation into associations between CpG sites associated with metabolic measures and clinical outcomes, and the effect on associations when adjusting for lipid-lowering drug intake**

CpG	Gene	Outcome	Model 1: Without adjustment for intake of lipid-lowering drugs				Model 2: Adjustment for intake of lipid-lowering drugs				Significant in
			Beta value M1	P_M1	Bonf P M1	OR (95% CI) M1	Beta value M2	P_M2	Bonf P M2	OR (95% CI) M2	
cg19693031	TXNIP	Diabetes	-0.56	4.33E-09	2.08E-07	0.57 (0.47-0.69)	1.32E-09	6.32E-08	0.55 (0.46-0.67)		M1 and M2
cg06500161 <sup>a</sup>	ABCG1	MI	0.70	2.87E-05	1.38E-03	2.02 (1.45-2.8)	2.56E-02	1	1.5 (1.05-2.15)		M1 only
cg17901584	DHCR24	MI	-0.65	4.37E-04	2.10E-02	0.52 (0.37-0.75)	4.04E-01	1	0.85 (0.57-1.25)		M1 only
cg06500161 <sup>b</sup>	ABCG1	obesity	0.35	3.94E-08	1.89E-06	1.42 (1.25-1.61)	2.54E-08	1.22E-06	1.43 (1.26-1.63)		M1 and M2
cg27243685 <sup>b</sup>	ABCG1	Obesity	0.29	5.56E-06	2.67E-04	1.34 (1.18-1.52)	3.96E-06	1.90E-04	1.35 (1.19-1.53)		M1 and M2
cg00574958 <sup>b</sup>	CPT1A	Obesity	-0.29	3.30E-06	1.58E-04	0.75 (0.66-0.85)	2.86E-06	1.37E-04	0.75 (0.66-0.84)		M1 and M2
cg07504977 <sup>b</sup>	LINC00263	Obesity	0.21	1.02E-03	4.87E-02	1.23 (1.09-1.39)	1.04E-03	4.98E-02	1.23 (1.09-1.39)		M1 and M2
cg16246545 <sup>b</sup>	PHGDH	Obesity	-0.29	5.88E-07	2.82E-05	0.74 (0.66-0.84)	1.27E-06	6.11E-05	0.75 (0.67-0.84)		M1 and M2
cg06690548 <sup>b</sup>	SLC7A11	Obesity	-0.31	2.75E-07	1.32E-05	0.73 (0.65-0.83)	3.76E-07	1.80E-05	0.74 (0.66-0.83)		M1 and M2
cg11024682 <sup>b</sup>	SREBF1	Obesity	0.34	1.26E-06	6.03E-05	1.4 (1.22-1.6)	1.34E-06	6.45E-05	1.4 (1.22-1.6)		M1 and M2

Associations of CpG site-clinical outcome in the KORA F4 cohort. Clinical outcomes: previous myocardial infarction, prevalent type 2 diabetes, obesity, and prevalent hypertension. Model 1: logistic regression model with clinical outcome as dependent variable and technically adjusted methylation value as independent variable, adjusted for the following covariates: age, sex, BMI (except for obesity model), C-reactive protein levels, hemoglobin A1c levels (except for diabetes model), history of myocardial infarction (except for the myocardial infarction model), smoking status, alcohol intake, current hypertension (except for the hypertension model), physical activity, white blood cell count, and estimated proportions of white blood cell types. Model 2: additionally adjusted for intake of lipid-lowering drugs (yes/no). Only statistically significant results are shown. P: P value; M1: model 1; M2: model 2; OR: odds ratio for a 1 standard deviation increase in methylation; CI: confidence interval

<sup>a</sup> Association using KORA F4 data in [15]

<sup>b</sup> Association using KORA F4 data in [25]

with obesity (smallest  $p$  value =  $2.8e-7$  for a site found in *SLC7A11*), among them one CpG site in *PHGDH*. The associations with both T2D and obesity were largely independent of lipid-lowering drug intake.

Associations with MI tended to be partially or completely mitigated when adjusting for the intake of lipid-lowering drugs. cg06500161 in *ABCG1* was positively associated with MI ( $p = 2.9e-5$ , OR = 2.02, CI = 1.45–2.80, model 1), but the association lost significance when adjusting for lipid-lowering drugs, with a large reduction in the odds ratio ( $p = 0.025$ , OR = 1.50; CI = 1.05–2.15). Similar effects were observed for cg17901584 in *DHCR24* ( $p = 4.4e-4$  for model 1, effect disappearing completely in model 2 with  $p = 0.4$ ).

No CpG sites were associated with hypertension. Adjusting for SNPs in *cis* of the CpG sites had little or no effect on the results with regard to clinical phenotypes.

We additionally generated receiver operating characteristic (ROC) curves for each of the CpG sites-clinical phenotype pairs significantly associated in model 1 or model 2 (Table 3; Additional file 17: Table S13; Additional file 18: Figure S4). For each pair and each model (M1 or M2), we plotted the ROC curve and calculated the area under the curve (AUC) for the model without the CpG site and for the model with the CpG site, to see whether the addition of the CpG site to the model significantly increases the model's ability to predict the outcome. After incorporating the associated CpG sites into the models, five CpG site-outcome pairs for M1 and four for M2 showed nominally significant ( $p < 0.05$ ) increases in AUC, e.g., cg19693031 in *TXNIP* showed an AUC = 0.813 without CpG site, and AUC = 0.834 with CpG site ( $p = 0.011$ ).

## Discussion

This EWAS of 226 mostly lipid-related serum metabolic measures is the largest to date incorporating the different lipid concentrations and lipid compositions of lipoprotein subclasses. Our EWAS revealed 161 replicated associations between 16 CpG sites in eleven loci and 57 unique metabolic measures. All of the eleven epigenetic loci have been previously found to be associated with metabolic traits and processes, primarily based on clinical and biochemical measurements of composite blood lipids [14–20, 22–25, 29, 37, 38, 42, 43]. Here, we uncover novel findings with regard to specific features of lipoprotein subclasses such as the lipid compositions and concentrations of each type of lipid in lipoproteins, giving deeper insights into the underlying biology of previous associations.

Results related to ApoB lipoproteins, particles involved in the endogenous lipoprotein pathway, suggest that DNA methylation is intertwined with the changes in lipid compositions and concentrations that all sizes of

VLDLs, IDLs, and LDLs undergo along the pathway. Methylation at CpG sites in *ABCG1*, *SREBF1*, *CPT1A*, and *TXNIP* had most of their associations with ApoB lipoproteins. Although methylation at CpG sites of those genes showed additional associations with HDLs, MUFAs, or isoleucine, their primary relationships seem to be with ApoB lipoproteins. *CPT1A* was the only gene whose CpG site showed associations with ApoB lipoproteins but not with HDLs. Methylation at cg00574958 in *CPT1A* was negatively associated with the concentration of almost all types of lipids and the TG composition in VLDLs, IDLs, and LDLs, adding evidence of a possible link between hypermethylation at this site and healthier metabolic phenotypes [44, 45]. Methylation at cg00574958 was also negatively associated with the TG composition of small LDLs. As smaller LDLs easily diffuse through the arterial wall, their low-TG load promotes increased cholesterol uptake and therefore impedes atherosclerosis development [46–48]. *CPT1A* is highly expressed in the liver, where it initiates mitochondrial oxidation of long-chain fatty acids and therefore contributes to lower serum TG levels, suggesting a relation between a higher *CPT1A* expression and lower serum TG levels [49]. Although no association between methylation at cg00574958 and expression of *CPT1A* was observed, this CpG site was negatively associated with serum TG and MUFA levels, the lower levels of which in turn limit TG acquisition by lipoproteins. Furthermore, albeit only nominally significant, sex-varying associations of methylation at cg00574958 with small and very small VLDLs could partially explain reported lower concentration and average size of circulating VLDLs in women compared to men [50]. In line with prior observations of sex-specific effects for *Cpt1a* in rodent lipid metabolism [51, 52], our results suggest a link between hypermethylation at cg00574958 and anti-atherogenic traits, possibly emphasized in females, further supporting that hypermethylation at this CpG site might be linked to healthier metabolic outcomes.

