

PROCEEDINGS

Open Access



# Disentangling associations between DNA methylation and blood lipids: a Mendelian randomization approach

Sergi Sayols-Baixeras<sup>1,2,3</sup>, Hemant K. Tiwari<sup>4</sup> and Stella W. Aslibekyan<sup>5\*</sup>

From Genetic Analysis Workshop 20  
San Diego, CA, USA. 4 - 8 March 2017

## Abstract

**Background:** DNA methylation is an epigenetic mechanism that has been proposed as a possible link between genetic and environmental determinants of disease. Prior studies reported robust associations between the methylation of specific cytosine-phosphate-guanine (CpG) sites and plasma lipids, namely triglycerides (TGs) and high-density lipoprotein cholesterol (HDL-C). However, the causality of the observed association remains elusive, hampered by weak instrumental variables for methylation status.

**Aim:** We present a novel application of the elastic net approach to implement a bidirectional Mendelian randomization approach to inferring causal relationships between candidate CpGs and plasma lipids in GAW20 data.

**Methods:** We used DNA methylation, TGs, and HDL-C measured during the visit 2. Based on prior findings, we selected 5 methylation markers (cg00574958, cg07504977, cg06690548, cg19693031, and cg03717755) related to TGs, 2 markers (cg09572125 and cg02650017) related to HDL-C, and 2 markers (cg06500161 and cg11024682) related to both traits. We implemented an elastic net approach to improve the selection of the genetic instrument for the methylation markers, followed by bidirectional Mendelian randomization 2-stage least-squares regression.

**Results:** We observed causal effects of blood fasting TGs on the methylation levels of cg00574958 (*CPT1A*) and cg06690548 (*SLC7A11*). For cg00574958, our findings were also consistent with the reverse direction of association, that is, from *CPT1A* methylation to TGs.

**Conclusions:** Current evidence does not rule out either direction of association between the methylation of the cg00574958 *CPT1A* locus and plasma TGs, highlighting the complexity of lipid homeostasis. We also demonstrated a novel approach to improve instrument selection in DNA methylation studies.

## Background

Fasting blood lipids are independent modifiable risk factors for cardiovascular disease, the leading cause of death worldwide [1, 2]. Like many other complex traits, fasting blood lipids have a heritable component, but known DNA sequence variants only explain a small (< 12% cumulatively) proportion of their variation [3]. An emerging body of evidence supports DNA methylation, which refers to the addition of a methyl group to the DNA molecule, as a

more promising contributor to the missing heritability of lipids [4–7]. For example, methylation of one locus in *CPT1A* explained 11.6% of plasma triglyceride variation in a prior epigenome-wide study in the Genetics of the Lipid Lowering Drugs and Diet Network (GOLDN) [4].

In contrast to DNA sequence variants that are inherited from parents and persist through the offspring's lifetime, methylation markers can be inherited as well as modified by lifestyle and environmental factors [8]. Therefore, the associations reported in previous cross-sectional epigenetic studies of fasting blood lipids have a variety of possible causal interpretations [9]. One method to test the specific causal scenarios (eg, lipids affecting methylation patterns

\* Correspondence: [saslibek@uab.edu](mailto:saslibek@uab.edu)

<sup>5</sup>Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, 1665 University Blvd, Birmingham, AL, USA  
Full list of author information is available at the end of the article



or vice versa) is Mendelian randomization (MR), which uses genetic markers (single-nucleotide polymorphisms [SNPs]) as instrumental variables, taking advantage of the natural randomization that occurs at conception [10]. A study by Dekkers et al. [7] implemented stepwise MR to establish the causal effect of lipids on methylation; however, the presented approach was not truly bidirectional as it was limited in selecting instrumental variables for methylation (ie, *cis*-methylation quantitative trait loci [*cis*-meQTL]). Therefore, the reverse effect of methylation on lipids has not been rigorously tested and cannot be ruled out.

Using data from the GAW20, we aimed to fully interrogate bidirectional relationships between plasma lipids and methylation at 5 methylation markers related to triglycerides (TGs), 2 related to high-density lipoprotein cholesterol (HDL-C), and 2 related to both traits, with selection based on prior evidence [6]. Furthermore, we present a novel approach for selecting SNP proxies for epigenomic variants, using GAW20 data to test the potential of penalized regression, specifically elastic net models, to identify *cis*-meQTL instruments.

