commentary

Robust validation of methylation levels association at *CPT1A* locus with lipid plasma levels[1]

France Gagnon,* Dylan Aïssi,[†,§,] Alain Carrié,[§,**,††] Pierre-Emmanuel Morange,[§§,***,†††] and David-Alexandre Trégouët[2,†,§,**]**

Division of Epidemiology,* Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1166, Team *Genomics & Pathophysiology of Cardiovascular Diseases*,[†] F-75013, Paris, France; INSERM,[§] UMR_S 1166, F-75013, Paris, France; ICAN Institute for Cardiometabolism and Nutrition,** F-75013, Paris, France; Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1166, Team *Integrative Biology of Atherosclerosis*,[††] F-75013, Paris, France; Aix-Marseille University, UMR_S 1062, Nutrition Obesity and Risk of Thrombosis,[§§] F-13385, Marseille, France; INSERM, UMR_S 1062, Nutrition Obesity and Risk of Thrombosis,*** F-13385, Marseille, France; and Laboratory of Haematology,[†††] La Timone Hospital, F-13385, Marseille, France

There is increasing enthusiasm regarding the use of biobanked whole blood DNA as a model to discover methylation marks associated with biological phenotypes and generate novel mechanistic hypotheses (1–3). DNA methylation has a critical role in cell functions and is cell-type specific. Such cell specificity makes DNA methylation particularly challenging for epidemiological epigenetic investigations because disease relevant cell types might not be accessible due to practical issues such as availability, ethics, and cost associated with more complex specimen collection.

Recent work suggests that agnostic methylation-wide association scan (MWAS) in peripheral blood can reflect phenotype-associated methylation marks in other tissues and cell types, with effects detected in established effector cells much stronger than effects detected in blood (4). These observations suggest that marks detected in blood are associated with functions in effector cells. The Illumina HumanMethylation450 (HM450K) array is a robust assay to measure DNA methylation across the genome (4–7). For any high-throughput technologies, and in particular for a novel assay such as the HM450K, rigorous quality control procedures are warranted and robustness of findings must be validated through independent replication to avoid reporting spurious associations.

In the current issue of the *Journal*, Frazier-Wood et al. (8) reported the novel findings of significant negative correlations between methylation levels at two CpG sites in the *CPT1A* locus and plasma levels of VLDL and LDL. Methylation levels were assessed in CD4+ T-cells isolated from peripheral blood DNA using the HM450K array. Given that no independent study samples were available for replication, to circumvent this challenge, the authors adopted an internal validation method by splitting the whole sample into "discovery and replication subsamples". This strategy provides arguments in favor of the discovered associations but does not provide evidence of robustness against spurious findings due to sampling or confounding biases or any other undetected biases present in the study sample. A robust and thorough validation strategy implies the use of independent study samples and variation in the study designs (9, 10). The validation phase is of particular importance in MWAS, as this technique is particularly subjected to confounders (3). Thus, we undertook to test for associations the two *CTP1A* CpG sites found associated with lipid-related traits by Frazier-Wood et al. using two independent study samples with considerable variations in their respective study design and with the design of the Frazier-Wood study.

The studies had differences in sampling scheme, DNA methylation specimen, and array preprocessing approaches. The notable differences in the design and sample characteristics between the three studies are shown in **Table 1**. Most notable is the method for lipid measurement, nuclear magnetic resonance spectroscopy in Frazier-Wood et al. and spectrophotometry in our studies. In addition

D.A. was supported by a PhD grant from the Région Ile de France (CORDDIM). The MARTHA project was supported by a grant from the Program Hospitalier de Recherche Clinique. The F5L Thrombophilia French-Canadian Pedigree study was supported by grants from the Canadian Institutes of Health Research (MOP86466) and by the Heart and Stroke Foundation of Canada (T6484). The Human450Meethylation epityping was partially funded by the Canadian Institutes of Health Research (MOP86466) and by the Heart and Stroke Foundation of Canada (grant T6484). F.G. holds a Canada Research Chair. Statistical analyses were performed using the C2BIG computing cluster, funded by the Région Ile de France, Pierre and Marie Curie University, and the ICAN Institute for Cardiometabolism and Nutrition (ANR-10-IAHU-05).

[1]See referenced article, *J. Lipid Res.* 2014, 55: 1324–1330.

