University of Groningen

Blood lipids influence DNA methylation in circulating cells

Dekkers, Koen F.; van Iterson, Maarten; Slieker, Roderick C.; Moed, Matthijs H.; Bonder, Marc Jan; van Galen, Michiel; Mei, Hailiang; Zhernakova, Daria V.; van den Berg, Leonard H.; Deelen, Joris

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

CrossMark

Blood lipids influence DNA methylation in circulating cells

Koen F. Dekkers[1], Maarten van Iterson[1], Roderick C. Slieker[1], Matthijs H. Moed[1], Marc Jan Bonder[2],
Michiel van Galen[3], Hailiang Mei[4], Daria V. Zhernakova[2], Leonard H. van den Berg[5], Joris Deelen[1],
Jenny van Dongen[6], Diana van Heemst[7], Albert Hofman[8], Jouke J. Hottenga[6], Carla J. H. van der Kallen[9],
Casper G. Schalkwijk[9], Coen D. A. Stehouwer[9], Ettje F. Tigchelaar[2], André G. Uitterlinden[10], Gonneke Willemsen[6],
Alexandra Zhernakova[2], Lude Franke[2], Peter A. C. 't Hoen[3], Rick Jansen[11], Joyce van Meurs[10], Dorret I. Boomsma[6],
Cornelia M. van Duijn[8], Marleen M. J. van Greevenbroek[9], Jan H. Veldink[5], Cisca Wijmenga[2], BIOS Consortium[12],
Erik W. van Zwet[13], P. Eline Slagboom[1], J. Wouter Jukema[14] and Bastiaan T. Heijmans[1*]

Abstract

Background: Cells can be primed by external stimuli to obtain a long-term epigenetic memory. We hypothesize that long-term exposure to elevated blood lipids can prime circulating immune cells through changes in DNA methylation, a process that may contribute to the development of atherosclerosis. To interrogate the causal relationship between triglyceride, low-density lipoprotein (LDL) cholesterol, and high-density lipoprotein (HDL) cholesterol levels and genome-wide DNA methylation while excluding confounding and pleiotropy, we perform a stepwise Mendelian randomization analysis in whole blood of 3296 individuals.

Results: This analysis shows that differential methylation is the consequence of inter-individual variation in blood lipid levels and not vice versa. Specifically, we observe an effect of triglycerides on DNA methylation at three CpGs, of LDL cholesterol at one CpG, and of HDL cholesterol at two CpGs using multivariable Mendelian randomization. Using RNA-seq data available for a large subset of individuals (N = 2044), DNA methylation of these six CpGs is associated with the expression of *CPT1A* and *SREBF1* (for triglycerides), *DHCR24* (for LDL cholesterol) and *ABCG1* (for HDL cholesterol), which are all key regulators of lipid metabolism.

Conclusions: Our analysis suggests a role for epigenetic priming in end-product feedback control of lipid metabolism and highlights Mendelian randomization as an effective tool to infer causal relationships in integrative genomics data.

Keywords: DNA methylation, Lipids, Mendelian randomization, Gene expression

Background

External stimuli, including tobacco smoke [1], prenatal malnutrition [2], and ultraviolet radiation [3], can induce persistent changes in the epigenome. Circulating immune cells are continuously exposed to a range of stimuli present in plasma. Infectious agents or oxidized low-density lipoprotein (LDL) have been shown to affect the epigenome of monocytes, which, as a result, became protected against secondary infections [4] or acquired a

pro-inflammatory phenotype [5], respectively. Here, we investigated whether circulating immune cells are primed by blood lipids *in vivo*, a process potentially relevant for the development of atherosclerosis, which is driven by the interaction between lipids and the immune system [6].

Various population studies have reported on the association between inter-individual variation in blood lipid levels and genome-wide DNA methylation, a key component of the epigenome, in circulating immune cells [7–9]. Although sometimes interpreted as a causal effect of DNA methylation on lipid levels [7], such epigenome-wide association studies (EWASs) cannot distinguish

* Correspondence: bas.heijmans@lumc.nl
[1]Molecular Epidemiology section, Leiden University Medical Center, Einthovenweg 20, Leiden, The Netherlands
Full list of author information is available at the end of the article

Dekkers *et al. Genome Biology* (2016) 17:138

Page 2 of 12

cause and consequence [10]. Large genome-wide association studies (GWASs), however, yielded sets of single nucleotide polymorphisms (SNPs) that are robustly associated with lipid levels [11]. Since genetic variants cannot be the consequence of phenotypic variation, lipid-associated SNPs can serve as causal anchors to test whether elevated lipid levels induce DNA methylation changes in immune cells using Mendelian randomization (MR) [12–14].

To study whether blood lipids can epigenetically prime circulating immune cells, we performed a systematic MR analysis of triglycerides (TGs), LDL cholesterol (LDL-C), and high-density lipoprotein (HDL) cholesterol (HDL-C) levels using whole blood samples of 3296 individuals to interrogate the causal relationship between lipid levels and DNA methylation, while excluding confounding and pleiotropy. Using this approach to analyze both DNA methylation and transcriptome data from 2044 individuals, we identified specific differences in DNA methylation that are induced by blood lipids and could be linked to expression of genes with a well-established role in lipid metabolism.

Results

The association between blood lipids and genome-wide DNA methylation in whole blood was investigated in 3296 individuals from six cohorts (Table 1; the six cohorts are in the Biobank-based Integrative Omics Study (BIOS) Consortium, a full list of the authors of which is available in Additional file 1). All cohorts were comprised of men and women (32–60 % men) and the age of the individuals ranged from 18 to 81 years.

First, we performed EWASs of TG, LDL-C, and HDL-C levels for 453,109 CpGs (Fig. 1a and Additional file 2). For TG, we observed 21 differentially methylated CpGs (false discovery rate (FDR)-adjusted P value <0.05; Table 2) with

effect sizes ranging from –0.9 to +2.4 % change in DNA methylation per standard deviation (SD) difference in TG levels. For LDL and HDL levels, we found three and four differentially methylated CpGs with effect sizes ranging from 0.4–1.0 % per SD and 0.2–0.7 % per SD, respectively. The direction of the observed associations were consistent across the six cohorts (Additional file 3). The EWAS effect size estimates were not sensitive to the exclusion of non-fasted samples (24 % of the individuals), the exclusion of samples with imputed cell counts (15 % of the individuals), or the addition of current smoking behavior, lipid-lowering medication, or body mass index (BMI) as covariates (Additional file 4).