The reverse cholesterol transport (RCT) pathway delivers cellular cholesterol back to the liver in both a direct and indirect manner. The players of the direct RCT are HDLs, which are loaded with effluxed cholesterol from cells [9]. Apart from one association of larger HDL diameters with methylation at a CpG site in *DHCR24*, all other HDL associations found involved the TG concentrations and compositions in small and medium HDLs, and CpG sites in *ABCG1*, *SREBF1*, *LINC00263*, and *TXNIP*. HDLs that are TG-rich promote the clearance of circulating HDLs [46, 53, 54]. Therefore our findings might be related to the impairment of the RCT through larger HDLs and to associations of CpG sites in *ABCG1*, *SREBF1*, and *TXNIP* with adiposity-related traits [20,

24, 44, 45, 55]. *ABCG1*, *SREBF1*, and *TXNIP* were additionally associated with ApoB lipoprotein features and fatty acids. While these three genes have recently caught a lot of attention as they have been frequently found to be associated with metabolic traits [15–18, 20, 22–25, 29, 37, 44, 55], only sparse attention has been paid to a potential role of *LINC00263* in metabolic features [25, 32, 43]. In this study, *LINC00263* was the only lncRNA associated with metabolic measures. Notably, methylation at cg07504977 in *LINC00263* was solely associated with the TG concentration in small HDLs and no other metabolic measures, suggesting a specific role underlying previous general associations. First studies propose that *LINC00263* is a sex-specific oncogene [56]. Its involvement in metabolic disturbances is plausible as other genes involved in both cancer and metabolic disturbances have been reported [57]. No associations between methylation at cg07504977 and expression of *LINC00263* or other transcripts have been identified by us or by others, perhaps because of its absence on most commercial arrays. *LINC00263* interacts with at least 100 mRNAs, lncRNAs, miRNAs, and transcription factors [56], a common characteristic of lncRNAs [58]. The location of cg07504977 overlaps an active histone mark region within the promoter of *LINC00263* [43], suggesting a role of methylation at this CpG site in the lncRNA transcription. Thus, it seems feasible that methylation at cg07504977 could affect *LINC00263* expression, which in turn could directly adsorb or regulate the expression of miRNAs, and indirectly regulate the expression of miRNA target genes, e.g., miR-128, which directly inactivates *ABCG1* [59, 60]. Our results link methylation at cg07504977 to obesity and support this CpG as an emerging target for metabolic outcomes.

The indirect RCT pathway exchanges TG and cholesterol esters between VLDLs, LDLs, and HDLs, and is promoted by high levels of VLDLs with subsequent HDL clearance [6, 9]. Our results highlight methylation at cg06500161 in *ABCG1* as a biomarker entangled not only in the endogenous lipoprotein and the direct RCT pathways, but also in the indirect RCT pathway. cg06500161 showed the highest number of associations involving ApoB lipoproteins and HDLs and the lowest *p* values among all assessed CpG sites. In this study, cg06500161 was one of two CpG sites that exhibited associations with their respective gene transcript, metabolic measure ratios, and disease endpoints. Moreover, transcripts of *ABCG1* were additionally associated with CpG sites located in *trans* and annotated to *DHCR24* and *CPT1A*. Novel results for methylation at cg06500161 include associations with the lipid concentrations of all types of lipids and the TG compositions in ApoB lipoproteins, and associations with the TG composition and

TG concentration in smaller HDLs. Additionally, in line with the known function of *ABCG1* in controlling the bioavailability and activity of the TG-hydrolysis enzyme lipoprotein lipase [61], the analysis of metabolic measure ratios showed the lowest *p* values for ratios involving serum total TG levels. CpG methylation has shown to be driven by TG levels and not vice versa [24]. Although *ABCG1* promotes the net cholesterol efflux to larger HDLs in the direct RCT pathway [62, 63], and associations with *ABCG1* transcription were found, no associations of cg06500161 with the cholesterol concentration or composition in HDLs were observed. Therefore, methylation at cg06500161 might be linked to a lower activity of TG-hydrolysis enzymes (e.g., lipoprotein lipase), which in turn enhances the indirect RCT pathway, and impedes observation of associations with cholesterol features in HDLs [15, 25]. However, a precise role of cg06500161 in lipid metabolism remains to be elucidated, as does that of lipid-lowering drugs in the relationship. The previously hypothesized influence of statins on methylation at this site [64], and its mediation on the association between statins and type 2 diabetes [65], suggest that methylation at cg06500161 could lie on the causal path between the apparent mitigating effect of lipid-lowering drugs and its association with MI.

The chemical structure of fatty acids (FA) allows their categorization according to their saturation into MUFAs or PUFAs. In blood, FAs are transported by ApoB lipoproteins and HDLs. Although not fully understood, it has been proposed that MUFAs are mainly transported through the TG content of these lipoproteins, while PUFAs are mainly transported through the phospholipids or cholesterol ester content of lipoproteins [66]. We identified several associations for CpG methylation in *SCD*, *FADS1/2*, *SLC7A11*, *TXNIP*, and *PHGDH* with PUFAs. However, in line with previous studies, those in the *FADS* region appear to have a complex (epi)genetic architecture [37, 67–70]. The only CpG site showing associations with PUFAs, and additionally with the respective gene transcript, metabolic measure ratios, and disease endpoints was cg16246545 in *PHGDH*. Since we found no associations between methylation at cg16246545 and lipoproteins, but we did see that methylation at this site was associated with omega-6 PUFAs such as linoleic acid (LA), PUFAs and DNA methylation might have an interrelation that does not involve lipoproteins. Omega-6 PUFAs intake has been associated with changes to DNA methylation and metabolic alterations [71]. Additionally, the inhibition of *PHGDH* induces changes in DNA methylation and broad changes in metabolism such as alterations in nucleotide metabolism [72, 73]. Higher LA consumption might thus be related to methylation of

cg16246545. As we also found an association of this site with gene expression, this may be a pathway through which LA consumption leads to adverse metabolic outcomes such as obesity. Previously, we demonstrated an association between *PHGDH* transcription and a CpG site only 50 bp downstream from cg16246545 [25, 74]. We hypothesize that not only a single CpG site, but rather a bigger genetic region which is overlapping active histone marks, contributes to the functional relevance of cg16246545. Although we do not confirm negative associations of methylation at cg16246545 with serum total TG levels, we found novel associations with omega-6 FAs, fatty acids that are involved in regulation of TG levels [18].

The major strength of this work is the detailed information presented by the serum metabolic measures relating to the sizes, lipid compositions, and lipid concentrations of the lipoprotein subclasses. All CpG sites found to be associated with metabolic measures have additionally been associated with many more cardiometabolic conditions such as liver enzymes and hepatic steatosis [75] among others, thus extending the relevance of our findings. The generalizability of our results to a broad range of populations seems plausible, as our study uses population-based cohorts with different ethnic backgrounds [76]. We have confidence in our results owing to the overall large sample size of our study, the high percentage of replicated associations in the largest replication cohort (LOLIPOP), and the consistent directions of effect and Pearson correlations with the discovery cohort coefficients across all replication cohorts. However, NFBC1966 and YFS replication cohorts showed smaller effect sizes than those found in the discovery cohort, perhaps due to the younger age and healthier status of NFBC1966 and YFS participants. Additionally, cross-sectional studies do not readily provide information on causation in the context of DNA methylation, although recent studies imply an effect of lipid levels on DNA methylation rather than vice versa [24], and DNA methylation has often been considered a biomarker rather than a predictor [20, 24, 25, 29, 31, 32]. Mendelian randomization studies involving meta-analyses of studies of larger sample sizes than investigated here are needed to unravel the causal structures of the associations presented in this work [30]. Further in vitro and/or in vivo studies could also clarify causes and functional consequences of lipid-related DNA methylation alterations. Another limitation is the fact that DNA methylation was analyzed in DNA extracted from whole blood, a mixture of different cell types, while the investigated metabolic measures largely originate from metabolic processes in the liver, muscle, and adipose tissue. Nevertheless, blood represents an easy-to-obtain human tissue that can be used for predictive,

prognostic, and intervention biomarkers, and so its detailed investigation is certainly warranted.

Our findings could potentially be used as part of a multifaceted approach that incorporates genetic data, epigenetic data, and genetic–epigenetic interactions for complex disease prediction [77], hence offering future researchers a building block for developing biomarkers for dyslipidemia and other cardiometabolic diseases.