## Methods

### Phenotypes and covariates

We used TGs and HDL-C measured during the visit 2 as the phenotypic traits of interest. Both traits were log transformed to normalize their distributions. We selected 5 cytosine-phosphate-guanine (CpG) sites (cg00574958, cg07504977, cg06690548, cg19693031, and cg03717755) related to TGs, 2 CpG sites (cg09572125 and cg02650017) related to HDL-C, and 2 CpG sites (cg06500161 and cg11024682) related to both lipid measures in a previous study from our group [6]. These CpG sites are located in the genes *CPT1A*, *SLC7A11*, *TXNIP*, *MYLIP*, *SYNGAP1*, *PHOSPHO1*, *ABCG1*, and *SREBF1*, and an intergenic region on chromosome 10, respectively. During the analyses, we included age, sex, center, and smoking status as fixed effects, and the family relatedness as a random effect.

### Analysis pipeline

We applied the MR method to interrogate the causal association between lipid traits and DNA methylation. The MR method is predicated upon several assumptions: (1) a reliable association between the genetic instrument and the exposure; (2) associations between the instrument and the outcome must only be mediated through exposure; (3) no pleiotropic effects of the instrument [7, 11]. In the first step of our analysis, we investigated associations between the selected CpG sites and the lipid traits of interest in the GAW20 data. Second, we verified assumption (1) by evaluating associations between a previously validated polygenic risk score as an instrument for lipids (PRS-L) [7] and DNA methylation in the GAW20 data set.

Third, we ensured that our polygenic risk score was not associated with methylation other than through its effect on lipid levels, testing assumption (2). To that end, we fitted 2 models, adjusted and unadjusted, for the lipids predicted by the PRS-L. Fourth, we investigated the possibility of reverse causality using a polygenic risk score as an instrumental variable for DNA methylation (PRS-M), which we built using an elastic net approach (detailed below), and testing its effect on lipids. Finally, we assessed the net unmeasured pleiotropic effects [assumption (3)] using the Egger test [12]. For a truly bidirectional approach, we applied these steps in the opposite direction (from methylation to lipids) for all CpG sites that met the Bonferroni threshold (0.05/number of tests) in the first step of the analysis.

### Associations between DNA methylation and lipids

Using the *nlme::R* package [13], we fitted a linear mixed model with DNA methylation beta score as the independent variable and the lipids as the dependent variables, adjusting for the covariates as described above. Methylation status of specific CpG sites was deemed to be significantly associated with lipids if the *p* values met the Bonferroni cutoff of 0.05/7 CpG sites = 0.0071.

### Causal effects of lipids on DNA methylation

We evaluated the causal effects of lipids on DNA methylation using the two-stage least-squares (TSLS) approach [10]. Briefly, TSLS comprises 2 regression stages. In the first stage, the exposure (lipids) is regressed on the genetic instrument (PRS-L) to obtain the values of the exposure predicted by the genetic instrument (lipids|PRS-L). In the second stage, the outcome (DNA methylation) is regressed on the predicted values for the exposure (lipids|PRS-L) from the first stage. Thus, in this second regression, the causal coefficient is estimated [14].

First, we modified a previously validated genetic risk score for lipids [7] based on the availability of its constituent SNPs in GAW20 data. We used 20 available SNPs on the GAW20 data, out of the 28 SNPs proposed by Dekkers et al. [7]. Of these 20 SNPs, 8 were genotyped in GAW20 and 12 were proxy SNPs selected by SNAPtool with an  $r^2 > 0.8$  [15]. Once we selected the SNPs, we built the PRS-L as  $\frac{\sum_{i=1}^N genotype_i \cdot ES_i}{\text{mean}(ES)}$ , where *genotype* is the number of risk alleles carried at a given locus, *N* is the number of SNPs used to build the PRS-L, and *ES* is the effect size. We scaled the PRS-L to obtain a mean of 0 and SE of 1.

Second, we applied the TSLS to estimate the causal effects of lipids on DNA methylation. The first regression was fit to test the association between PRS-L and lipids using a linear mixed-model approach adjusted for the covariates according to the following equation:

$$\text{predict}_L = \overline{\beta_0 + \beta_1 * \text{PRS-L} + \beta_2 * \text{age} + \beta_a * \text{sex} + \beta_4 * \text{center} + \beta_5 * \text{smoking} + \beta_6 * \text{family}}$$

(throughout this article, the single line over the text refers to fixed effects and the double line refers to random effects).