EWASs are prone to bias due to unmeasured confounding and, crucially, cannot be used to infer whether the associations arise from an influence of lipids on DNA methylation or DNA methylation on lipids. To address these limitations, we performed a MR analysis to determine causality while excluding confounding (Fig. 1b). We constructed weighted polygenic scores (PSs) for TG, LDL-C, and HDL-C levels (Additional file 5) using genetic variants identified in a large-scale meta-analysis of GWASs of lipid levels [11] and evaluated their validity in our own data (i.e., an F-statistic >10 and no association with known confounders). The PSs for TG (30 SNPs), LDL-C (28 SNPs), and HDL-C (60 SNPs) explained a relatively small, but highly significant, proportion of the variance (5.6 % [F = 157, $P = 1.7 \times 10^{-40}$], 3.1 % [F = 99, $P = 6.2 \times 10^{-27}$] and 5.1 % [F = 164, $P = 1.1 \times 10^{-36}$]; Additional file 6) and were not associated with confounders (Additional file 7) with the exception of the HDL-C PS, which was associated with confounder neutrophil counts (P value = 1.8×10^{-2}). However, the level of statistical significance was very weak when compared with the association between HDL-C PS and HDL-C levels. Using the PSs as unbiased predictors of lipids levels, we

Table 1 Characteristics of the six cohorts in the BIOS Consortium

	CODAM	LL	LLS	NTR	PAN	RS
Number of individuals	164	748	785	692	184	723
RNA-Seq[a]	159	616	650			619
Gender (% male)	50	42	48	32	60	42
Age (years) [SD]	65.6 [6.8]	45.6 [13.3]	58.4 [7.5]	34 [12.1]	62.4 [9.4]	67.6 [6.0]
TG (mmol/L) [SD]	1.6 [0.8]	1.2 [0.9]	1.9 [1.2]	1.3 [0.7]	1.9 [1.1]	1.5 [0.9]
LDL-C (mmol/L) [SD]	3.6 [1.0]	3.0 [0.9]	3.4 [1.0]	2.8 [0.8]	3.4 [0.9]	3.3 [0.9]
HDL-C (mmol/L) [SD]	1.3 [0.3]	1.5 [0.4]	1.4 [0.4]	1.4 [0.9]	1.4 [0.4]	1.5 [0.4]
Monocytes (%) [SD]	7.9 [1.6]	8.5 [2.2]	5.5 [1.5]	7.7 [2.0]	6.8 [1.1]	7.1 [2.1]
Lymphocytes (%) [SD]	51.2 [14.0]	34.5 [7.4]	29.2 [6.9]	35.1 [7.8]	31.8 [7.1]	36.3 [7.9]
Neutrophils (%) [SD]	40.7 [6.9]	53.3 [8.0]	59.8 [7.7]	52.6 [7.7]	55.3 [7.0]	48.3 [7.1]

[a]The number of individuals for which RNA-Seq data were available

CODAM Cohort on Diabetes and Atherosclerosis Maastricht, *LL* LifeLines, *LLS* Leiden Longevity Study, *NTR* Netherlands Twin Register, *PAN* Prospective ALS Study Netherlands, *RS* Rotterdam Study, *SD* standard deviation

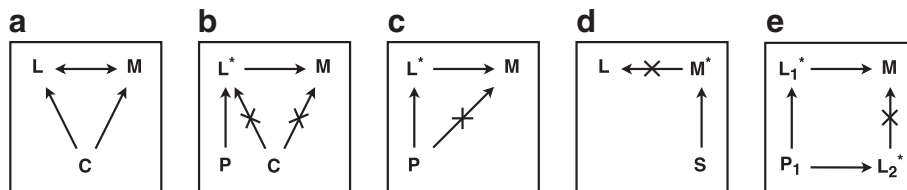Dekkers *et al. Genome Biology* (2016) 17:138

Page 3 of 12

Fig. 1 Illustration of the stepwise Mendelian randomization approach. **a** In a conventional EWAS, associations observed are potentially confounded (*C*) and the direction of the association between lipids (*L*) and DNA methylation (*M*) cannot be inferred. **b** Using Mendelian randomization, polygenic scores (*P*) are used to obtain an unconfounded proxy for lipid levels and, since M cannot influence P, an effect of L on M can be inferred. **c** An additional analysis is required to exclude a direct effect of P on M (i.e., *cis*-methylation quantitative trait loci (QTL) effect of polygenic score SNPs) not mediated through L. **d** Reverse causation, i.e., an effect of M on L, is excluded by evaluating the association of local genetic variation (*S*) affecting M (*cis*-methylation QTL) on lipid levels. **e** Pleiotropic effects are excluded using a multivariate model that incorporates all lipids and their polygenic scores

Table 2 Associations between log TG, LDL-C, and HDL-C levels and DNA methylation

		Chromosome	Position	Mean DNAm (%)	Estimate (%/SD)[a]	P_{FDR}
TG	cg14476101	1	120,255,993	63	−0.8 (−1.0, −0.6)	2.6×10^{-8}
	cg19693031	1	145,441,553	73	−0.6 (−0.8, −0.5)	5.2×10^{-9}
	cg06690548	4	139,162,809	85	−0.6 (−0.7, −0.5)	5.7×10^{-12}
	cg05575921	5	373,379	79	−0.9 (−1.3, −0.6)	4.3×10^{-3}
	cg14817490	5	392,921	32	−0.5 (−0.6, −0.3)	3.3×10^{-2}
	cg06560379	6	44,231,306	12	−0.2 (−0.2, −0.1)	4.0×10^{-2}
	cg19589396	8	103,937,375	69	−0.4 (−0.5, −0.2)	7.6×10^{-3}
	cg07504977	10	102,131,013	41	0.5 (0.3, 0.6)	2.4×10^{-4}
	cg11376147	11	57,261,199	21	−0.2 (−0.3, −0.1)	7.3×10^{-3}
	cg00574958	11	68,607,623	12	−0.4 (−0.5, −0.3)	2.4×10^{-23}
	cg17058475	11	68,607,738	14	−0.4 (−0.5, −0.3)	7.9×10^{-7}
	cg12556569	11	116,664,040	26	2.4 (1.5, 3.4)	1.4×10^{-2}
	cg15863539	17	17,716,951	79	0.2 (0.1, 0.3)	1.3×10^{-3}
	cg20544516	17	17,717,184	73	0.3 (0.2, 0.4)	1.1×10^{-3}
	cg11024682	17	17,730,095	46	0.5 (0.3, 0.6)	3.8×10^{-11}
	cg08857797	17	40,927,700	57	0.4 (0.3, 0.6)	7.3×10^{-3}
	cg03636183	19	17,000,586	68	−0.6 (−0.8, −0.3)	1.3×10^{-2}
	cg02711608	19	47,287,965	24	−0.3 (−0.4, −0.2)	3.9×10^{-2}
	cg27243685	21	43,642,367	83	0.4 (0.4, 0.5)	8.4×10^{-20}
	cg01881899	21	43,652,705	13	0.3 (0.2, 0.3)	1.7×10^{-4}
	cg06500161	21	43,656,588	60	0.7 (0.6, 0.8)	1.5×10^{-29}
LDL-C	cg27168858	1	55,351,660	78	0.4 (0.3, 0.6)	5.3×10^{-3}
	cg00908766	1	109,817,497	59	1.0 (0.6, 1.3)	2.3×10^{-2}
	cg05119988	4	166,251,190	59	0.5 (0.3, 0.6)	1.4×10^{-4}
HDL-C	cg17901584	1	55,353,707	53	0.7 (0.5, 0.9)	1.2×10^{-4}
	cg26313301	19	11,219,616	87	−0.2 (−0.3, −0.1)	3.6×10^{-2}
	cg27243685	21	43,642,367	83	−0.4 (−0.5, −0.3)	1.2×10^{-9}
	cg06500161	21	43,656,588	60	−0.7 (−0.8, −0.5)	1.3×10^{-13}

[a]Estimate is the percentage change in DNA methylation per standard deviation change in lipid levels
DNAm DNA methylation

Dekkers *et al. Genome Biology* (2016) 17:138

Page 4 of 12

confirmed an effect of TG on 9 out of 21 EWAS-identified CpGs (*P* value <0.05; Additional file 8). For LDL-C and HDL-C, an effect was found for two of the three and two of the four CpGs, respectively. Since the CpGs with higher *P* values (i.e., lowest level of statistical evidence) in the EWAS analysis were the ones that did not survive the MR approach, the multiple testing correction applied in the EWAS step (P_{FDR} <0.05) for the association with lipid levels remains applicable to the MR-identified CpG set. Interestingly, MR and EWAS estimates for the effect sizes were highly concordant for the EWAS-identified CpGs, also for CpGs not statistically significant in the MR step. The confidence intervals of the MR estimates were wider, however, in line with the moderate proportion of variance explained by the PSs (Additional file 9). Post hoc power calculations [15] were performed for three scenarios where the EWAS effect size estimate is (1) the true causal effect size, (2) half the true causal effect size, and (3) double the true causal effect size (Additional file 10). The power estimates indicated that the causal effects we identified in our MR analysis were the ones for which we had the highest statistical power.