## Conclusion

In summary, serum metabolic measures were found to be associated with the methylation levels of interrelating genes involved in lipid metabolism and cardiometabolic disturbances. We observed that DNA methylation is linked to the sizes, lipid compositions, and lipid concentrations of apolipoprotein B-containing lipoprotein and HDL subclasses. No evidence of a link between DNA methylation and PUFAs involving lipoproteins was obtained. Our results provide in-depth insights into previous metabolic trait-DNA methylation associations based on total concentrations of serum lipids and indicate a complex regulation of the human metabolism possibly closely interrelated with epigenetic processes. We demonstrate the power of detailed metabolic measure profiling in large population-based cohorts to improve the molecular understanding of dyslipidemia and related disease mechanisms. Further studies are needed to clarify underlying functional mechanisms and identify pharmaceutical interventions for cardiometabolic disturbances.

## Methods

### Study design

The aim of this study was to identify the association of DNA methylation and a set of 226 mostly lipid-related NMR-measured serum metabolic measures. The design of the study comprised discovery and replication stages with subsequent follow-up analyses of the CpG sites associated with metabolic measures (Fig. 3). The discovery stage consisted of an EWAS of metabolic measures from the KORA cohort with subsequent validation of robust associations in the LOLIPOP, NFBC1966, and YFS replication cohorts. In the follow-up studies we assessed potential genetic confounding of the obtained associations, whether the associations varied between the sexes, and whether the CpG sites were associated with metabolic measure ratios, gene expression, and disease endpoints. Follow-up studies were performed in the KORA cohort using only those CpG sites that showed replicated associations with metabolic measures.

### Discovery cohort

The Cooperative Health Research in the Region of Augsburg (KORA) study is a series of independent



population-based epidemiological surveys and follow-up studies of participants living in the region of Augsburg, Southern Germany. The studies have been conducted according to the principles expressed in the Declaration of Helsinki. The KORA F4 study, a seven-year follow-up study of the KORA S4 survey (examined 1999–2001), was conducted between 2006 and 2008. The standardized examinations applied in the survey have been described in detail elsewhere [78]. A total of 3080 subjects with ages ranging from 32 to 81 years participated in the examination. Anthropometric and serum measures were measured concurrently. Aliquots of whole blood were stored at  $-80^{\circ}\text{C}$  for extraction of genomic DNA. In a random subgroup of 1802 KORA F4 subjects DNA methylation patterns were analyzed. Of the 1790 subjects who also had serum metabolic measures, 36 had detection rates of less than 95% over all measures and were eliminated from further analysis. Eight individuals were further eliminated due to non-fasting status at the time of blood sampling and 84 due to lack of valid methylation data, leaving a final sample size of 1662 subjects. Clinical phenotypes were defined as follows: type 2 diabetes (T2D), self-report, or intake of glucose-lowering medication, excluding metformin; hypertension,  $\geq 140/90$  mmHg, or intake of anti-hypertensive medication; obesity,  $\text{BMI} \geq 30$ ; previous myocardial infarction (MI), self-report.

#### Replication cohorts

The London Life Sciences Prospective Population Study (LOLIPOP) is a prospective cohort study of ~28000 Indian Asian and European men and women, recruited from the lists of 58 General Practitioners in West London, UK, between 2003 and 2008. In 4060 samples of Indian-Asian subjects, anthropometric and serum measures were measured concurrently, and aliquots of whole blood were stored at  $-80^{\circ}\text{C}$  for extraction of genomic DNA. DNA methylation was quantified in a subset of 2805 participants.

The Northern Finland Birth Cohort 1966 (NFBC1966) is a prospective population-based birth cohort, in the two northernmost provinces of Finland ( $N=12055$ ) with children whose expected date of birth was in the year 1966. In 1997–1998, a postal questionnaire on health, social status, and lifestyle was sent to the living cohort members, and those living in the original target area or in the capital area were invited for a clinical examination, including blood sample collection. Aliquots of whole blood were stored at  $-80^{\circ}\text{C}$  for later extraction of genomic DNA. DNA methylation patterns were analyzed for 807 subjects randomly selected.

The Cardiovascular Risk in Young Finns Study (YFS) is an ongoing multicentre Finnish longitudinal population study sample on the evolution of cardiovascular risk

factors from childhood to adulthood. The study began in 1980, when 3596 participants between the ages of 3 and 18 were randomly selected from the national population registers. Anthropometric and serum measures were measured concurrently, and aliquots of whole blood were stored at  $-80^{\circ}\text{C}$  for extraction of genomic DNA. In a subsample of 184 individuals randomly assigned from a follow-up in 2011, DNA methylation patterns were determined.

#### DNA methylation quantification

DNA methylation was quantified in bisulfite-converted genomic DNA from whole blood samples of all participants in both the discovery ( $N=1662$ ) and replication cohorts ( $N=3752$ ), using the Infinium HumanMethylation450 BeadChip (450 K BeadChip) (Illumina Inc, San Diego, CA, USA) in the discovery and replication cohorts. Further details on processing of the methylation data can be found in the supplemental methods (Additional file 4: Supplemental Results and Methods).

#### Serum metabolomics

Participants of all cohorts were in a state of fasting when blood samples were collected. Metabolite detection and quantification were performed on a high-throughput nuclear magnetic resonance (NMR) spectroscopy-based platform (Nightingale Ltd, Helsinki, Finland) [79, 80]. A total of 228 serum metabolic measures were assessed, and after data quality control 226 remained: 147 directly measured, mostly given in concentration units, and 79 derived ratios, mostly given in percentage, such as the ratios of specific types of lipids to total lipids in lipoprotein subclasses. The metabolic measures included six VLDL-, one IDL-, three LDL-, and four HDL-lipoprotein size-subclasses. Each lipoprotein size-subclass was measured for concentration and composition of phospholipids, total and free cholesterol, cholesterol esters, triglycerides, and total lipids. Additionally, two apolipoproteins, eight fatty acids, eight glycerides and phospholipids, nine cholesterols, nine amino acids, one inflammatory marker, and ten small molecules involved in glycolysis, citric acid cycle, or urea cycle were measured. Further details on sample preparation and the metabolic measure data can be found in the supplemental methods (Additional file 4: Supplemental Results and Methods).

#### Statistical analysis

##### **Epigenome-wide association studies: discovery and meta-analysis**

The discovery stage in the KORA F4 ( $N=1662$ ) cohort was made up of 226 epigenome-wide association studies, one per investigated metabolic measure passing quality

control (Additional file 1: Table S1). Specifically, for each metabolic measure, 468151 linear regression models were examined, one per CpG site. Each model used the natural logarithm of the metabolic measure as the dependent variable and technically adjusted beta values (i.e., proportion of methylation at the given CpG site) and covariates (detailed in Additional file 4: Supplemental Results and Methods) as explanatory variables. The covariates used in the linear models as potential confounders were age, sex, body mass index ( $\text{kg}/\text{m}^2$ ), c-reactive protein ( $\text{mg}/\text{l}$ ), HbA1c (%), smoking status (current smoker, ex-smoker, or never smoker), alcohol consumption ( $\text{g}/\text{day}$ ), lipid-lowering drug use (yes/no), presence of hypertension (yes/no), history of self-reported myocardial infarction (yes/no), level of physical activity (high/low), total white blood cell count (/nl), and proportions of white blood cell types as estimated using the Houseman method [81]. Statistical significance was determined using a Bonferroni-corrected threshold ( $p < 0.05 / (468151 \times 226) \approx 4.7 \times 10^{-10}$ ). Following the discovery, we ran sensitivity analyses to examine the model assumptions and the robustness of the results (Additional file 4: Supplemental Results and Methods). We then tested for replication of those significant CpG-metabolic measure pairs showing robustness using a meta-analysis of the results of the three participating replication studies (LOLIPOP,  $N = 2805$ ; NFBC1966,  $N = 771$ ; YFS,  $N = 176$ ; statistical significance  $p < 0.05 / 274 \approx 1.8 \times 10^{-4}$ ).

#### Genetic effects analysis

Multi-omics analyses were performed in the discovery cohort for associated CpG sites. Conditional analyses were performed to investigate whether genetic variation (single nucleotide polymorphisms, SNPs) within 1 Mb of the CpG sites could drive the relationships between the metabolic measures and methylation. For each CpG-metabolic measure pair, associated SNPs within 1 Mb of the CpG site were added singly to the models to determine the effect of the SNP on the association. Full details are given in the supplemental methods (Additional file 4: Supplemental Results and Methods).