The second regression model estimated the causal effect of circulating lipids on DNA methylation:

CpG site methylation

$$= \overline{\beta_0 + \beta_1 * \text{predict}_L + \beta_2 * \text{age} + \beta_3 * \text{sex} + \beta_4 * \text{center} + \beta_5 * \text{smoking} + \beta_6 * \text{family}}$$

We also tested whether PRS-L was associated with methylation independently of predicted lipids using the following model:

CpG site methylation

$$= \overline{\beta_0 + \beta_1 * \text{PRS-L} + \beta_2 * \text{age} + \beta_3 * \text{sex} + \beta_4 * \text{center} + \beta_5 * \text{smoking} + \beta_6 * \text{predict}_L + \beta_7 * \text{family}}$$

### Causal effects of DNA methylation on lipids

To determine the causal effect of methylation on lipids, we followed the same TSLS approach, starting with selecting the appropriate instrument for methylation. We selected all the SNPs located  $\pm 50$  kb from the methylation marker as possible *cis*-meQTL. Then we fitted the linear mixed models to obtain the residuals of the association between methylation and the covariates as follows:

CpG site methylation

$$= \overline{\beta_0 + \beta_1 * \text{age} + \beta_2 * \text{sex} + \beta_3 * \text{center} + \beta_4 * \text{smoking} + \beta_5 * \text{family}}$$

Subsequently, we used an elastic net approach to find the SNPs associated with the methylation marker with a coefficient that is statistically significantly different from zero. We set the elastic net algorithm to the following options: alpha = 0.5, lambda = lambda.min obtained from the cross-validation model and the seed = "123".

*Elastic net model:* CpG site methylation\* =  $\beta_0 + \beta_1 * \text{SNP}_1 + \beta_2 * \text{SNP}_2 \dots + \beta_n * \text{SNP}_n$  where CpG site methylation\* is the residual from the previous equation, and  $n$  refers to all the SNPs located  $\pm 50$  kb from the CpG site that are not directly on the probe.

We tested the relationship between our selected *cis*-meQTL and the CpG site methylation as follows:

CpG site methylation

$$= \overline{\beta_0 + \beta_1 * \text{meQTL} + \beta_2 * \text{age} + \beta_3 * \text{sex} + \beta_4 * \text{center} + \beta_5 * \text{smoking} + \beta_6 * \text{family}}$$

Once the SNPs were selected, we created and standardized a PRS-M using the approach outlined in our description of PRS-L above.

Subsequently, we applied the TSLS approach with lipids as the outcome to estimate the causal effect of DNA methylation on lipids, and tested whether PRS-M was related to lipids independently of predicted methylation.

As the final step, we tested for net pleiotropic effects using the MR-Egger test implemented in the *MendelianRandomization:R* package [16].

## Results

### Associations between DNA methylation and lipids

After removing the individuals with missing data, 993 individuals remained in the analyses. Of all tested CpG sites, five (cg00574958, cg11024682, cg07504977, cg06690548, and cg06500161) were associated with TGs (Table 1) and none were associated with HDL-C in GAW20 data (data not shown). Consequently, all subsequent analyses were restricted to the TG phenotype.

### Causal effects of lipids on DNA methylation

Data from 655 individuals were available for MR analyses. The polygenic risk score for TG was robustly associated with the trait and associated with methylation of 2 (cg00574958 and cg06690548) of the 5 CpG sites (see Table 1). PRS-L was not associated with methylation of these 2 loci independently of the predicted TG levels (data not shown). PRS-L was not significantly associated with the other CpG sites. Thus, those results do not support a causal effect of TG on DNA methylation at cg07504977, cg11024682, and cg06500161.

### Causal effects of DNA methylation on lipids

We implemented the elastic net approach and created 2 PRS-Ms for cg00574958 (*CPT1A*; 3 SNPs), and cg06500161 (*ABCG1*; 5 SNPs) (see Table 1). The respective PRS-Ms were associated with the methylation of the cg00574958

and cg06500161 sites (see Table 1). The predicted methylation of the cg00574958 was associated with TG (see Table 1), but predicted methylation of the cg06500161 was not associated with TG ( $p$  value = 0.47).

### Pleiotropic effects

We tested the pleiotropy for the genetic instruments for the cg00574958 (*CPT1A*) using the MR-Egger test, which suggested no pleiotropic effect across the genetic variants in PRS-L and PRS-M for cg00574958.