SNPs in the PS may directly influence DNA methylation (i.e., a methylation QTL effect) [16], violating the assumption of MR that the effect is mediated through lipid levels (Fig. 1c). Of the MR-identified CpGs, three mapped within 1 Mbp of a PS SNP, for which analyses were re-run with the SNP added as covariate to the model (Additional file 11). The association of CpGs cg12556569 with TG and cg00908766 with LDL-C was mediated by a direct effect of a PS SNP *in cis* (rs964184 [mapping 15,121 bp from the CpG] and rs629301 [808 bp from the CpG]). The association of CpG cg27168858 with LDL-C, however, was not explained by a nearby PS SNP (rs2479409 [152,989 bp from the CpG]). Thus, for TG, LDL-C, and HDL-C, eight, one and two CpGs remained, respectively. Of interest, CpGs cg06500161 and cg27243685 were associated with both TG and HDL-C.

Next, we evaluated the possibility that DNA methylation affected lipid levels, i.e., the reverse of the association implied by the MR analysis. To this end, for every MR-identified CpG we established the SNP *in cis* (<100 kb) with the strongest association with methylation level defined as the lowest *P* value (Additional file 12) and used this SNP as a proxy for DNA methylation (Fig. 1d; Additional file 13). We found no evidence of an effect of DNA methylation on lipid levels at these CpGs. In addition to the MR-identified CpGs, we performed the same analysis for all EWAS-identified CpGs (Additional files 12 and 13). Two SNPs could not be evaluated because the SNP was either located in the CpG site (cg12556569) or was in linkage disequilibrium (R^2 = 0.52) with a PS SNP (cg00908766), violating MR assumptions. For all the other six CpGs, no effect of DNA

methylation on lipid levels was found. Thus, no evidence for an effect on lipid levels was observed for any of the CpGs (either surviving the previous MR steps or identified in the initial EWAS analysis) and reverse causation is, therefore, unlikely.

A final limitation precluding an unequivocal interpretation of the analysis so far is the possibility of pleiotropy. The SNPs used to construct PSs are primarily, but not solely, associated with their respective lipid level [11]. Using a multivariable MR analysis with all three lipids and their PSs (Fig. 1e), we found that the two CpGs overlapping between TG and HDL-C were driven by HDL-C while their association with TG could be attributed to pleiotropy (Table 3). For an additional three CpGs associated with TG, pleiotropic effects could not be formally excluded (*P* value >0.05). The effect of TG, LDL-C, and HDL-C on the remaining three, one, and two CpGs, respectively, was not driven by pleiotropic effects. To exclude the occurrence of unmeasured and unknown pleiotropic effects, we applied Egger regression [17]. This analysis showed that there was no net pleiotropic effect (Table 4). However, while multivariable MR indicated that the association for the two CpGs overlapping between TG and HDL-C could be attributed to an effect of HDL-C, Egger regression indicated that it was due to TG. Table 5 summarizes the outcome from the sequential steps of our stepwise MR approach on EWAS-identified CpGs.

The CpGs resulting from the stepwise MR analysis were annotated to blood cell chromatin states [18], which inform on the biological role of the genomic region harboring the CpG. This analysis showed that all CpGs were located in regulatory regions active in blood cells (Table 6). To link the differentially methylated CpGs to a specific gene, we tested for an association between the methylation level of the six CpGs resulting from the MR analysis and gene expression *in cis* (<100 kb from transcription start site (TSS)). Remarkably, all CpGs were associated with gene expression (*P* value <0.05; Table 6; Additional file 14) and, in all cases, it was the gene in which the CpG was located. The TG-influenced CpGs cg00574958 and cg17058475 (located in a region flanking an active TSS and in an enhancer, respectively) were both associated with expression of *CPT1A* and CpG cg11024682 (located in an enhancer region) was linked to *SREBF1* expression. The LDL-C-influenced CpG cg27168858 (located in an active TSS flanking region) mapped to *DHCR42*. Finally, the TG- (Egger regression) or HDL-C- (multivariable MR) influenced CpGs cg06500161 and cg27243685 (located in a region of weak transcription and flanking an active TSS, respectively) were associated with *ABCG1* expression. All these genes have a central role in lipid metabolism.

Dekkers *et al. Genome Biology* (2016) 17:138

Page 5 of 12

Table 3 Estimated effect of TG, LDL-C, and HDL-C levels on DNA methylation using a multivariable MR analysis to exclude pleiotropic effects

		Chromosome	Position	Mean DNAm (%)	Estimate (%/SD)[a]	P value
TG	cg00574958	11	68,607,623	12	−0.7 (−1.2, −0.3)	1.5×10^{-3}
	cg17058475	11	68,607,738	14	−1.0 (−1.6, −0.3)	2.5×10^{-3}
	cg11024682	17	17,730,095	26	1.0 (0.3, 1.7)	5.4×10^{-3}
LDL-C	cg27168858	1	55,351,660	78	1.4 (0.5, 2.3)	3.1×10^{-3}
HDL-C	cg27243685	21	43,642,367	83	−0.7 (−1.1, −0.3)	2.2×10^{-3}
	cg06500161	21	43,656,588	60	−1.1 (−1.8, −0.5)	3.5×10^{-4}

[a]*Estimate* is the percentage change in DNA methylation per standard deviation change in lipid levels
DNAm DNA methylation

Discussion

We performed EWASs of blood lipids in a meta-analysis comprising 3296 individuals followed by a stepwise MR approach to evaluate the association of blood lipids with DNA methylation in circulating immune cells and infer the direction of causality. Our analysis showed that differential methylation was induced by TG at three CpGs, by LDL-C at one CpG, and by either TG or HDL-C at two CpGs. We did not observe evidence for the reverse relationship, that is, an effect of DNA methylation on lipid levels. These data indicate that blood lipids can epigenetically prime immune cells, which may prove relevant for disease phenotypes with an inflammatory component [4, 5].