#### Gene expression analysis

To study the interplay between the identified CpG sites and gene expression we examined associations with gene expression probes lying within 1 Mb of the significant CpG sites and extended these investigations to tissues beyond whole blood using data extracted from the ArrayExpress database [82] for both subcutaneous fat (Twin-sUK study, ArrayExpress references E-TABM-1140, and E-MTAB-1866 [83, 84]) and liver (Karolinska Liver Bank cohort ArrayExpress reference E-GEOD-61279 [85]). Statistical significance was determined in the discovery

cohort as  $p < 0.05 / 480 \approx 1.0 \times 10^{-4}$ ;  $p < 0.05 / 521 \approx 9.6 \times 10^{-5}$  for subcutaneous fat; and  $p < 0.05 / 271 \approx 1.8 \times 10^{-4}$  for liver; based on the total number of CpG-expression probe pairs examined per tissue. Results from the BIOS QTL browser (FDR  $< 0.05$ ), a database presenting whole blood expression-methylation associations [41, 86], were integrated into significant CpG-transcript associations. Additional details on the gene expression analysis can be found in the supplemental methods (Additional file 4: Supplemental Results and Methods).

#### Sex interaction analysis

Sex interaction analysis was performed in the discovery cohort for each replicated CpG site-metabolic measure association, excluding those involving CpG sites from the *FADS* region. The models were identical to the discovery models, but with a “sex  $\times$  methylation” interaction term (males as reference sex). Statistical significance for the interaction coefficient was judged at a Bonferroni-corrected threshold of  $p < 0.05 / 148 \approx 3.4 \times 10^{-4}$ .

#### Associations with serum metabolic measures ratios implicated in different pathways

In an attempt to identify specific steps of metabolic pathways that might be linked to DNA methylation in the discovery cohort KORA F4, we assessed association of CpG sites with additional ratios beyond those provided by the platform, using linear regression and adjusting for the same covariates used in the discovery EWAS. We calculated additional ratios related to the lipolysis, proteolysis, glycolysis, and ketogenesis pathways. Only associated metabolic measures and CpG sites from the replicated results were included. Statistical significance was determined as  $p < 0.05 / (60 \times 12) \approx 6.9 \times 10^{-5}$ . The tests for the associations of the 12 CpG sites with 30 transcripts of proteins directly involved in lipoprotein metabolism such as enzymes, transfer proteins, and lipid transporters was performed as for the gene expression analysis described for KORA F4 above and in the supplemental methods, but only looking additionally at probes within 500 bp, rather than 1 Mb. Statistical significance was determined as  $p < 0.05 / (62 \times 12) \approx 6.7 \times 10^{-5}$ . Further details on the selection criteria of ratios and transcripts can be found in the supplemental methods (Additional file 4: Supplemental Results and Methods).

#### Associations with clinical phenotypes

We next determined whether replicated CpG sites were associated with prevalent type 2 diabetes (T2D, self-report, or intake of glucose-lowering medication, excluding metformin;  $N = 148$  cases, 1516 controls), hypertension ( $\geq 140/90$  mmHg, or intake of anti-hypertensive medication;  $N = 757$  cases, 901 controls), obesity

(BMI  $\geq 30$ ;  $N=497$  cases, 1159 controls), or previous myocardial infarction (MI, self-report;  $N=59$  cases, 1602 controls) in the discovery cohort. For each CpG site and each outcome, we performed logistic regression adjusted for all covariates as in the discovery analysis (excluding HbA1c for diabetes and BMI for obesity) (model 1). To explore the effects of lipid-lowering drugs on the relationships, we ran two models: one without this covariate (model 1) and one with (model 2). Statistical significance was determined at a Bonferroni-corrected threshold of  $p < 0.05/(4 * 12) \approx 1.0e-3$ .

Additional details on the data preparation, statistical analysis of the discovery and meta-analysis, multi-omics analyses, and associations with ratios and clinical phenotypes can be found in the supplemental methods (Additional file 4: Supplemental Results and Methods). All statistical significance was determined using Bonferroni-corrected thresholds based on family-wise type I error rates of 0.05 and the number of relevant tests, except where noted.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13148-020-00957-8>.

**Additional file 1: Table S1.** List of the metabolic measures ( $N=228$ ) assessed in all cohorts.

**Additional file 2: Table S2.** Epigenome-wide association study (EWAS) of metabolic measures. Presented are CpG site-metabolic measurement pairs statistically significantly associated in the KORA F4 discovery study. Results for the discovery analysis in KORA ("KORA"), the sensitivity analysis in KORA ("\_KORA\_Sens"), LOLIPOP ("\_LOLIPOP"), NFBC1966 ("\_NFBC"), and Young Finns ("\_YFS") studies, as well as the results for the meta-analysis ("\_MA") of these three replication studies, are presented. The scaled coefficients of the sensitivity analysis in KORA ("Coef\_KORA\_scaled", i.e., both the log-transformed metabolite measure and methylation beta value were z-transformed prior to analysis) are also presented. The coefficients and  $p$  values for the discovery KORA analysis were calculated based on the results of 10 MICE imputed datasets and combined using the commands `pool.scalar` and `micombine.chisquare`, from the R packages `mice` and `micetools`, respectively. Discovery (KORA F4) significance based on  $p$  value  $< 4.73e-10$ ; meta-analysis significance based on  $p$  value  $< 1.80e-4$  (based on 274 pairs tested for replication). Gene, CHR, and Pos: gene, chromosome, and position annotation for the CpG site taken from the Illumina 450 K annotation file; Coef: coefficient of the CpG site from the regression analysis; SE: standard error of the coefficient; P:  $p$  value for the regression coefficient; N: number of observations; P\_Bonf: Bonferroni-corrected  $p$  value for the given analysis; Explained\_variance\_KORA: percentage of explained variance of the log-transformed metabolic measure by the CpG; Stat\_Sig\_MA: statistically significant in the meta-analysis of the three replication studies. Unless otherwise specified, all coefficients are change in natural log-transformed metabolite measurement unit (as given in Additional file 1: Table S1) per unit increase in methylation (beta value on 0–1 scale).

**Additional file 3: Table S3.** Summary statistics comparing the 274 KORA-significant associations across the replication cohorts. Results are given as total number (proportion) or as  $p$  values calculated using the binomial distribution ( $n$  = number of valid models in the study), where the probability for a single Bernoulli trial is given as  $*p=0.5$ ,  $**0.05$ ,  $***0.05/274$ . The number of valid models is the number of pairs for which results were available.

**Additional file 4: Supplemental Results and Methods.** In the Supplemental Results section a detailed comparison of the results obtained across cohorts can be found. In the Supplemental Methods section a detailed description of data processing for each of the population based cohorts can be found, as well as detailed specification on statistical and multi-omics analyses.

**Additional file 5: Table S4.** Replication tables. Replicated associations are displayed by CpG site, CpG gene location, metabolite (metabolic measure), and metabolite type (metabolic measure type). A total of 274 significant, robust associations in the discovery cohort (KORA F4) were found.

**Additional file 6: Figure S1.** Pearson correlations between metabolic measures associated with DNA methylation, KORA F4 data. Correlations among DNA methylation-associated metabolic measures are shown. All metabolic measures found associated with methylation in the discovery cohort are included.

**Additional file 7: Figure S2.** Explained variance of the first 30 principal components of the metabolite measure principal component analysis, performed in the discovery cohort KORA F4.

**Additional file 8: Figure S3.** Presented are the first 8 principal components (PCs) of the metabolite data in KORA F4 (after scaling of the individual metabolites) coloured according to sex, intake of lipid-lowering drugs, and smoking habits. "Expl. Var" is the variance explained by the given PC. Some clustering is observed for sex within the first 2 PCs, but no other obvious clusters emerge for any of the other phenotypes or PCs.

**Additional file 9: Table S5.** Results of the EWAS of metabolite measure principal components. Presented are the statistically significant (Bonferroni-adjusted  $P < 1.34e-8$ ) results of the EWAS of the first 8 principal components of the metabolite measures data in the discovery cohort KORA F4. CHR: annotated chromosome of the CpG site; Pos: annotated chromosomal position of the CpG site; Coef: coefficient of the CpG site in the regression model; SE: standard error of the coefficient; P:  $p$  value; PC\_explained\_var: Explained variance of the metabolite principal component.

**Additional file 10: Table S6.** Results of look-ups in two epigenome-wide association study (EWAS) catalogues, the EWAS Atlas [39] (Source: "EWAS\_Atlas") and the EWAS Catalog [40] (Source: "EWAS\_Cat"). Shown are the catalogue results (significantly associated CpGs and traits) for all CpGs from our replicated metabolic measure-CpG associations (16 CpG sites total). Chromosome (CHR) and chromosomal position (Pos) taken from the Illumina HumanMethylation450 v1.2 Manifest File, available from the Illumina website.