[2]To whom correspondence should be addressed.
e-mail: david.tregouet@upmc.fr

DOI 10.1194/jlr.E051276

to sampling variation, and of particular interest for MWAS studies, Frazier-Wood et al. assessed DNA methylation in isolated CD4+ T-cells, while we assessed methylation in peripheral whole blood, which includes CD4+ T-cell (<30%) and several other leukocyte subtypes. Finally, different normalization procedures were used: we applied the SWAN methodology (4, 11) to globally normalize β values from the Infinium I and II probes, while separate normalization by probe type was applied by Frazier-Wood et al.

Despite the nontrivial differences between these studies, we observed strong statistical evidence for a negative association between the two *CTP1A* CpG sites (cg00574958 and cg17058475) identified by Frazier-Wood et al. and plasma levels of both LDL and triglycerides (TG) in two independent studies, the MARTHA (4, 12) and F5L-pedigree studies (13). In our two samples totaling 526 individuals, increased DNA methylation levels at *CPTA1* CpG sites were associated with both decreased LDL and TG (**Table 2**). A 1% increase in cg00574958 DNA methylation levels was associated with a 0.057 ± 0.011 decrease in log TG levels ($P = 5.71\ 10^{-8}$). Corresponding values for a 1% increase in cg17058475 levels were 0.030 ± 0.008 ($P = 9.83\ 10^{-5}$).

Of note, cg00574958 and cg17058475 were highly correlated ($\rho_{spearman} = 0.67$ in both studies, $P < 10^{-16}$); adjusting for cg00574958 in the model abolished the effect observed for cg17058475 on log TG levels. Finally, after adjustment for key covariates (age, sex, BMI, cell type composition, batch, and chip effects), cg00574958 explained ~4% of log TG plasma levels, both in MARTHA and F5L-pedigrees. Negative association was also observed between plasma LDL levels and cg17058475 ($P = 1.7\ 10^{-2}$) but not with cg00574958 ($P = 0.11$). No association was observed with HDL-cholesterol levels ($P = 0.96$ for cg00574958 and $P = 0.75$ for cg17058475), nor with total cholesterol levels ($P = 0.16$ for cg00574958 and $P = 0.53$ for cg17058475).

The CPT1A protein is essential for fatty acid oxidation (a multistep process that metabolizes fats and converts them into energy) and is expressed in the liver and glandular tissues (14). This pivotal role in fatty acid metabolism makes *CPT1A* DNA methylation marks relevant to many metabolic disorders (from lipids to glucose homeostasis). The lipid-related DNA methylation probes in this study (cg00574958 and cg17058475) are designated as falling in a single "CpG shore", and are flanked by two CpG islands. Human ENCODE HM450K studies performed on over 40 cell lines suggest these two probes show more variable methylation levels than the two CpG islands that flank them. The uncoupled methylation levels at these probes versus the flanking islands suggest that the observed variation

TABLE 1. Main design and sample characteristics of the three MWAS studies on lipids

Study name	MARTHA	F5L-Pedigrees	GOLDEN (Frazier-Wood et al.)	
Study design	Unrelated individuals	Extended pedigrees	Extended pedigrees	
Subjects Origin	Caucasians from Marseille area (South of France)	French-Canadians from Ottawa area (Canada)	European descent from Minneapolis (Minnesota) and Salt Lake City (Utah)	
			Discovery	Validation
N	327	199	663	331
Age	44.1 (14.23)	39.6 (16.9)	48.6 (16.4)	47.7 (16.6)
Sex (% male)	21.7	46.7	47	49.2
Total cholesterol	5.452 (1.019) (g/L)	4.896 (1.079) (g/L)	NA	NA
HDL-cholesterol	1.476 (0.435) (g/L)	1.359 (0.353) (g/L)	40.0 (5.6) (nmol/L)	37.0 (5.8) (nmol/L)
LDL-cholesterol	3.647 (0.980) (g/L)	3.111 (0.901) (g/L)	1393.8 (460.0) (nmol/L)	1369.5 (1369.5) (nmol/L)
Triglycerides	1.058 (0.772) (mmol/L)	1.487 (0.905) (mmol/L)	NA	NA
Lipid measurement technology	Spectrophotometry except for LDL that was derived from the Friedewald's formula		Nuclear Magnetic Resonance spectroscopy	
Blood collection	Fasting	Fasting and non smoking	Fasting	
DNA specimen	Whole blood		Isolated CD4+ T-cells	
Medication	No exclusion	Exclusion if on medication	Asked to discontinue the use of lipid lowering drugs and over-the-counter medication that could affect lipid levels.	
HumanMethylation450k Normalization	Noob[a] and SWAN[b]		Separately normalized probes from the Infinium I and II using ComBat[c]	
Adjustment	Age, sex, batch effect, chip effect, cell type composition, dyslipidemia	Age, sex, batch effect, chip effect, cell type composition,[d] family structure	Age, sex, study site, T-cell purity (based on the first 4 principal components), family structure	