All the CpGs we identified mapped to regions with active regulatory roles in blood cells and were associated with the expression of genes with a central role in lipid metabolism. Higher TG levels induced lower methylation of CpGs cg00574958 and cg17058475, which was associated with higher expression of *CPT1A*. Higher TG levels also induced higher methylation of CpG cg11024682, which was associated with lower expression of *SREBF1*. *CPT1A* attaches carnitine to long-chain fatty acids, which is required for entry into the

mitochondria and subsequent catabolism [19]. *SREBF1* regulates energy homeostasis, including genes involved in the synthesis, import, and efflux of lipids. Sterols, such as cholesterol, inhibit the effect of *SREBF1* [20]. Higher LDL-C levels induced higher methylation of a CpG in *DHCR24*, which was linked to lower *DHCR24* expression. *DHCR24* catalyzes the reduction of desmosterol to cholesterol, the last step in the Bloch cholesterol biosynthesis pathway [21]. Either lower TG or higher HDL-C levels induced lower methylation of CpGs cg06500161 and cg27243685, which in turn were associated with higher expression of *ABCG1*. *ABCG1* mediates cellular cholesterol efflux to HDL in the reverse cholesterol transport pathway. The reverse cholesterol transport pathway transfers cholesterol from peripheral tissues via the blood to the liver [22]. Taken together, higher TG and LDL-C and lower HDL-C levels lead to DNA methylation changes that are associated with, on the one hand, increased catabolism and cellular export and, on the other hand, decreased cellular import of lipids. These findings suggest a role for epigenetic priming in end-product feedback control, a process where the end-product inhibits its own synthesis and that has been observed for cholesterol [23].

Table 4 Estimated effect of TG, LDL-C, and HDL-C levels on DNA methylation using an approach based on Egger regression to exclude pleiotropic effects, including P values for an estimate of net pleiotropic effects of the PS SNPs

		Chr	Position	Mean DNAm (%)	Estimate (%/SD)[a]	P value	Pleiotropy P value
TG	cg00574958	11	68,607,623	12	−1.1 (-1.4, -0.7)	5.7×10^{-9}	0.37
	cg17058475	11	68,607,738	14	−1.3 (-1.8, -0.8)	2.3×10^{-7}	0.82
	cg11024682	17	17,730,095	26	0.9 (0.3, 1.5)	1.9×10^{-3}	0.50
	cg27243685	21	43,642,367	83	0.7 (0.3, 1.1)	6.4×10^{-4}	0.18
	cg06500161	21	43,656,588	60	1.4 (0.8, 2.0)	2.9×10^{-6}	0.75
LDL-C	cg27168858	1	55,351,660	78	0.7 (0.1, 1.3)	3.1×10^{-2}	0.36
HDL-C	cg27243685	21	43,642,367	83	0.3 (-0.9, 1.5)	0.62	0.79
	cg06500161	21	43,656,588	60	−0.1 (-1.8, 1.7)	0.93	0.15

[a]*Estimate* is the percentage change in DNA methylation per standard deviation change in lipid levels
DNAm DNA methylation

Dekkers *et al. Genome Biology* (2016) 17:138

Page 6 of 12

Table 5 Overview of the stepwise Mendelian randomization approach

		b	c	d	e
TG	cg14476101	✗			
	cg19693031	✓	✓	✓	✗
	cg07504977	✓	✓	✓	✗
	cg11376147	✗			
	cg00574958	✓	✓	✓	✓
	cg17058475	✓	✓	✓	✓
	cg12556569	✓	✗		
	cg15863539	✗			
	cg20544516	✗			
	cg11024682	✓	✓	✓	✓
	cg08857797	✗			
	cg03636183	✗			
	cg02711608	✗			
	cg27243685	✓	✓	✓	✓[1]
	cg01881899	✓	✓	✓	✗
	cg06500161	✓	✓	✓	✓[1]
	cg06690548	✗			
	cg05575921	✗			
	cg14817490	✗			
	cg06560379	✗			
	cg19589396	✗			
LDL-C	cg27168858	✓	✓	✓	✓
	cg00908766	✓	✗		
	cg05119988	✗			
HDL-C	cg17901584	✗			
	cg26313301	✗			
	cg27243685	✓	✓	✓	✓[2]
	cg06500161	✓	✓	✓	✓[2]

Starting from associations between lipid levels and DNA methylation, a causal and unconfounded effect of lipid levels on DNA methylation was estimated (b), a direct effect of genetic variants on DNA methylation was excluded (c), reverse causation was excluded (d) and pleiotropy was excluded (e)
A *tick* indicates a CpG passed step; a *cross* indicates a CpG failed step
[1]Egger regression only
[2]Multivariable MR only

Our analysis replicated previous EWASs reporting the association between TG and two CpGs in *CPT1A* in CD4+ T cells [7] and whole blood [8]. In addition, eight out of ten differentially methylated CpGs found in the most recent EWAS of various blood lipids were replicated [9] (7/8 CpGs replicated for TG, 0/1 for LDL-C, 1/1 for HDL-C). The CpGs that we did not replicate were cg20544516 associated with TG (rank 111 in our analysis, $P = 1 \times 10^{-6}$) and cg22178392 associated with LDL-C (rank 308,850 in our analysis, $P = 0.68$). In contrast to earlier interpretations of these findings [9], our MR analysis indicates that differential methylation is the

consequence of inter-individual variation in blood lipid levels instead of its cause. Interestingly, EWASs of lipid levels also overlapped with those of other outcomes. An EWAS of BMI reported differential methylation at *CPT1A* and *ABCG1* [24], and an EWAS of type 2 diabetes at *SREBF1* and *ABCG1* [25]. Our MR analysis using polygenic scores of lipid levels predicts that the altered lipid profile in blood that is associated with BMI and type 2 diabetes was driving these outcomes.

The stepwise MR approach we implemented circumvents various issues commonly encountered in MR. We used a weighted polygenic score (PS) for the MR analysis with weights estimated in a large lipid GWAS [11] to test the hypothesis that lipid levels affected DNA methylation. The PS optimally captured the inter-individual variance in lipid levels and is less prone to violations of MR assumptions compared with single genetic variants [26]. Furthermore, we could exclude the violation of the assumption in MR that genetic variants in the PS have a direct effect on DNA methylation by testing the association of lipid-associated differential methylation with PS SNPs *in cis*. This contrasts with most MR studies where a direct effect of genetic variants on the outcome is unknown [12]. Remaining key issues in MR, namely reverse causation and pleiotropic effects, were addressed using a bidirectional multivariable model [27] and an approach based on Egger regression, which accounts for unmeasured and unknown pleiotropic effects [17]. Together these analyses excluded the occurrence of reverse causation (here an effect of DNA methylation on lipid levels) and an involvement of pleiotropy, supporting a causal role of lipid levels in its association with DNA methylation in circulating immune cells. Surprisingly, multivariable MR and Egger regression yielded opposite results for the two *ABCG1* CpGs that were associated with both TG and HDL-C. While multivariable MR indicated that the association could be attributed to an effect of HDL-C, Egger regression indicated that it was due to TG. These discrepant outcomes may be indicative of the limits of MR analysis in the presence of a substantial correlation between evaluated phenotypes, as is the case for lipid levels. We presume that the result of the multivariable MR analysis is closer to the biological mechanism in this case because the analysis uses the actual measured lipid levels. Moreover, this interpretation is consistent with the biological role of *ABCG1*, which is a key mediator of cellular cholesterol efflux to HDL particles.