**Additional file 11: Table S7.** Genetically influenced CpG site-metabolic measure associations. Results for the *cis*-SNP analysis are presented. "Coef\_discovery" and "P\_discovery": coefficient and  $p$  value for the discovery analysis for the given CpG site-trait pair. "Count\_conf\_SNPs": number of SNPs which, when added singly to the metabolic measure-CpG site regression model, cause the pair to lose its statistical significance, as defined by the discovery threshold ( $p=4.73e-10$ ). "Coef\_adj" and "p\_adj": coefficient and  $p$  value of the CpG in the model with the addition of the SNP causing the greatest effect (largest change in  $p$  value). "top\_SNP": name of the SNP causing the largest effect. "Loses\_significance" indicates whether the addition of any SNP to the CpG-metabolic measure regression model causes the association to lose significance (i.e., "Count\_conf\_SNPs"  $> 0$ ). "Probable\_SNP\_confounding" indicates the addition of at least one single SNP to the model renders the association insignificant and drastically alters the results, indicated likely SNP confounding by 1 or more SNPs. NAs within the table indicate the absence of a SNP fulfilling the requirements of our conditional analyses for the CpG-metabolic measure pair (i.e., the absence of a SNP being associated with both the CpG site and the metabolic measure). CHR: chromosome, pos: position, UCSC\_RefGene\_Name\_CpG: annotated gene name of the CpG site, all as given by the Illumina manifest file. N: number of observations in the model incorporating the SNP. All coefficients are change in natural-log transformed metabolite measurement unit (as given in Additional file 1: Table S1) per unit increase in methylation (beta value on 0–1 scale). \*Drastic increase (factor  $> 100$ ) of  $p$  value after addition of the SNP to the regression model.

**Additional file 12: Table S8.** Results of gene expression analysis of CpG sites associated with metabolic measures. Displayed are the FDR (Benjamini-Hochberg,  $< 0.05$ ) statistically significant associations between metabolic measure-associated CpG sites and expression transcripts in *cis* in whole blood (KORA F4, 480 CpG-transcript pairs examined), subcutaneous fat (TwinsUK study, 521 pairs examined) and liver (Karolinska Liver Bank cohort and the Dutch tissue cohort MORE/BBMRI obesity cohort, 271 pairs examined). FDR-adjusted  $p$  values were calculated for each tissue separately. "Probe ID" and "Transcript\_annotated\_gene": transcript ID and annotated gene from the Illumina annotation files. "Distance": distance between the CpG and the transcript according to the annotation files. CHR: chromosome; Distance: distance between the CpG and the transcript based on positions given in the annotation files; Coef: beta-coefficient; P:  $p$  value; FDR: FDR-corrected  $p$  value; Bonf\_P: Bonferroni-corrected  $p$  value; N: number of observations in the model; "replicated BIOS": whether the association (where the CpG is matched directly, but the transcript is matched by annotated gene only), with consistent direction of effect, is found in the FDR  $< 0.05$  significant BIOS QTL database. An "-" indicates the CpG-gene expression pair is not found in the FDR  $< 0.05$  significant results in BIOS. All coefficients are change in log<sub>2</sub>-transformed expression intensity per unit increase in methylation (beta value on 0–1 scale), except for subcutaneous fat, which is correlation assessed using the R-package rmcrr.

**Additional file 13: Table S9.** Results of the sex interaction analysis. Results for all 148 replication CpG site-metabolic measure associations are shown for models identical to the discovery, but with additional "sex  $\times$  methylation" interaction term. Presented are the obtained coefficient (interaction\_coef) for the term of the interaction between methylation and sex (0: male, 1: female), standard error of the coefficient (interaction\_SE),  $p$  value of the coefficient (interaction\_P), the Bonferroni-corrected  $p$  value of the coefficient (interaction\_bonf\_p), the methylation-metabolite measure coefficient for males (coef\_CpG\_Met\_male), with the calculated females coefficient (coef\_CpG\_Met\_female, i.e., coef\_CpG\_Met\_male + interaction\_coef), and number of observations for the model (N). No interaction coefficient results pass a Bonferroni-corrected  $p$  value threshold of  $0.05/148 \sim 3.4e-4$  for statistical significance. All coefficients are change in natural-log transformed metabolite measurement unit (as given in Additional file 1: Table S1) per unit increase in methylation (beta value on 0–1 scale).

**Additional file 14: Table S10.** Results of the associations with additional metabolic measure ratios as proxies of enzymatic activity or linked to metabolic disease. Results for all 12 replicated metabolic measure-associated CpG sites associations are shown. Presented are the coefficient (Coef) of the CpG site from the regression analysis, standard error of the coefficient (SE),  $p$  value of the coefficient (P), Bonferroni-corrected  $p$  value (Bonf\_P) and number of observations for the model (N). A Bonferroni-corrected  $p$  value threshold of  $p < 0.05/(60 * 12) \approx 6.9e-5$  is used for statistical significance. All coefficients are change in difference of natural-log transformed metabolite measurement units (as given in Additional file 1: Table S1, i.e.,  $\log(\text{metabolite 1}) - \log(\text{metabolite 2})$ ) per unit increase in methylation (beta value on 0–1 scale).

**Additional file 15: Table S11.** Results of the associations between CpG sites and expression of transcripts of genes coding for enzymes or proteins directly involved with lipoprotein metabolism. Results are shown for those pairs passing a false discovery rate (Benjamini-Hochberg) threshold of FDR  $< 0.05$ . Presented are the CpG sites and expression probe IDs; the annotated genes of the CpG site and expression probe; the chromosomes of the CpG site and expression probe; the chromosomal location of the CpG site; the distance between the CpG site and expression probe (where Inf indicates they are on different chromosomes); the coefficient (Coef) of the CpG site from the regression analysis; standard error of the coefficient (SE);  $p$  value of the coefficient (P); Bonferroni-corrected  $p$  value (Bonf\_P); FDR-corrected  $p$  value (FDR); number of observations for the model (N); and whether the association is significant at Bonferroni-corrected threshold (based on 12 CpGs  $\times$  62 expression probes:  $p < 0.05/(12 * 62) = 6.7e-5$ ). All coefficients are change in log<sub>2</sub>-transformed expression intensity per unit increase in methylation (beta value on 0–1 scale).

**Additional file 16: Table S12.** Results of the investigation into associations between metabolic measures-related CpG sites and clinical outcomes. Listed are the CpG site-clinical outcome (previous myocardial infarction, prevalent type 2 diabetes, obesity and prevalent hypertension) results for both model 1: logistic regression model with clinical outcome as dependent variable and technically adjusted methylation value as independent variable, adjusted for the following covariates: age, sex, BMI, C-reactive protein levels, hemoglobin A1c levels (except for diabetes model), history of myocardial infarction (except for the myocardial infarction model), smoking status, current hypertension (except for the hypertension model), physical activity, white blood cell count and estimated proportions of white blood cell type; or model 2, which is additionally adjusted for intake of lipid-lowering drugs. P:  $p$  value of the association; OR: odds ratio for a 1 standard deviation increase in methylation; CI: confidence interval; FDR: false discovery rate  $p$  value based on the Benjamini-Hochberg method applied to each outcome and model separately; M1: model 1; M2: model 2. \*Association investigated using the KORA F4 data in [15]; \*\*Association investigated using the KORA F4 data in [25].

**Additional file 17: Table S13.** ROC curve analysis for significant CpG-outcome associations: Presented are the areas under the curve (AUC) for the receiver operating characteristic (ROC) curve analysis for each outcome-CpG pair for which there exists a statistically significant association for either M1 or M2 (Table 3). Presented are the results for the intercept (mean) model and the model with the CpG sites and no other covariates; for M1 without and with the CpG site; and for M2 without and with the CpG site. Presented are the AUCs for the respective ROCs, and a  $p$  value for the null hypothesis that the addition of the CpG site to the model has no effect on the predictive performance of the model. The  $p$  value was determined using the R package pROC [87], command roc.test, method "bootstrap". The analysis was run in the KORA F4 dataset, and, to ensure comparability of the results, the ROC curves were generated using individuals with no missing values in any of the outcomes or covariates, and the methylation data were mean imputed.