## Discussion

Using GAW20 data, we assessed causal relations between fasting blood lipids and methylation from lipids to methylation. We observed causal effects of lipids on 2 methylation loci, but we could only investigate reverse causation for 1 locus because of the lack of appropriate instruments. The estimated associations between methylation and lipids were consistent with previous observational studies [4–6], but our conclusions diverged from prior MR findings [7].

Specifically, we established that methylation levels of cg00574958 (*CPT1A*) and cg06690548 (*SLC7A11*) can be affected by circulating TGs. The largest study of lipid epigenomics to date [7] also reported a causal effect of TGs on the methylation of the *CPT1A* locus, but not vice versa (ie, from cg00574958 methylation to TG). In contrast, we present novel evidence for a causal effect of

**Table 1** Summary of the statistically significant results in the GAW20 data

TG–DNA methylation					
CpG	Chr: Pos	Gene	B	SE	$p$ Value
cg06690548	Chr4: 139162809	<i>SLC7A11</i>	−0.97	0.07	$2.99 \times 10^{-4}$
cg07504977	Chr10: 102131013	NA	0.36	0.008	$6.19 \times 10^{-5}$
cg00574958	Chr11: 68607622	<i>CPT1A</i>	−1.00	0.02	$1.29 \times 10^{-14}$
cg11024682	Chr17: 17730095	<i>SREBF1</i>	0.53	0.02	$4.71 \times 10^{-5}$
cg06500161	Chr21: 43656587	<i>ABCG1</i>	0.54	0.02	$5.61 \times 10^{-4}$
TG–PRS-L					
PRS-L			B	SE	$p$ Value
PRS-L			0.05	$7.68 \times 10^{-6}$	$5.99 \times 10^{-9}$
DNA methylation–Predicted TG					
CpG	Chr: Pos	Gene	B	SE	$p$ Value
cg06690548	Chr4: 139162809	<i>SLC7A11</i>	−0.05	$2.86 \times 10^{-4}$	$2.84 \times 10^{-3}$
cg00574958	Chr11: 68607622	<i>CPT1A</i>	−0.12	$1.17 \times 10^{-3}$	$3.97 \times 10^{-4}$
DNA methylation–PRS-M					
PRS-M	CpG	Gene	B	SE	$p$ Value
PRS-M <sub><i>CPT1A</i></sub>	cg00574958	<i>CPT1A</i>	0.006	$5.80 \times 10^{-6}$	$1.08 \times 10^{-2}$
PRS-M <sub><i>ABCG1</i></sub>	cg06500161	<i>ABCG1</i>	0.008	$3.77 \times 10^{-6}$	$2.61 \times 10^{-5}$
TG–Predicted Methylation					
CpG	Chr: Pos	Gene	B	SE	$p$ Value
cg00574958	Chr11: 68607622	<i>CPT1A</i>	−2.36	0.37	$1.28 \times 10^{-4}$

cg00574958 methylation on fasting TGs. Our comprehensive bidirectional approach was enabled by a novel application of elastic net models to create a comprehensive polygenic methylation score. Although we were able to replicate and expand on the *CPT1A* finding, we did not detect other previously reported causal effects, possibly as a result of our smaller sample size: TG → cg11024682 and TG/HDL → cg06500161 [7]; additionally, we did not have robust genetic instruments to interrogate DNA methylation effects on lipids for other loci.

All 2 regions harboring CpG sites that emerged as causally associated in our analyses have extensive biological implications for lipid homeostasis. *CPT1A* encodes the liver isoform of carnitine palmitoyltransferase 1, a key enzyme in the fatty acid metabolism pathway; the cg00574958 locus specifically has been linked to plasma lipid levels [4–7] and lipoprotein subfractions [17]. In the same way, the *SLC7A11* (Solute carrier family 7 member 11) has been related to TGs [6, 7] and it has an important role protecting cells from oxidative stress [18].

## Conclusions

To conclude, we cannot rule out either direction of association between DNA methylation loci (namely in *CPT1A*) and TG blood levels, illustrating the complexity of biological regulation of lipid traits. Our findings likely paint only a part of the underlying causal picture. We did not have strong genetic instruments to test reverse causation for other lipid-associated CpG sites, highlighting the limitations of MR. Future studies should consider expanding the regions included in the elastic net (eg, to ±100 kb) and integrating publicly available bioinformatics data to improve the capture of *cis*-meQTLs to create robust genetic instruments for DNA methylation.

## Funding

Publication of this article was supported by NIH R01 GM031575. SSB was funded by the Instituto de Salud Carlos III-Fondos FEDER (IFI14/00007) and a grant from Fundació Privada Daniel Bravo Andreu. SWA was funded by NIH NHLBI K01 HL136700.