We performed our analysis in a large meta-analysis of cohorts with matched genotype and genome-wide DNA methylation data. However, MR analysis has a relatively low power; the lipid PSs capture only about 5 % of the total variance in lipid levels. Therefore, we adopted an approach in which we performed a discovery phase using a regular EWAS approach followed by validation

Dekkers *et al. Genome Biology* (2016) 17:138

Page 7 of 12

Table 6 DNA methylation of MR-identified CpGs was associated with gene expression of genes involved in lipid metabolism

		Chr	Position	Mean DNAm (%)	Gene	Estimate (logCPM/β)[a]	*P* value	Chromatin state
TG	cg00574958	11	68,607,623	12	*CPT1A*	−3.3 (−4.1,−2.5)	1.5×10^{-14}	Flanking active TSS
	cg17058475	11	68,607,738	14	*CPT1A*	−1.7 (−2.3,−1.1)	1.8×10^{-8}	Enhancer
	cg11024682	17	17,730,095	26	*SREBF1*	−2.4 (−2.8,−2.0)	5.2×10^{-33}	Enhancer
LDL-C	cg27168858	1	55,351,660	78	*DHCR24*	−4.1 (−4.9,−3.3)	2.4×10^{-25}	Flanking active TSS
HDL-C	cg27243685	21	43,642,367	83	*ABCG1*	−4.5 (−5.4,−3.6)	4.0×10^{-24}	Flanking active TSS
	cg06500161	21	43,656,588	60	*ABCG1*	−5.1 (−5.7,−4.5)	1.2×10^{-56}	Weak transcription

CpGs are located in regions of active gene regulation
[a]*Estimate* is the log counts per million change in gene expression per β (0–1) change in DNA methylation
DNAm DNA methylation

and causal inference using a stepwise MR approach. Only for a subset of EWAS findings were we able to infer the direction of causality in the MR approach. These were the ones for which our study had a relatively high statistical power. However, it was striking that the effect size estimates from the EWASs were generally highly consistent with those from the MR analysis, which is compatible with an effect of lipid levels on DNA methylation for more CpGs. One may argue that individual steps in our MR approach require additional multiple testing. This would have further limited our findings in the multivariable MR approach to those for LDL-C and HDL-C, for which fewer associated CpGs were observed and which are thus less affected by additional multiple testing adjustment. We believe that this would be too conservative, since the association of the lipid levels with DNA methylation of specific CpGs was already corrected for multiple testing in the EWAS. These considerations have two implications. First, larger studies of blood lipid are expected to detect additional CpGs influenced by lipid levels. Second, sample sizes should ideally be increased to achieve sufficient statistical power to directly evaluate the association of PSs with DNA methylation on a genome-wide scale. This would remove the necessity of a discovery phase using EWAS to preserve statistical power and would resolve the issue pertaining to the question of additional adjustment for multiple testing in a stepwise approach. Second, most of the resulting associations will be explained by an effect of blood lipids on DNA methylation. The power to test reverse causal relationships (that is of DNA methylation on blood lipids) can be improved by the availability of comprehensive catalogues of methylation QTL enabling testing both *in cis* and *in trans*.

Cell counts are known confounders in EWASs measuring whole blood [10, 28] and were included as a covariate in the analyses. However, this also impedes the discovery of cell type-specific processes. Hence, it remains to be determined if specific immune cells are particularly prone to epigenetic priming by lipid levels. However, previous EWASs of lipid levels in adipose

tissue (cg27243685 and cg11024682) and in skin biopsies (cg11024682) reported differential methylation at CpG sites also identified in whole blood [9]. This suggests that the lipid priming does not occur exclusively in circulating immune cells. A recent study of the TG-lowering drug fenofibrate suggested that a 3-week daily treatment was not sufficient to reverse lipid-associated DNA methylation changes [29] despite the short half-life of various blood cell types. Together, the presence of lipid-induced changes across tissues and the inability to reverse these changes in circulating cells in the short-term open the possibility that these changes did not arise in the circulation but occurred already in hematopoietic stem cells, where lipid priming has previously been observed [30].

Conclusions

We implemented a systematic MR approach to uncover the occurrence of epigenetic priming on immune cells and show that it can be used effectively when genetic variants for the stimulus of interest are available. Our findings imply that differential methylation observed in EWASs may frequently be a consequence rather than a cause of an outcome of interest. Future studies should establish whether epigenetic priming plays a role in the etiological path from risk factor to disease outcome. We anticipate that with the rapid increase in the availability of genomics, epigenomics, transcriptomics and metabolomics data, approaches like ours will be increasingly used to unravel causal relationships in integrative genomics approaches [14].

Methods
Cohorts

The Biobank-based Integrative Omics Study (BIOS) Consortium [31, 32] consists of the six Dutch cohorts Cohort on Diabetes and Atherosclerosis Maastricht (CODAM) [33], LifeLines (LL) [34], Leiden Longevity Study (LLS) [35], the Netherlands Twin Register (NTR) [36], Rotterdam Study (RS) [37], and Prospective ALS Study Netherlands (PAN) [38]. For all 3296 unrelated

Dekkers *et al. Genome Biology* (2016) 17:138

Page 8 of 12

individuals genotypes, DNA methylation and blood profiles (including lipid levels and cell counts) were measured in whole blood, which was collected simultaneously for all measurements. RNA-Seq data were also available for 2044 individuals. DNA methylation and RNA-Seq data were generated by the Human Genotyping facility (HugeF) of ErasmusMC, the Netherlands (http://www.glimdna.org/). Characteristics of the cohorts can be found in Table 1.

Lipid measurements and cell counts

Triglyceride (TG), HDL cholesterol (HDL-C), and total cholesterol levels (TC) were measured after a fasting period of 12 h for CODAM, LL, NTR (for 678/692 individuals), RS, and PAN; for LLS non-fasted lipids were measured for 757 of the 785 individuals. LDL cholesterol (LDL-C) was calculated using Friedewald's method [39]. For all analyses, TG levels were log-transformed and all lipid levels were scaled to have mean 0 and standard deviation 1.

White blood cell counts (WBC), i.e. neutrophils, lymphocytes, monocytes, eosinophils, and basophils, were measured by the standard WBC differential as part of the complete blood count (CBC). However, a minority of samples were lacking CBC measurements (15 %) or did not differentiate between granulocyte subtypes (neutrophils, eosinophils, and basophils; 37 %). Since DNA methylation levels are informative for the white blood cell composition [40], a predictor was built to infer the white blood cell composition of those samples using partial least squares, which can handle both multivariate responses and high-dimensional covariates. The R package *pls* [41] was used to fit the model and to optimize the number of pls components using fivefold cross-validation, including age and gender as covariates. The predictor was trained on two-thirds of the samples with complete WBC data (N = 1364) and tested on one-third (N = 682) of samples across cohorts to obtain a validated predictor. The correlations between measured and predicted leukocyte subtype percentages in the test set were 0.84, 0.86, and 0.68 for neutrophils, lymphocytes, and monocytes, respectively. The R code and detailed documentation for the function *wbccPredictor* are available from https://github.com/mvaniterson/wbccPredictor.

Genotypes

SNPs were measured per cohort (for data generation details see Tigchelaar et al. [34] for LL, Deelen et al. [42] for LLS, Willemsen et al. [36] for NTR, and Hofman et al. [37] for RS), harmonized (*Genotype Harmonizer* [43]), and imputed (*Impute2* [44]) using GoNL5 [45] as reference. SNPs with an imputation info-score <0.5, Hardy–Weinberg equilibrium *P* value $<10^{-4}$, call rate <95 % or minor allele frequency <0.05 were removed.