**Additional file 18: Figure S4.** ROC curves for significant CpG-outcome associations: Presented are the receiver operating characteristic (ROC) curves for each outcome-CpG pair for which there exists a statistically significant association for either M1 or M2 (Table 3). The red line is the ROC curve for the model without the CpG, and the green line is the model with the CpG. Presented are also the areas under the curve (AUC) for the respective ROCs, and a  $p$  value for the null hypothesis that the addition of the CpG to the model has no effect on the predictive performance of the model. The  $p$  value was determined using the R package pROC [87], command roc.test, method "bootstrap". The analysis was run in the KORA F4 dataset, and, to ensure comparability of the results, the ROC curves were generated using individuals with no missing values in any of the outcomes or covariates, and the methylation data were mean imputed.

**Additional file 19: Table S14.** Inflation factors for each epigenome-wide association analysis. Presented are the genomic inflation factors (lambda) for all 226 EWAS run in the discovery analysis in KORA F4. All inflation factors were calculated using complete case analysis.

## Abbreviations

DHA: Docosahexaenoic acid; DNA: Deoxyribonucleic acid; EWAS: Epigenome-wide association study; FA: Fatty acids; HbA1c: Hemoglobin A1c; HDL: High-density lipoproteins; IDL: Intermediate-density lipoproteins; KORA: Cooperative Health Research in the Region of Augsburg; LA: Linoleic acid; LDL: Low-density lipoproteins; lncRNA: Long non-coding RNA; LOLIPOP: London Life Sciences Prospective Population Study; Mb: Mega-base pairs; MI: Myocardial infarction; miRNA: Micro-ribonucleic acid; mRNA: Messenger ribonucleic acid; MUFA: Monounsaturated fatty acids; NFBC1966: Northern Finland Birth Cohort 1966; NMR: Nuclear magnetic resonance; PUFA: Polyunsaturated fatty acids; RCT: Reverse cholesterol transport; SNP: Single nucleotide polymorphisms; T2D: Type 2 diabetes; TG: Triglycerides; VLDL: Very low-density lipoproteins; YFS: Cardiovascular Risk in Young Finns.



### Acknowledgements

We thank the participants and research staff from all cohorts and studies who made the study possible. NFBC1966 would like to thank the late professor Paula Rantakallio (launch of NFBC1966).

### Authors' contributions

LP, SW, MW, RW designed the study. RW, SH, VK analyzed the data. MCGA, AK, RW, LP, KM, SW, PRMG, LRSG, MAK, CG, MW interpreted the data. MCGA, AK, RW, MW wrote the manuscript. RW, AK, MCGA, LRSG prepared figures and tables. TL, MH, MMN, CH, WR, MR, MAK, IS, WZ, HP, TM, SS, OR, ML, MRJ, JSK, AP, JC, CG, JK contributed to population-based cohorts. All authors reviewed and approved the manuscript prior to submission. All authors read and approved the final manuscript.

### Funding

The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research has been supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. This work was supported by a Grant (WA 4081/1-1) from the German Research Foundation and by the German Federal Ministry of Education and Research (BMBF) within the framework of the EU Joint Programming Initiative “A Healthy Diet for a Healthy Life” (01EA1902A: Dimension). The German Diabetes Center (DDZ) is funded by the German Federal Ministry of Health and the Ministry of Science and Culture of the State North Rhine-Westphalia. This study was supported in part by a Grant from the German Federal Ministry of Education and Research to the German Center for Diabetes Research (DZD). The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z, 090532 & 098381), the NIHR (RP-PG-0407-10371), the NIHR Official Development Assistance (ODA, award 16/136/68), the European Union FP7 (EpiMigrant, 279143), and H2020 programs (iHealth-T2D, 643774). The Young Finns Study has been financially supported by the Academy of Finland: Grants 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere, and Turku University Hospitals (Grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (Grant 755320 for TAXINOMISIS; Grant 848146 for To\_Aition); European Research Council (Grant 742927 for MULTIEPIGEN project); and Tampere University Hospital Supporting Foundation. The NFBC1966 team acknowledges funding from the following: European Union's Horizon 2020 research and innovation programme [DYNAHEALTH 633595, LIFECYCLE 733206, EUCANCONNECT 824989, LongITools 874739, EarlyCause 848458], Academy of Finland [EGEA 285547], and the JPI-HDHL program [PRECISE – MRC-UK P75416]. MCGA holds a scholarship from the Consejo Nacional de Ciencia y Tecnología (CONACYT)-México. MAK was supported by a research Grant from the Sigrid Juselius Foundation, Finland. JC is supported by the Singapore Ministry of Health's National Medical Research Council under its Singapore Translational Research Investigator (STaR) Award (NMRC/STaR/0028/2017). The funders had no role in study design or data collection, analysis, and interpretation. The funding sources had no influence in the study design, collection, analysis, interpretation of data, writing of the report, and in the decision to submit the article.

### Availability of data and materials

The publicly available datasets analyzed during the current study included subcutaneous fat data (TwinsUK study) and liver data (Karolinska Liver Bank cohort) which are available in the following repositories: ArrayExpress reference E-TABM-1140 <https://www.ebi.ac.uk/arrayexpress/experiments/E-TABM-1140/>. ArrayExpress reference E-MTAB-1866 <https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1866/>. ArrayExpress reference E-GEOD-61279 <https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-61279/>.

### Ethics approval and consent to participate

The KORA cohort is approved by the local ethics committee Bayerische Landesärztekammer in Bayern, Germany. The LOLIPOP study is approved by the National Research Ethics Service (07/H0712/150). The NFBC1966 cohort is approved by the Ethics Committee of the Northern Ostrobothnia Hospital District. The YFS study is approved by the Ethics Committee of the Hospital District of Southwest Finland. All participants provided written informed consent to participate in the respective study.

### Consent for publication

Not applicable.

### Competing interests

CH received honoraria from Lilly and Sanofi and a research Grant from Sanofi outside the submitted work.

### Author details

<sup>1</sup> Research Unit of Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München German Research Center for Environmental Health, Ingolstaedter Landstraße 1, 85764 Neuherberg, Germany. <sup>2</sup> Institute of Epidemiology, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg, Germany. <sup>3</sup> Department of Epidemiology and Biostatistics, Imperial College London, London, UK. <sup>4</sup> Department of Clinical Chemistry, Pirkanmaa Hospital District, Fimlab Laboratories, and Finnish Cardiovascular Research Center, Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. <sup>5</sup> Department of Cardiology, Ealing Hospital, London North West University Healthcare NHS Trust, London, Middlesex, UK. <sup>6</sup> Institute of Human Genetics, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg, Germany. <sup>7</sup> Institute of Human Genetics, School of Medicine, Technical University Munich, Munich, Germany. <sup>8</sup> German Center for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany. <sup>9</sup> Chair of Microbiology, Technical University of Munich, Freising, Germany. <sup>10</sup> Center for Life Course Health Research, University of Oulu, Oulu University Hospital, Oulu, Finland. <sup>11</sup> Centre for Population Health Research, University of Turku, Turku University Hospital, Turku, Finland. <sup>12</sup> Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. <sup>13</sup> Department of Clinical Physiology and Nuclear Medicine, University of Turku, Turku University Hospital, Turku, Finland. <sup>14</sup> Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore. <sup>15</sup> Institute for Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>16</sup> German Center for Diabetes Research (DZD), Munich-Neuherberg, Germany. <sup>17</sup> Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany. <sup>18</sup> Institute of Genetic Epidemiology, Helmholtz Zentrum München German Research Center for Environmental Health, Neuherberg, Germany. <sup>19</sup> Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>20</sup> Division of Endocrinology and Diabetology, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>21</sup> Department of Microbiology and Immunology, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. <sup>22</sup> UKMRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. <sup>23</sup> Department of Life Sciences, College of Health and Life Sciences, Brunel University London, London, UK. <sup>24</sup> NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland. <sup>25</sup> Computational Medicine, Faculty of Medicine, University of Oulu, Oulu, Finland. <sup>26</sup> Biocenter Oulu, University of Oulu, Oulu, Finland. <sup>27</sup> National Heart and Lung Institute, Imperial College London, London, UK. <sup>28</sup> Imperial College Healthcare NHS Trust, London, UK. <sup>29</sup> MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. <sup>30</sup> Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, 55101 Mainz, Germany.