## Availability of data and materials

The data that support the findings of this study are available from the Genetic Analysis Workshop (GAW) but restrictions apply to the availability of these data, which were used under links for the current study. Qualified researchers may request these data directly from GAW.

## About this supplement

This article has been published as part of BMC Proceedings Volume 12 Supplement 9, 2018: Genetic Analysis Workshop 20: envisioning the future of statistical genetics by exploring methods for epigenetic and pharmacogenomic data. The full contents of the supplement are available online at <https://bmcpoc.biomedcentral.com/articles/supplements/volume-12-supplement-9>.

## Authors' contributions

SSB, HKT, and SWA implemented the method, SSB analyzed the data, SSB and SWA wrote the paper. SSB, HKT, and SWA read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

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

## Author details

<sup>1</sup>Cardiovascular Epidemiology and Genetics Research Group, IMIM (Hospital del Mar Medical Research Institute), 88 Dr. Aiguader Street, 08003 Barcelona, Catalonia, Spain. <sup>2</sup>Universitat Pompeu Fabra (UPF), 80 Dr. Aiguader Street, 08003 Barcelona, Catalonia, Spain. <sup>3</sup>CIBER Cardiovascular Diseases (CIVERC), 88 Dr. Aiguader Street, 08003 Barcelona, Catalonia, Spain. <sup>4</sup>Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, 1665 University Blvd, Birmingham, AL, USA. <sup>5</sup>Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, 1665 University Blvd, Birmingham, AL, USA.

Published: 17 September 2018

## References

- O'Donnell CJ, Elosua R. Cardiovascular risk factors insights from Framingham heart study. *Rev Esp Cardiol.* 2008;61(3):299–310.
- Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, de Ferranti S, Després JP, Fullerton HJ, Howard VJ, et al. Heart disease and stroke statistics—2015 update: a report from the American Heart Association. *Circulation.* 2015;131(4):e29–e322.
- Surakka I, Horikoshi M, Mägi R, Sarin AP, Mahajan A, Lagou V, Marullo L, Ferreira T, Miraglio B, Timonen S, et al. The impact of low-frequency and rare variants on lipid levels. *Nat Genet.* 2015;47(6):589–97.
- Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, Thibault KS, Patel N, Day K, Jones LW, 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, Holdt LM, Kretschmer A, Schramm K, Adamski J, et al. DNA methylation of lipid-related genes affects blood lipid levels. *Circ Cardiovasc Genet.* 2008;8(2):334–42.
- Sayols-Baixeras S, Subirana I, Lluís-Ganella C, Civeira F, Roquer J, Do AN, Absher D, Cenarro A, Muñoz D, Soriano-Tárraga C, 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.
- Dekkers KF, van Iterson M, Slieker RC, Moed MH, Bonder MJ, van Galen M, Mei H, Zhernakova DV, van den Berg LH, Deelen J, et al. Blood lipids influence DNA methylation in circulating cells. *Genome Biol.* 2016;17(1):138.
- Trerotola M, Relli V, Simeone P, Alberti S. Epigenetic inheritance and the missing heritability problem. *Hum Genomics.* 2015;9(1):17.
- Mill J, Heijmans BT. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet.* 2013;14(8):585–94.
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23(R1):R89–98.
- Latvala A, Ollikainen M. Mendelian randomization in (epi)genetic epidemiology: an effective tool to be handled with care. *Genome Biol.* 2016;17(1):156.
- 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(2):512–25.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3; 2016. p. 1–128.
- Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res.* 2015; [Epub ahead of print]

15. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008;24(24):2938–9.
16. Yavorska O, Burgess S: MendelianRandomization: Mendelian Randomization Package. R package version 0.2.0, 2016.
17. Frazier-Wood AC, Aslibekyan S, Absher DM, Hopkins PN, Sha J, Tsai MY, Tiwari HK, Waite LL, Zhi D, Arnett DK. Methylation at CPT1A locus is associated with lipoprotein subfraction profiles. *J Lipid Res*. 2014;55(7):1324–30.
18. Mandal PK, Seiler A, Perisic T, Kölle P, Banjac Canak A, Förster H, Weiss N, Kremmer E, Lieberman MW, Bannai S, et al. System xc<sup>-</sup> and thioredoxin reductase 1 cooperatively rescue glutathione deficiency. *J Biol Chem*. 2010; 29:22224–53.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