DNA methylation

Using the Zymo EZ DNA methylation kit (Zymo Research, Irvine, CA, USA), 500 ng of genomic DNA was bisulfite-converted and 4 µl of bisulfite-converted DNA was measured on the Illumina 450 k array using the manufacturer's protocol (Illumina, San Diego, CA, USA). Quality control and normalization of the data were done according to Tobi et al. [46]. In brief, sample outliers were detected and removed using *MethylAid* [47] (95 samples removed), probes with a detection *P* value >0.01, bead number <3, or zero intensity were removed, as well as the ambiguously mapped probes (34,064 probes removed) [48]. This yielded a set of 453,109 CpGs measured in 3296 samples. Next, the DNA methylation data were normalized using the functional normalization [49] approach as implemented in *minfi* [50] (with five principal components of the control probes). Samples where 5 % of the probes failed or probes where 5 % of the samples failed were excluded (0 samples and 0 probes removed). The final dataset consisted of 453,109 CpGs from 3296 individuals. The R code for the quality control and normalization pipeline is available at https://git.lumc.nl/molepi/Leiden450K.

Gene expression

Total RNA libraries were generated using the TruSeq v2 library protocol and 2 × 50-bp paired-end sequencing was performed on the Illumina Hiseq2000. Reads passing Illumina's Chastity filter were produced using *CASAVA* and quality control was done with *FastQC* [51], *cutadapt* [52] (adapter trimming), and *Sickle* [53] (removal of low-quality read ends). Reads were aligned to the human genome (build NCBI37) using *STAR* [54]. Gene quantifications were obtained as the total number of reads that aligned to the exons of a gene as annotated by Ensembl v.71. Subsequently, gene counts were normalized using *cqn* [55] to correct for gene- and sample-specific GC biases. Linear model fitting and inference were performed using *voom* [56] and *limma* [57]. This resulted in gene expression data for 2044 of the 3296 individuals (Table 1).

Statistical analysis

All analyses were performed using R [58]. For the EWAS of lipid levels, the association of DNA methylation with TG, HDL-C, and LDL-C was evaluated per cohort using the linear model defined by Eq. 1:

$$DNAm_i = \beta_0 + \beta_1 \cdot lipid_i + \beta_2 \cdot gender_i + \beta_3 \cdot age_i \\ + \beta_4 \cdot \%mono_i + \beta_5 \cdot \%lymph_i \\ + \beta_6 \cdot \%neut_i + \beta_7 \cdot bisplate_i \\ + \beta_8 \cdot position_i + \varepsilon_i \ (i = 1, ..., 3296) \quad (1)$$

where %mono, %lymph, and %neut are the percentages of monocytes, lymphocytes, and neutrophils, bisplate is the bisulfite plate number, and position is the position on the 450 k array.

Dekkers *et al. Genome Biology* (2016) 17:138

Page 9 of 12

To minimize possible inflation in the test statistic, the genomic control procedure [59] was applied to achieve an inflation factor λ of 1.0. Nominal *P* values and standard errors were calculated from the resulting corrected t-statistics. The outcomes of the six cohorts were combined in a fixed-effect meta-analysis using *metaphor* [60], followed by genomic control on the resulting z-statistics (λ = 1.0). *P* values were calculated from corrected z-values and adjusted for multiple testing using the Benjamini–Hochberg (FDR) procedure.

Polygenic scores (PSs) were constructed from GWAS [11] SNPs to serve as an instrument in MR analysis. For each SNP, dosages were calculated using Eq. 2:

$$Dosage = 0 \cdot AA + 1 \cdot AB + 2 \cdot BB \tag{2}$$

where AA, AB, and BB are the measured allele frequencies.

SNPs primarily associated with their respective lipid level (Additional file 5) were used to calculate the PSs as defined by Eq. 3:

$$PS = \frac{dosage_1 \cdot ES_1 + dosage_2 \cdot ES_2 + \ldots + dosage_N \cdot ES_N}{\text{mean}\left(ES_{1, \ldots, N}\right)} \tag{3}$$

where ES is the effect size of the reported SNP–lipid association.

The PSs were scaled to have mean 0 and standard deviation 1. The association of the PS with lipid levels was confirmed using a meta-analysis of the outcomes of a model similar to the model described in Eq. 1 with PS as outcome instead of DNA methylation.

MR analysis was done using *AER* [61]. For each cohort a two-stage least-squares model (Eq. 4 and Eq. 5), a model suitable for epidemiological studies [12], was fitted per CpG followed by a meta-analysis of the results.

Stage 1:

$$
\begin{aligned}
predlipid_i = {} & \gamma_0 + \gamma_1 \cdot PS_i + \gamma_2 \cdot gender_i + \gamma_3 \cdot age_i \\
& + \gamma_4 \cdot \%mono_i + \gamma_5 \cdot \%lymph_i \\
& + \gamma_6 \cdot \%neut_i + \gamma_7 \cdot bisplate_i \\
& + \gamma_8 \cdot position_i \\
& + v_i \ (i = 1, \ldots, 3296)
\end{aligned}
\tag{4}
$$

Stage 2:

$$
\begin{aligned}
DNAm_i = {} & \beta_0 + \beta_1 \cdot predlipid_i + \beta_2 \cdot gender_i \\
& + \beta_3 \cdot age_i + \beta_4 \cdot \%mono_i \\
& + \beta_5 \cdot \%lymph_i + \beta_6 \cdot \%neut_i \\
& + \beta_7 \cdot bisplate_i + \beta_8 \cdot position_i \\
& + \varepsilon_i (i = 1, \ldots, 3296)
\end{aligned}
\tag{5}
$$

where predlipid is the predicted lipid level (TG, LDL-C, HDL-C) based on its polygenic score.

Power calculations for two-stage least-squares-based MR studies [15] were performed using the R implementation (https://github.com/kn3in/mRnd).

To address whether the PS SNPs directly affect DNA methylation as a *cis*-methylation QTL effect instead of through an effect on lipid levels, genotypes of SNPs (as dosage) within 1 Mb of CpGs resulting from the MR analysis were included as an additional covariate in the MR model (Eq. 4 and Eq. 5).

To test whether DNA methylation affected lipid levels, instead of lipid levels affecting DNA methylation (the direction of the effect evaluated in the PS-based MR analysis), SNPs were identified that were *cis*-methylation QTLs of CpGs associated with lipid levels found using the model defined by Eq. 6:

$$
\begin{aligned}
DNAm_i = {} & \beta_0 + \beta_1 \cdot dosage_i + \beta_2 \cdot gender_i \\
& + \beta_3 \cdot age_i + \beta_4 \cdot \%mono_i \\
& + \beta_5 \cdot \%lymph_i + \beta_6 \cdot \%neut_i \\
& + \beta_7 \cdot bisplate_i + \beta_8 \cdot position_i \\
& + \varepsilon_i (i = 1, \ldots, 3296)
\end{aligned}
\tag{6}
$$

Subsequently, the dosage of these meQTL SNPs was entered into the MR model as an instrument for DNA methylation (Eq. 7 and Eq. 8) and the results were meta-analyzed.

Stage 1:

$$
\begin{aligned}
predDNAm_i = {} & \gamma_0 + \gamma_1 \cdot meqtl_i + \gamma_2 \cdot gender_i \\
& + \gamma_3 \cdot age_i + \gamma_4 \cdot \%mono_i \\
& + \gamma_5 \cdot \%lymph_i + \gamma_6 \cdot \%neut_i \\
& + \gamma_7 \cdot bisplate_i + \gamma_8 \cdot position_i \\
& + v_i \ (i = 1, \ldots, 3296)
\end{aligned}
\tag{7}
$$

Stage 2:

$$
\begin{aligned}
lipid_i = {} & \beta_0 + \beta_1 \cdot predDNAm_i + \beta_2 \cdot gender_i \\
& + \beta_3 \cdot age_i + \beta_4 \cdot \%mono_i + \beta_5 \cdot \%lymph_i \\
& + \beta_6 \cdot \%neut_i + \beta_7 \cdot bisplate_i \\
& + \beta_8 \cdot position_i + \varepsilon_i (i = 1, \ldots, 3296)
\end{aligned}
\tag{8}
$$

where predDNAm is the predicted DNA methylation level based on its methylation QTL and meqtl is the dosage of the methylation QTL.