Received: 25 August 2020 Accepted: 22 October 2020

Published online: 07 January 2021

## References

- Petitti DB, Imperatore G, Palla SL, Daniels SR, Dolan LM, Kershner AK, et al. Serum lipids and glucose control: the SEARCH for Diabetes in Youth study. *Arch Pediatr Adolesc Med*. 2007;161(2):159–65.
- Castelli WP. Cholesterol and lipids in the risk of coronary artery disease—the Framingham Heart Study. *Can J Cardiol*. 1988;4 Suppl A:5A-10A.
- Wang J, Stancakova A, Soininen P, Kangas AJ, Paananen J, Kuusisto J, et al. Lipoprotein subclass profiles in individuals with varying degrees of glucose tolerance: a population-based study of 9399 Finnish men. *J Intern Med*. 2012;272(6):562–72.
- Flora GD, Nayak MK. A brief review of cardiovascular diseases, associated risk factors and current treatment regimes. *Curr Pharm Des*. 2019;25(38):4063–84.
- Klop B, Elte JW, Cabezas MC. Dyslipidemia in obesity: mechanisms and potential targets. *Nutrients*. 2013;5(4):1218–40.
- Feingold KR, Grunfeld C. Introduction to lipids and lipoproteins. In: Feingold KR, Anawalt B, Boyce A, Chrousos G, Dungan K, Grossman A, et al., editors. *South Dartmouth: Endotext*; 2000.
- Kumpula LS, Kumpula JM, Taskinen MR, Jauhiainen M, Kaski K, Ala-Korpela M. Reconsideration of hydrophobic lipid distributions in lipoprotein particles. *Chem Phys Lipids*. 2008;155(1):57–62.
- Ginsberg HN. Lipoprotein physiology. *Endocrinol Metab Clin North Am*. 1998;27(3):503–19.
- Ouimet M, Barrett TJ, Fisher EA. HDL and reverse cholesterol transport. *Circ Res*. 2019;124(10):1505–18.
- Jang H, Park K. Omega-3 and omega-6 polyunsaturated fatty acids and metabolic syndrome: a systematic review and meta-analysis. *Clin Nutr*. 2020;39(3):765–73.
- Deelen J, Kettunen J, Fischer K, van der Spek A, Trompet S, Kastenmuller G, et al. A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat Commun*. 2019;10(1):3346.
- Muzsik A, Jelen HH, Chmurzynska A. Metabolic syndrome in postmenopausal women is associated with lower erythrocyte PUFA/MUFA and n-3/n-6 ratio: a case-control study. *Prostaglandins Leukot Essent Fatty Acids*. 2020;159:102155.
- Pietiläinen KH, Naukkarinen J, Rissanen A, Saharinen J, Ellonen P, Keranen H, et al. Global transcript profiles of fat in monozygotic twins discordant for BMI: pathways behind acquired obesity. *PLoS Med*. 2008;5(3):e151.
- Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, et al. Epigenome-wide association study of fasting blood lipids in the genetics of lipid-lowering drugs and diet network study. *Circulation*. 2014;130(7):565–72.
- Pfeiffer L, Wahl S, Pilling LC, Reischl E, Sandling JK, Kunze S, et al. DNA methylation of lipid-related genes affects blood lipid levels. *Circ Cardiovasc Genet*. 2015;8(2):334–42.
- Sayols-Baixeras S, Subirana I, Lluís-Ganella C, Civeira F, Roquer J, Do AN, 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*. 2016;25(20):4556–65.
- Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, Uitterlinden AG, et al. Epigenome-wide association study (EWAS) on lipids: the Rotterdam study. *Clin Epigenet*. 2017;9:15.
- Hedman AK, Mendelson MM, Marioni RE, Gustafsson S, Joehanes R, Irvin MR, et al. Epigenetic patterns in blood associated with lipid traits predict incident coronary heart disease events and are enriched for results from genome-wide association studies. *Circ Cardiovasc Genet*. 2017;10(1):e001487.
- Mittelstrass K, Waldenberger M. DNA methylation in human lipid metabolism and related diseases. *Curr Opin Lipidol*. 2018;29(2):116–24.
- Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman AK, Aslibekyan S, 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(1):e1002215.
- Guay SP, Voisin G, Brisson D, Munger J, Lamarche B, Gaudet D, et al. Epigenome-wide analysis in familial hypercholesterolemia identified new loci associated with high-density lipoprotein cholesterol concentration. *Epigenomics*. 2012;4(6):623–39.
- Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, 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*. 2015;3(7):526–34.
- Kriebel J, Herder C, Rathmann W, Wahl S, Kunze S, Molnos S, et al. Association between DNA methylation in whole blood and measures of glucose metabolism: KORA F4 study. *PLoS ONE*. 2016;11(3):e0152314.
- Dekkers KF, van Iterson M, Sliker RC, Moed MH, Bonder MJ, van Galen M, et al. Blood lipids influence DNA methylation in circulating cells. *Genome Biol*. 2016;17(1):138.
- Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*. 2017;541(7635):81–6.
- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. *PLoS ONE*. 2012;7(7):e41361.
- Tammen SA, Friso S, Choi SW. Epigenetics: the link between nature and nurture. *Mol Aspects Med*. 2013;34(4):753–64.
- van der Harst P, de Windt LJ, Chambers JC. Translational perspective on epigenetics in cardiovascular disease. *J Am Coll Cardiol*. 2017;70(5):590–606.
- Zaghloul SB, Mook-Kanamori DO, Kader S, Stephan N, Halama A, Engelke R, et al. Deep molecular phenotypes link complex disorders and physiological insult to CpG methylation. *Hum Mol Genet*. 2018;27(6):1106–21.
- Dekkers KF, Slagboom PE, Jukema JW, Heijmans BT. The multifaceted interplay between lipids and epigenetics. *Curr Opin Lipidol*. 2016;27(3):288–94.
- Reed ZE, Suderman MJ, Relton CL, Davis OSP, Hemani G. The association of DNA methylation with body mass index: distinguishing between predictors and biomarkers. *Clin Epigenet*. 2020;12(1):50.
- Sayols-Baixeras S, Tiwari HK, Aslibekyan SW. Disentangling associations between DNA methylation and blood lipids: a Mendelian randomization approach. *BMC Proc*. 2018;12(Suppl 9):23.
- Kettunen J, Holmes MV, Allara E, Anufrieva O, Ohukainen P, Oliver-Williams C, et al. Lipoprotein signatures of cholesteryl ester transfer protein and HMG-CoA reductase inhibition. *PLoS Biol*. 2019;17(12):e3000572.
- Rankin NJ, Preiss D, Welsh P, Burgess KE, Nelson SM, Lawlor DA, et al. The emergence of proton nuclear magnetic resonance metabolomics in the cardiovascular arena as viewed from a clinical perspective. *Atherosclerosis*. 2014;237(1):287–300.
- Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet*. 2015;8(1):192–206.
- Wurtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -omic technologies. *Am J Epidemiol*. 2017;186(9):1084–96.
- Petersen AK, Zeilinger S, Kastenmuller G, Romisch-Margl W, Brügger M, Peters A, et al. Epigenetics meets metabolomics: an epigenome-wide association study with blood serum metabolic traits. *Hum Mol Genet*. 2014;23(2):534–45.
- Frazier-Wood AC, Aslibekyan S, Absher DM, Hopkins PN, Sha J, Tsai MY, et al. Methylation at CPT1A locus is associated with lipoprotein subfraction profiles. *J Lipid Res*. 2014;55(7):1324–30.
- Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res*. 2019;47(D1):D983–8.
- Bristol UO. The MRC-IEU catalog of epigenome-wide association studies 2018. <https://www.ewascatalog.org/>. Accessed March 2019.
- Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet*. 2017;49(1):131–8.
- Chitrana KN, Hernandez DG, Nalls MA, Mode NA, Zonderman AB, Ezike N, et al. Race-specific alterations in DNA methylation among middle-aged African Americans and Whites with metabolic syndrome. *Epigenetics*. 2020;15(5):462–82.
- Aslibekyan S, Demerath EW, Mendelson M, Zhi D, Guan W, Liang L, et al. Epigenome-wide study identifies novel methylation loci associated with body mass index and waist circumference. *Obesity (Silver Spring)*. 2015;23(7):1493–501.
- Cardona A, Day FR, Perry JRB, Loh M, Chu AY, Lehne B, et al. Epigenome-wide association study of incident type 2 diabetes in a British population: EPIC-norfolk study. *Diabetes*. 2019;68(12):2315–26.

