

PROCEEDINGS

Open Access



# Integrative methylation score to identify epigenetic modifications associated with lipid changes resulting from fenofibrate treatment in families

Biqi Wang<sup>1</sup>, Anita L. DeStefano<sup>1</sup> and Honghuang Lin<sup>2,3\*</sup>

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

## Abstract

Epigenome-wide association studies (EWAS) have traditionally focused on the association test of single epigenetic markers with complex traits. However, it is possible that multiple cytosine-phosphate-guanine (CpG) sites at the same locus could jointly exert their effects on human traits. Therefore, a region-based test that combines multiple markers could be more powerful. We used 2 different region-based tests to investigate the association between changes in DNA methylation and drug response, including the median methylation level test (MMLT) and sequence kernel association test (SKAT). No genes were found to be significantly associated with the drug response (for triglycerides, the false discovery rate ranged from 0.855 to 0.999; for high-density lipoprotein cholesterol, and the false discovery rate ranged from 0.584 to 0.915). Further evidence is needed to explore potential application of gene-level methylation association analysis.

## Background

Considerable interindividual variabilities of responses have been observed in people who take lipid-lowering drugs, underscoring the importance of genetic variants to the drug response. For instance, common genetic polymorphisms in the apolipoprotein B (*APOB*) gene are associated with the capability of transferring triglyceride and cholesteryl esters during lipoprotein metabolism [1]. Moreover, DNA methylation could be altered by lipid-lowering drugs, which can interrupt methionine biosynthesis through multiple pathways [2]. It was found that the methylation level on *CPT1A* was significantly associated with triglyceride and high-density lipoprotein cholesterol (HDL-C) in European populations [3].

Das et al. performed an epigenome-wide association study (EWAS) to test the association of each individual cytosine-phosphate-guanine (CpG) site with lipid levels before and after 3 weeks of fenofibrate treatment [2]. No significant association was found. It is possible that multiple CpG sites on the same locus, each one with small effect, could jointly exert their effects, which could not be identified by the single-CpG site analysis [4]. It is thus interesting to explore region-based tests, which could use the information from multiple CpG sites and examine their joint effects on drug response. Such analysis can be a complement to current single-CpG association studies. The objective of this study is to investigate whether region-based changes of methylation profiles are associated with lipid changes induced by fenofibrate treatment in a family-based population.

\* Correspondence: [hhlin@bu.edu](mailto:hhlin@bu.edu)

<sup>2</sup>National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, 73 Mount Wayte Avenue, Framingham, MA 01702, USA

<sup>3</sup>Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, 72 E Concord St, B-616, Boston, MA 02118, USA

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



## Methods

### Data processing

The family data from the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study was obtained via GAW20. A detailed description of the GOLDN study design can be found elsewhere [3]. Briefly, the study is a family-based, open-label, and one arm clinical trial assessing the lipid-lowering-drug effect on human genomics after a treatment of fenofibrate (160 mg) for 3 weeks. Participants included in current analyses met these 3 criteria: (a) having qualified methylation data both before and after the 3 weeks of treatment; (b) having 1 or 2 lipid measurements at both pre- and posttreatment times; (c) having all the covariate data (such as age, sex, field center, and smoking status) at entry of the trial. There were 446 participants of European ancestry from 139 families that met these criteria.

DNA methylation was quantified by the Illumina Infinium HumanMethylation450K BeadChip using CD4+ T cells both pre- and posttreatment. The methylation level of each CpG site was estimated as a continuous variable ranging from 0 (not methylated) to 1 (fully methylated). Quality control of the methylation data was conducted by samples and probes separately. A multidimensional scaling plot was used to identify outliers among samples; no samples were excluded. For quality control of probes, the following filters were applied: exclusion of probes with single-nucleotide polymorphisms (SNPs) with minor allele frequency > 0.01 under the actual CpG sites or at the single base extension (17,297 probes), or cross-reacted to multiple targets (29,233 probes) [5]. After quality control, 446 samples with 423,180 probes remained. Beta-mixture quantile normalization was conducted to control biases from 2 different types of probes of the 450 K methylation array [6]. In addition, the first 20 principal components were calculated separately in pre- and posttreatment methylation data to control for potential confounders.