[a] Triche et al. 2013. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucl. Acids Res.* 41: e90.
[b] Ref. 11.
[c] Johnson et al. 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 8: 118–127.
[d] In MARTHA, specific measured biological counts of lymphocytes, monocytes, neutrophils, eosinophils and basophils were used to characterize leukocytes composition. In F5L-pedigrees, adjustment for cell type composition was handled by the methods described in Houseman et al. (BMC Bioinformatics 2012;13:86)

TABLE 2. Association of cg00574958 and cg17058475 *CPT1A* CpG variability with plasma TG and LDL levels in the MARTHA and F5L-pedigrees

		TG (log)	LDL
cg00574958	MARTHA	−0.059 (0.013) $P = 8.28\ 10^{-6}$	−0.023 (0.040) $P = 0.57$
	F5L-pedigrees	−0.054 (0.018) $P = 3.28\ 10^{-3}$	−0.046 (0.019) $P = 0.12$
	Combined[a]	−0.057 (0.011) $P = 5.71\ 10^{-8}$	−0.038 (0.024) $P = 0.11$
cg17058475	MARTHA	−0.025 (0.009) $P = 8.86\ 10^{-3}$	−0.051 (0.029) $P = 8.57\ 10^{-2}$
	F5L-pedigrees	−0.041 (0.013) $P = 2.88\ 10^{-3}$	−0.037 (0.022) $P = 9.36\ 10^{-2}$
	Combined	−0.030 (0.008) $P = 9.83\ 10^{-5}$	−0.042 (0.017) $P = 1.7\ 10^{-2}$

Association was tested using a linear regression model (mixed linear model in F5L-Pedigrees) where log(TG) (LDL, resp.) was the outcome and the CpG site the predictor variable. Analyses were adjusted for age, sex, cell type, batch and chip effects. Reported coefficients (standard error) represent the increase in outcome value associated with a 1% increase in CpG site variability. In MARTHA, TG and LDL phenotypes were measured in 327 and 180 individuals, respectively. In the F5L-pedigrees study, lipid phenotypes were measured in 199 individuals.

[a] Results of the MARTHA and F5L-pedigrees studies were combined into a random effect meta-analysis based on the inverse-variance weighting method.

is more likely to be regulatory. This region also shows evidence of open chromatin through DNase I hypersensitivity assays (15) and gene regulatory potential through chromatin immunoprecipitation sequencing of the epigenetic modification H3K27ac (16). More work is needed to understand the functional impact of DNA methylation on *CPT1A* gene regulation.

Three important conclusions emerge from this validation study. First, despite limitations in the Frazier-Wood et al. replication approach, the published results are robust to variation in sample, study design, normalization procedures, and even DNA blood specimen type. Second, inter-individual variation in lipid-related traits appears to be under the influence of DNA methylation regulation at the *CPT1A* locus. This epidemiological evidence now requires technical validation and functional work to confirm that these methylation marks are causes rather than consequences of lipid levels variation. Given that DNA methylation marks are potentially reversible, evidence for their role in the regulation of such a key enzyme is of great interest as it could lead to new therapeutic approaches (e.g., drug and/or diet supplementation) to modulate *CPT1A* expression. Finally, and of major importance for MWAS studies, peripheral whole blood DNA methylation marks were detected in an enzyme gene expressed in the liver and glandular tissues, suggesting that such marks could serve as surrogates for methylation at more closely-related effector cells, such as hepatocytes. The latter adds to the recent paper by Dick et al. (4) also supporting the value of peripheral whole blood DNA methylation marks as biomarkers of methylation in other tissues.