To account for pleiotropic effects (as SNPs in the PSs are often associated with multiple lipids), multivariable instrumental variable analysis was done using the three PSs and three lipids in a two-stage least-squares model similar to the model defined by Eq. 4 and Eq. 5, followed by a meta-analysis of the results. This method performs well for lipid pleiotropy correction [27]. We compared the multivariable MR approach with an approach based on Egger regression [17]. For the association between lipids and PS SNPs required for Egger regression we used the estimates reported in a GWAS [11] and for the

Dekkers *et al. Genome Biology* (2016) 17:138

Page 10 of 12

association between PS SNPs and DNA methylation we used Eq. 9. We combined the results in a meta-analysis.

$$
\begin{aligned}
DNAm_i = \gamma_0 &+ \gamma_1 \cdot dosage_i + \gamma_2 \cdot gender_i \\
&+ \gamma_3 \cdot age_i + \gamma_4 \cdot \%mono_i \\
&+ \gamma_5 \cdot \%lymph_i + \gamma_6 \cdot \%neut_i \\
&+ \gamma_7 \cdot bisplate_i + \gamma_8 \cdot position_i \\
&+ v_i \ (i = 1, ..., 3296)
\end{aligned} \quad (9)
$$

DNA methylation was associated with expression of genes within 100 kb in *limma* [57] using the linear model defined by Eq. 10 per cohort, after which the results were meta-analyzed.

$$
\begin{aligned}
Expression_i = \beta_0 &+ \beta_1 \cdot DNAm_i + \beta_2 \cdot gender_i \\
&+ \beta_3 \cdot age_i + \beta_4 \cdot \%mono_i \\
&+ \beta_5 \cdot \%lymph_i + \beta_6 \cdot \%neut_i \\
&+ \beta_7 \cdot bisplate_i + \beta_8 \cdot position_i \\
&+ flowcell_i + \varepsilon_i (i = 1, ..., 2044)
\end{aligned} \quad (10)
$$

where flowcell is the Hiseq2000 flowcell.

Additional files

Additional file 1: Full list of authors of the BIOS Consortium. (PDF 68 kb)

Additional file 2: Quantile–quantile plots depicting the associations between TG, LDL-C, and HDL-C and genome-wide DNA methylation when compared with a uniform distribution expected by chance for a uncorrected meta-analysis z-values and b meta-analysis z-values corrected for genome-wide inflation using genomic control ($\lambda = 1.0$). (PDF 14800 kb)

Additional file 3: Estimates and 95 % confidence intervals for the EWAS-identified CpGs for each cohort contributing to the meta-analysis for TG, LDL-C, and HDL-C. (PDF 297 kb)

Additional file 4: The EWAS effect size estimates were not sensitive to the exclusion of non-fasted samples, the exclusion of samples with imputed cell counts, or the addition of current smoking behavior, lipid-lowering medication, or BMI as covariates. (PDF 273 kb)

Additional file 5: The lipid-associated SNPs used to construct the polygenic scores. (CSV 1 kb)

Additional file 6: Polygenic scores constructed from lipid GWAS-identified SNPs associate strongly with their respective lipid levels but explain only a small amount of their variance. (CSV 104 bytes)

Additional file 7: The HDL-C PS is nominally associated with confounder neutrophil counts. (CSV 202 bytes)

Additional file 8: Mendelian randomization reveals an effect of lipid levels on DNA methylation for some of the EWAS-identified CpGs. Estimate is the percentage change in DNA methylation per standard deviation change in lipid levels. (CSV 1 kb)

Additional file 9: Comparison between EWAS and MR estimates (with 95 % confidence intervals). (PDF 109 kb)

Additional file 10: Power calculations were performed for the situations where the EWAS effect size estimate is (1) the true causal effect size, (2) half the true causal effect size, and (3) double the true causal effect size. (CSV 1 kb)

Additional file 11: Effect of lipid levels on DNA methylation corrected for nearest PS SNP within 1 Mb. Estimate is the percentage change in DNA methylation per standard deviation change in lipid levels. (CSV 289 bytes)

Additional file 12: Association between DNA methylation and genetic variation for the CpGs and their strongest associating SNP within 100 kb. (CSV 1 kb)

Additional file 13: No effect of DNA methylation on lipid levels was found using the strongest associating SNP within 100 kb for each CpG as an instrument. Estimate is the standard deviation change in lipid levels per percentage change in DNA methylation. (CSV 1 kb)

Additional file 14: DNA methylation was associated with gene expression. (PDF 21974 kb)

Availability of data and materials
Discovery data, generated within the BIOS consortium, are available from the European Genome-phenome Archive (EGA) under accession number EGAC00001000277.

Authors' contributions
KFD, JWJ, and BTH designed the study. MB, MvG, HM, DVZ, LHvdB, JD, JvD, DvH, AH, JJH, CJHvdK, CGS, CDAS, EFT, AGU, GW, AZ, LF, PACtH, RJ, JvM, DIB, CMvD, MMJvG, JHV, CW, EWvZ, and PES contributed to establishing the data infrastructure. KFD, MvI, RCS, MHM, and BTH performed the analyses. KFD and BTH drafted the manuscript. MvI, RCS, MB, JvD, PACtH, DIB, MMJvG, JHV, PES, and JWJ were involved in critical revisions of the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Ethics approval and consent to participate
The study was approved by the institutional review boards of the participating centers (CODAM, Medical Ethical Committee of the Maastricht University; LL, Ethics committee of the University Medical Centre Groningen; LLS, Ethical committee of the Leiden University Medical Center; PAN, Institutional review board of the University Medical Centre Utrecht; NTR, Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre; RS, Institutional review board (Medical Ethics Committee) of the Erasmus Medical Center). All participants have given written informed consent and the experimental methods comply with the Helsinki Declaration.

Author details
[1]Molecular Epidemiology section, Leiden University Medical Center, Einthovenweg 20, Leiden, The Netherlands. [2]Department of Genetics, University of Groningen, University Medical Centre Groningen, Broerstraat 5, Groningen, The Netherlands. [3]Department of Human Genetics, Leiden University Medical Center, Einthovenweg 20, Leiden, The Netherlands. [4]Sequence Analysis Support Core, Leiden University Medical Center,

Dekkers *et al. Genome Biology* (2016) 17:138

Page 11 of 12

Einthovenweg 20, Leiden, The Netherlands. [5]Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, The Netherlands. [6]Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, De Boelelaan 1117, Amsterdam, The Netherlands. [7]Department of Gerontology and Geriatrics, Leiden University Medical Center, Einthovenweg 20, Leiden, The Netherlands. [8]Department of Genetic Epidemiology, ErasmusMC, 's-Gravendijkwal 230, Rotterdam, The Netherlands. [9]Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, P. Debyelaan 25, Maastricht, The Netherlands. [10]Department of Internal Medicine, ErasmusMC, 's-Gravendijkwal 230, Rotterdam, The Netherlands. [11]Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, De Boelelaan 1117, Amsterdam, The Netherlands. [12]BIOS Consortium, Einthovenweg 20, Leiden, The Netherlands. [13]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Einthovenweg 20, Leiden, The Netherlands. [14]Department of Cardiology, Leiden University Medical Center, Einthovenweg 20, Leiden, The Netherlands.