Lipid profiles were measured from blood after an overnight fast both before and after 3 weeks of intervention. Details about lipid measurement have been described in the previous GOLDN publication [2]. Gender, field center (Minnesota or Utah), and metabolic syndrome (yes or no) were binary variables. Metabolic syndrome was defined according to National Cholesterol Education Program (NCEP) guidelines from American Heart Association (AHA)/National Heart, Lung, and Blood Institute (NHLBI) 2004 meeting. Smoking status was categorized by never, past, and current smokers. Age and lipid were continuous variables in our analysis.

### Median methylation level test of genes

To better capture the effect of gene-level methylation, we defined a gene region by including all the CpG sites between 100 kb of the first transcript start site (TSS), and 100 kb downstream of the last TSS of the gene [7]. The

median methylation level of all the CpG sites within a region was calculated as a representation of the overall methylation pattern of the gene [7]. A total of 21,231 unique gene regions were identified, which encompassed more than 300,000 CpG sites. The TSSs were retrieved from GENCODE release 19, which was downloaded from the UCSC Genome Browser. Among the 21,231 unique genes, 894 were not among the UCSC known genes, and an additional 4602 genes did not contain any qualified CpG sites. Thus only median methylation levels for 15,735 genes were calculated for each sample. Both the pre- and posttreatment methylation matrices were composed of 446 samples.

The change of median gene region methylation was calculated as pretreatment median methylation minus posttreatment median methylation. A linear mixed-effect regression model was used to investigate the association between change in methylation median and natural logarithm transformed change of either triglyceride or HDL-C after 3 weeks of fenofibrate intervention. The outcome variable was natural log-transformed lipid change, while the change in median methylation was treated as an independent variable. Age, gender, field center, metabolic syndrome, smoking status, and pre- or post-principal components that were significantly associated with changes of lipid were treated as fixed effects, and the family structure was treated as a random effect. The analysis was conducted using the *lme4* function in the R *comex* package, and the significance threshold was defined as Benjamin-Hochberg false discovery rate  $q$  value equal to 0.05 [8].

**Table 1** Characteristics of 446 participants included in this analysis

Characteristics	N (%) or mean $\pm$ SD
Number of participants	446
Age (years)	47 $\pm$ 16
Women	230 (51.6)
Field centers	
Minnesota	201 (45.1)
Utah	245 (54.9)
Metabolic syndrome <sup>a</sup>	179 (40.1)
Smoking status	
Never smoker	316 (70.9)
Past smoker	100 (22.4)
Current smoker	30 (6.7)
Changes of triglycerides (mg/dL)	48.9 $\pm$ 54.6
Changes of high-density lipoprotein cholesterol (mg/dL)	-2.9 $\pm$ 5.1

<sup>a</sup>Metabolic syndrome was defined according to NCEP guidelines from AHA/NHLBI 2004 meeting

**Table 2** Top 10 genes associated with natural logarithm change in triglycerides

MMLT						SKAT <sup>a</sup>				
Genes	N	Effect size	SE	p Value	FDR q value	Genes	N	Q-statistic	p Value	FDR q value
<i>CSHL1</i>	3	1.29	0.35	2.02E-04	0.999	<i>RPL29</i>	1	117.39	1.59E-03	0.855
<i>C22orf39</i>	4	-1.89	0.56	7.74E-04	0.999	<i>GJB5</i>	2	104.51	2.95E-03	0.855
<i>C14orf166</i>	4	2.63	0.81	1.19E-03	0.999	<i>TBX20</i>	2	104.51	2.95E-03	0.855
<i>RPL29</i>	1	0.91	0.29	1.41E-03	0.999	<i>WDR92</i>	7	436.86	3.75E-03	0.855
<i>PRPH</i>	42	0.95	0.30	1.57E-03	0.999	<i>CPT1A</i>	6	439.50	3.92E-03	0.855
<i>GSG1</i>	2	-1.28	0.41	2.07E-03	0.999	<i>MIR1266</i>	2	51.56	5.29E-03	0.855
<i>LPIN2</i>	41	-0.95	0.32	2.55E-03	0.999	<i>TIMP4</i>	5	191.94	7.11E-03	0.855
<i>IP6K3</i>	124	-1.26	0.42	2.61E-03	0.999	<i>ZNF788</i>	5	191.94	7.11E-03	0.855
<i>PPIC</i>	53	-0.90	0.31	3.63E-03	0.999	<i>GSG1</i>	2	105.07	8.14E-03	0.855
<i>RHBDL3</i>	21	0.49	0.17	3.94E-03	0.999	<i>CRYGA</i>	3	115.23	1.06E-02	0.855