We thank Dr. Michael D. Wilson for his judicious comments on the manuscript and for the many fruitful discussions about epigenetic regulation.

REFERENCES

1. Murphy, T.M., and J. Mill. 2014. Epigenetics in health and disease: heralding the EWAS era. *The Lancet.* **S0140–6736**: 60269–60275.
2. Osório, J. 2014. Obesity: Looking at the epigenetic link between obesity and its consequences-the promise of EWAS. *Nat. Rev. Endocrinol.* **10:** 249.
3. Callaway, E. 2014. Epigenomics starts to make its mark. *Nature.* **508:** 22.
4. Dick, K.J., C.P. Nelson, L. Tsaprouni, J.K. Sandling, D. Aïssi, S. Wahl, E. Meduri, P.E. Morange, F. Gagnon, H. Grallert, et al. 2014. DNA methylation and body-mass index: a genome-wide analysis. *The Lancet.***S0140–6736**: 62674–62682.
5. Zeilinger, S., B. Kühnel, N. Klopp, H. Baurecht, A. Kleinschmidt, C. Gieger, S. Weidinger, E. Lattka, J. Adamski, A. Peters, et al. 2013. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS ONE.* **8:** e63812.
6. Bell, J.T., P-C. Tsai, T-P. Yang, R. Pidsley, J. Nisbet, D. Glass, M. Mangino, G. Zhai, F. Zhang, A. Valdes, et al. 2012. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet.* **8:** e1002629.
7. Dayeh, T., P. Volkov, S. Salö, E. Hall, E. Nilsson, A. H. Olsson, C. L. Kirkpatrick, C. B. Wollheim, L. Eliasson, T. Rönn, et al. 2014. Genome-wide DNA methylation analysis of human pancreatic islets from type 2 diabetic and non-diabetic donors identifies candidate genes that influence insulin secretion. *PLoS Genet.* **10:** e1004160.
8. Frazier-Wood, A. C., S. Aslibekyan, D. M. Absher, P. H. Hopkins, J. Sha, M. Y. Tsai, H. K. Tiwari, L. L. Waite, D. Zhi, and D. K. Arnett. 2014. Methylation at CPT1A locus is associated with lipoprotein subfraction profiles. *J. Lipid Res.* **55:** 1324–1330.
9. Rosenbaum, P. R. 2001. Replicating effects and biases. *Am. Stat.* **55:** 223–227.
10. Kraft, P., E. Zeggini, and J. P. A. Ioannidis. 2009. Replication in genome-wide association studies. *Stat. Sci.* **24:** 561–573.
11. Maksimovic, J., L. Gordon, and A. Oshlack. 2012. SWAN: Subset-quantile within array normalization for Illumina Infinium HumanMethylation450 BeadChips. *Genome Biol.* **13:** R44.
12. Oudot-Mellakh, T., W. Cohen, M. Germain, N. Saut, C. Kallel, D. Zelenika, M. Lathrop, D-A. Trégouët, and P-E. Morange. 2012. Genome wide association study for plasma levels of natural anticoagulant inhibitors and protein C anticoagulant pathway: the MARTHA project. *Br. J. Haematol.* **157:** 230–239.
13. Antoni, G., P-E. Morange, Y. Luo, N. Saut, G. Burgos, S. Heath, M. Germain, C. Biron-Andreani, J. F. Schved, G. Pernod, et al. 2010. A multi-stage multi-design strategy provides strong evidence that the BAI3 locus is associated with early-onset venous thromboembolism. *J. Thromb. Haemost.* **8:** 2671–2679.
14. Uhlen, M., P. Oksvold, L. Fagerberg, E. Lundberg, K. Jonasson, M. Forsberg, M. Zwahlen, C. Kampf, K. Wester, S. Hober, et al. 2010. Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **28:** 1248–1250.
15. Thurman, R.E., E. Rynes, R. Humbert, J. Vierstra, M.T. Maurano, E. Haugen, N.C. Sheffield, A.B Stergachis, H. Wang, B. Vernot, K. Garg, et al. 2012. The accessible chromatin landscape of the human genome. *Nature.* **489:** 75–82.
16. ENCODE Project Consortium, Bernstein, B.E., E. Birney, I. Dunham, E.D. Green, C. Gunter, and M. Snyder. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature.* **489:** 57–74.