References

1. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One. 2013;8, e63812.
2. Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, et al. DNA methylation signatures link prenatal famine exposure to growth and metabolism. Nat Commun. 2014;5:5592.
3. Vandiver AR, Irizarry RA, Hansen KD, Garza LA, Runarsson A, Li X, et al. Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. Genome Biol. 2015;16:80.
4. Saeed S, Quintin J, Kerstens HH, Rao NA, Aghajanirefah A, Matarese F, et al. Epigenetic programming of monocyte-to-macrophage differentiation and trained innate immunity. Science. 2014;345:1251086.
5. Bekkering S, Quintin J, Joosten LA, van der Meer JW, Netea MG, Riksen NP. Oxidized low-density lipoprotein induces long-term proinflammatory cytokine production and foam cell formation via epigenetic reprogramming of monocytes. Arterioscler Thromb Vasc Biol. 2014;34:1731–8.
6. Hansson GK, Hermansson A. The immune system in atherosclerosis. Nat Immunol. 2011;12:204–12.
7. 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;55:1324–30.
8. Gagnon F, Aissi D, Carrie A, Morange PE, Tregouet DA. Robust validation of methylation levels association at CPT1A locus with lipid plasma levels. J Lipid Res. 2014;55:1189–91.
9. 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;8:334–42.
10. Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet. 2013;14:585–94.
11. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45:1274–83.
12. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet. 2014;23:R89–98.
13. Relton CL, Davey SG. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol. 2012;41:161–76.
14. Dekkers KF, Slagboom PE, Jukema JW, Heijmans BT. The multifaceted interplay between lipids and epigenetics. Curr Opin Lipidol. 2016;27:288–94.
15. Brion MJ, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. Int J Epidemiol. 2013;42:1497–501.
16. Bell JT, Pai AA, Pickrell JK, Gaffney DJ, Pique-Regi R, Degner JF, et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. Genome Biol. 2011;12:R10.
17. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512–25.
18. Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518:317–30.
19. McGarry JD, Brown NF. The mitochondrial carnitine palmitoyltransferase system. From concept to molecular analysis. Eur J Biochem. 1997;244:1–14.
20. Espenshade PJ, Hughes AL. Regulation of sterol synthesis in eukaryotes. Annu Rev Genet. 2007;41:401–27.
21. Zerenturk EJ, Sharpe LJ, Ikonen E, Brown AJ. Desmosterol and DHCR24: unexpected new directions for a terminal step in cholesterol synthesis. Prog Lipid Res. 2013;52:666–80.
22. Rader DJ, Tall AR. The not-so-simple HDL story: is it time to revise the HDL cholesterol hypothesis? Nat Med. 2012;18:1344–6.
23. Brown MS, Goldstein JL. Cholesterol feedback: from Schoenheimer's bottle to Scap's MELADL. J Lipid Res. 2009;50(Suppl):S15–27.
24. Demerath EW, Guan W, Grove ML, Aslibekyan S, Mendelson M, Zhou YH, et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. Hum Mol Genet. 2015;24:4464–79.
25. 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;3:526–34.
26. Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol. 2013;42:1134–44.
27. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. Am J Epidemiol. 2015;181:251–60.
28. Paul DS, Beck S. Advances in epigenome-wide association studies for common diseases. Trends Mol Med. 2014;20:541–3.
29. Das M, Irvin MR, Sha J, Aslibekyan S, Hidalgo B, Perry RT, et al. Lipid changes due to fenofibrate treatment are not associated with changes in DNA methylation patterns in the GOLDN study. Front Genet. 2015;6:304.
30. Seijkens T, Hoeksema MA, Beckers L, Smeets E, Meiler S, Levels J, et al. Hypercholesterolemia-induced priming of hematopoietic stem and progenitor cells aggravates atherosclerosis. FASEB J. 2014;28:2202–13.
31. Bonder MJ, Luijk R, Zhernakova D, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. bioRxiv. 2015:033084.
32. Zhernakova D, Deelen P, Vermaat M, van Iterson M, van Galen M, Arindrarto W, et al. Hypothesis-free identification of modulators of genetic risk factors. bioRxiv. 2015:033217.
33. van Greevenbroek MM, Jacobs M, van der Kallen CJ, Vermeulen VM, Jansen EH, Schalkwijk CG, et al. The cross-sectional association between insulin resistance and circulating complement C3 is partly explained by plasma alanine aminotransferase, independent of central obesity and general inflammation (the CODAM study). Eur J Clin Invest. 2011;41:372–9.
34. Tigchelaar EF, Zhernakova A, Dekens JA, Hermes G, Baranska A, Mujagic Z, et al. Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. BMJ Open. 2015;5, e006772.
35. Schoenmaker M, de Craen AJ, de Meijer PH, Beekman M, Blauw GJ, Slagboom PE, et al. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. Eur J Hum Genet. 2006;14:79–84.
36. Willemsen G, Vink JM, Abdellaoui A, den Braber A, van Beek JH, Draisma HH, et al. The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. Twin Res Hum Genet. 2013;16:271–81.
37. Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA, et al. The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol. 2014;2013(28):889–926.
38. Huisman MHB, de Jong SW, van Doormaal PTC, Weinreich SS, Schelhaas HJ, van der Kooi AJ, et al. Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. J Neurol Neurosurg Psychiatry. 2011;82:1165–70.
39. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem. 1972;18:499–502.
40. 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 Bioinformatics. 2012;13:86.

Dekkers *et al. Genome Biology* (2016) 17:138

Page 12 of 12

41. Mevik B-H, Wehrens R. The pls Package: Principal Component and Partial Least Squares Regression in R. J Stat Softw. 2007;18:2.

42. Deelen J, Beekman M, Uh HW, Broer L, Ayers KL, Tan Q, et al. Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. Hum Mol Genet. 2014;23:4420–32.

43. Deelen P, Bonder MJ, van der Velde KJ, Westra HJ, Winder E, Hendriksen D, et al. Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. BMC Res Notes. 2014;7:901.

44. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5, e1000529.

45. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat Genet. 2014;46:818–25.

46. Tobi EW, Slieker RC, Stein AD, Suchiman HE, Slagboom PE, van Zwet EW, et al. Early gestation as the critical time-window for changes in the prenatal environment to affect the adult human blood methylome. Int J Epidemiol. 2015;44:1211–23.

47. van Iterson M, Tobi EW, Slieker RC, den Hollander W, Luijk R, Slagboom PE, et al. MethylAid: visual and interactive quality control of large Illumina 450 k datasets. Bioinformatics. 2014;30:3435–7.

48. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8:203–9.

49. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450 k methylation array data improves replication in large cancer studies. Genome Biol. 2014;15:503.

50. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30:1363–9.

51. Andrews S. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

52. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2.

53. Joshi N, Fass J. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). 2011. https://github.com/najoshi/sickle.

54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

55. Hansen KD, Irizarry RA, Wu ZJ. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics. 2012;13:204–16.

56. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15:R29.

57. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

58. R Core Team. R: A Language and Environment for Statistical Computing. 2014. http://www.r-project.org/.

59. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55:997–1004.

60. Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36:1–48.

61. Kleiber C, Zeileis A. Applied econometrics with R. Springer-Verlag; 2008