45. Karlsson IK, Ericsson M, Wang Y, Jylhava J, Hagg S, Pedersen NL, et al. Replicating associations between DNA methylation and body mass index in a longitudinal sample of older twins. *Int J Obes (Lond)*. 2020;44(6):1397–405.
46. Nordestgaard BG. Triglyceride-rich lipoproteins and atherosclerotic cardiovascular disease: new insights from epidemiology, genetics, and biology. *Circ Res*. 2016;118(4):547–63.
47. Davidson MH. Triglyceride-rich lipoprotein cholesterol (TRL-C): the ugly stepsister of LDL-C. *Eur Heart J*. 2018;39(7):620–2.
48. Srisawasdi P, Vanavanan S, Rochanawutanon M, Kruthkul K, Kotani K, Kroll MH. Small-dense LDL/large-buoyant LDL ratio associates with the metabolic syndrome. *Clin Biochem*. 2015;48(7–8):495–502.
49. Schlaepfer IR, Joshi M. CPT1A-mediated fat oxidation, mechanisms, and therapeutic potential. *Endocrinology*. 2020;161(2):bqz046.
50. Wang X, Magkos F, Mittendorfer B. Sex differences in lipid and lipoprotein metabolism: it's not just about sex hormones. *J Clin Endocrinol Metab*. 2011;96(4):885–93.
51. Bazhan N, Jakovleva T, Feofanova N, Denisova E, Dubinina A, Sitnikova N, et al. Sex differences in liver, adipose tissue, and muscle transcriptional response to fasting and refeeding in mice. *Cells*. 2019;8(12):1529.
52. Jiang Z, Huang X, Huang S, Guo H, Wang L, Li X, et al. Sex-related differences of lipid metabolism induced by triptolide: the possible role of the LXRalpha/SREBP-1 signaling pathway. *Front Pharmacol*. 2016;7:87.
53. Voloshyna I, Reiss AB. The ABC transporters in lipid flux and atherosclerosis. *Prog Lipid Res*. 2011;50(3):213–24.
54. Zhou L, Li C, Gao L, Wang A. High-density lipoprotein synthesis and metabolism (review). *Mol Med Rep*. 2015;12(3):4015–21.
55. Krause C, Sievert H, Geissler C, Grohs M, El Gammal AT, Wolter S, et al. Critical evaluation of the DNA-methylation markers ABCG1 and SREBF1 for Type 2 diabetes stratification. *Epigenomics*. 2019;11(8):885–97.
56. Liu S, Lai W, Shi Y, Liu N, Ouyang L, Zhang Z, et al. Annotation and cluster analysis of long noncoding RNA linked to male sex and estrogen in cancers. *NPJ Precis Oncol*. 2020;4:5.
57. Dragic D, Ennour-Idrissi K, Michaud A, Chang SL, Durocher F, Diorio C. Association between BMI and DNA methylation in blood or normal adult breast tissue: a systematic review. *Anticancer Res*. 2020;40(4):1797–808.
58. Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell*. 2013;152(6):1298–307.
59. Monchusi B, Kaur M. microRNAs targeting cellular cholesterol: implications for combating anticancer drug resistance. *Genes Cancer*. 2020;11(1–2):20–42.
60. Zhang Q, Ma XF, Dong MZ, Tan J, Zhang J, Zhuang LK, et al. MiR-30b-5p regulates the lipid metabolism by targeting PPARC1A in Huh-7 cell line. *Lipids Health Dis*. 2020;19(1):76.
61. Olivier M, Tanck MW, Out R, Villard EF, Lammers B, Bouchareychas L, et al. Human ATP-binding cassette G1 controls macrophage lipoprotein lipase bioavailability and promotes foam cell formation. *Arterioscler Thromb Vasc Biol*. 2012;32(9):2223–31.
62. Wang N, Lan D, Chen W, Matsuura F, Tall AR. ATP-binding cassette transporters G1 and G4 mediate cellular cholesterol efflux to high-density lipoproteins. *Proc Natl Acad Sci USA*. 2004;101(26):9774–9.
63. Hardy LM, Frisdal E, Le Goff W. Critical role of the human ATP-binding cassette G1 transporter in cardiometabolic diseases. *Int J Mol Sci*. 2017;18(9):1892.
64. Ochoa-Rosales C, Portilla-Fernandez E, Nano J, Wilson R, Lehne B, Mishra PP, et al. Epigenetic link between statin therapy and type 2 diabetes. *Diabetes Care*. 2020;43(4):875–84.
65. Liu Y, Shen Y, Guo T, Parnell LD, Westerman KE, Smith CE, et al. Statin use associates with risk of type 2 diabetes via epigenetic patterns at ABCG1. *Front Genet*. 2020;11:622.
66. Arterburn LM, Hall EB, Oken H. Distribution, interconversion, and dose response of n-3 fatty acids in humans. *Am J Clin Nutr*. 2006;83(6 Suppl):1467S–1476S.
67. Rahbar E, Waits CMK, Kirby EH Jr, Miller LR, Ainsworth HC, Cui T, et al. Allele-specific methylation in the FADS genomic region in DNA from human saliva, CD4+ cells, and total leukocytes. *Clin Epigenet*. 2018;10:46.
68. Rahbar E, Ainsworth HC, Howard TD, Hawkins GA, Ruczinski I, Mathias R, et al. Uncovering the DNA methylation landscape in key regulatory regions within the FADS cluster. *PLoS ONE*. 2017;12(9):e0180903.
69. He Z, Zhang R, Jiang F, Zhang H, Zhao A, Xu B, et al. FADS1-FADS2 genetic polymorphisms are associated with fatty acid metabolism through changes in DNA methylation and gene expression. *Clin Epigenet*. 2018;10(1):113.
70. Veenstra J, Kalsbeek A, Koster K, Ryder N, Bos A, Huisman J, et al. Epigenome wide association study of SNP-CpG interactions on changes in triglyceride levels after pharmaceutical intervention: a GAW20 analysis. *BMC Proc*. 2018;12(Suppl 9):58.
71. Gonzalez-Becerra K, Ramos-Lopez O, Barron-Cabrera E, Riezu-Boj JJ, Milagro FI, Martinez-Lopez E, et al. Fatty acids, epigenetic mechanisms and chronic diseases: a systematic review. *Lipids Health Dis*. 2019;18(1):178.
72. Mullarky E, Mattaini KR, Vander Heiden MG, Cantley LC, Locasale JW. PHGDH amplification and altered glucose metabolism in human melanoma. *Pigment Cell Melanoma Res*. 2011;24(6):1112–5.
73. Reid MA, Allen AE, Liu S, Libert MV, Liu P, Liu X, et al. Serine synthesis through PHGDH coordinates nucleotide levels by maintaining central carbon metabolism. *Nat Commun*. 2018;9(1):5442.
74. Truong V, Huang S, Dennis J, Lemire M, Zwingerman N, Aissi D, et al. Blood triglyceride levels are associated with DNA methylation at the serine metabolism gene PHGDH. *Sci Rep*. 2017;7(1):11207.
75. Nano J, Ghanbari M, Wang W, de Vries PS, Dhana K, Muka T, et al. Epigenome-wide association study identifies methylation sites associated with liver enzymes and hepatic steatosis. *Gastroenterology*. 2017;153(4):1096.
76. Sorlie P, Wei GS. Population-based cohort studies: still relevant? *J Am Coll Cardiol*. 2011;58(19):2010–3.
77. Shah S, Bonder MJ, Marioni RE, Zhu Z, McRae AF, Zhernakova A, et al. Improving phenotypic prediction by combining genetic and epigenetic associations. *Am J Hum Genet*. 2015;97(1):75–85.
78. Wichmann HE, Gieger C, Illig T, Group MKS. KORA-gen—resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*. 2005;67(Suppl 1):S26–30.
79. Soininen P, Kangas AJ, Wurtz P, Tukiainen T, Tynkkynen T, Laatikainen R, et al. High-throughput serum NMR metabolomics for cost-effective holistic studies on systemic metabolism. *Analyst*. 2009;134(9):1781–5.
80. Inouye M, Kettunen J, Soininen P, Silander K, Ripatti S, Kumpula LS, et al. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol Syst Biol*. 2010;6:441.
81. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinform*. 2012;13:86.
82. Parkinson H, Kapushesky M, Shojatalab M, Abegunawardena N, Coulson R, Farne A, et al. ArrayExpress—a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res*. 2007;35(Database issue):D747–50.
83. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*. 2012;44(10):1084–9.
84. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am J Hum Genet*. 2013;93(5):876–90.
85. Bonder MJ, Kasela S, Kals M, Tamm R, Lolk K, Barragan I, et al. Genetic and epigenetic regulation of gene expression in fetal and adult human livers. *BMC Genomics*. 2014;15:860.
86. Zhernakova DV, Deelen P, Vermaat M, van Iterson M, van Galen M, Arindrarto W, et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet*. 2017;49(1):139–45.
87. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform*. 2011;12:77.

## Publisher's Note

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