FDR, false discovery rate; N, number of CpG sites in the gene set

<sup>a</sup>The weight of SKAT all equaled 1**Sequence kernel association test of genes**

The sequence kernel association test (SKAT) method is one of the variance-component approaches that tests for associations by evaluating the distribution of genetic effects for a group of variants. It can be used to assess the association of a continuous or binary trait with a set of CpG sites in a similar fashion to a gene-level analysis in genome-wide genotyping or sequencing data [9]. For example, given a continuous trait such as the natural log-transformed change in triglyceride or HDL-C, the change of methylation level at single CpG sites would have a score statistic that was calculated by the marginal linear mixed-effect model adjusting for covariates and family structure. Multiple CpG sites could then be combined as a single CpG set. Considering such a set, SKAT for methylation sites uses a joint statistic that is an unweighted (all weights were equal to 1) sum of squares of the single CpG sites score statistics for the set. The joint statistic asymptotically follows a

mixture chi-square distribution and its *p* value can be computed analytically quickly. Here we used the same 15,735 gene regions as above to identify differentially methylated genes. The analysis was conducted using R package *seqMeta* and the significance level was defined as the same as in the median methylation analysis [10].

Sensitivity analysis was performed to address whether results would change when we applied more weights to the CpG sites that changed as a consequence of drug response. For example, the changes of methylation were treated as a continuous variable ranging from -1 to 1 (where the closer to 0, the less likely the methylation changed). We also computed the “methylation changing dosage,” which was defined as methylation changes plus 1, which ranged from 0 to 2. Thus a CpG site with most dosage around 1 meant that the methylation level remained the same (and methylation dosage frequency would close to 0.5), whereas a CpG

**Table 3** Top 10 genes associated with natural logarithm change in HDL-C

MMLT						SKAT <sup>a</sup>				
Genes	N	Effect size	SE	p Value	FDR q value	Genes	N	Q-statistic	p Value	FDR q value
<i>ZKSCAN8</i>	4	0.62	0.16	7.42E-05	0.584	<i>KRT222</i>	6	1271.43	5.97E-03	0.915
<i>ZSCAN16</i>	4	0.62	0.16	7.42E-05	0.584	<i>FAM211A</i>	5	840.88	1.00E-02	0.915
<i>MAGI2</i>	67	0.27	0.08	8.12E-04	0.999	<i>ZKSCAN8</i>	4	1095.82	1.13E-02	0.915
<i>CAPN3</i>	30	0.29	0.09	8.72E-04	0.999	<i>ZSCAN16</i>	4	1095.82	1.13E-02	0.915
<i>AK5</i>	37	0.27	0.08	8.78E-04	0.999	<i>HEATR5B</i>	1	552.38	1.28E-02	0.915
<i>KCTD14</i>	37	0.27	0.08	8.78E-04	0.999	<i>IFT27</i>	1	552.38	1.28E-02	0.915
<i>PTP4A2</i>	91	0.29	0.09	1.44E-03	0.999	<i>ZNF567</i>	1	552.38	1.28E-02	0.915
<i>GSTTP1</i>	28	-0.45	0.14	1.45E-03	0.999	<i>ZNF337</i>	4	983.95	1.28E-02	0.915
<i>FOXI1</i>	27	0.31	0.10	1.64E-03	0.999	<i>CD300E</i>	3	1014.54	1.40E-02	0.915
<i>VPS37A</i>	4	0.88	0.28	1.76E-03	0.999	<i>SORCS1</i>	9	1010.61	1.60E-02	0.915

FDR, false discovery rate; N, number of CpG sites in the gene set

<sup>a</sup>The weight of SKAT all equaled 1

with most dosages at 0 (or 2) (the methylation dosage frequency would be small, or even rare) meant that the methylation level actually changed between pre- and post-treatment. So, the default setting in *seqMeta* (which gave the rare alleles more weights) will add more weight to CpG sites with significant change of methylation levels, and less weight to CpG sites with little or change of methylation levels.

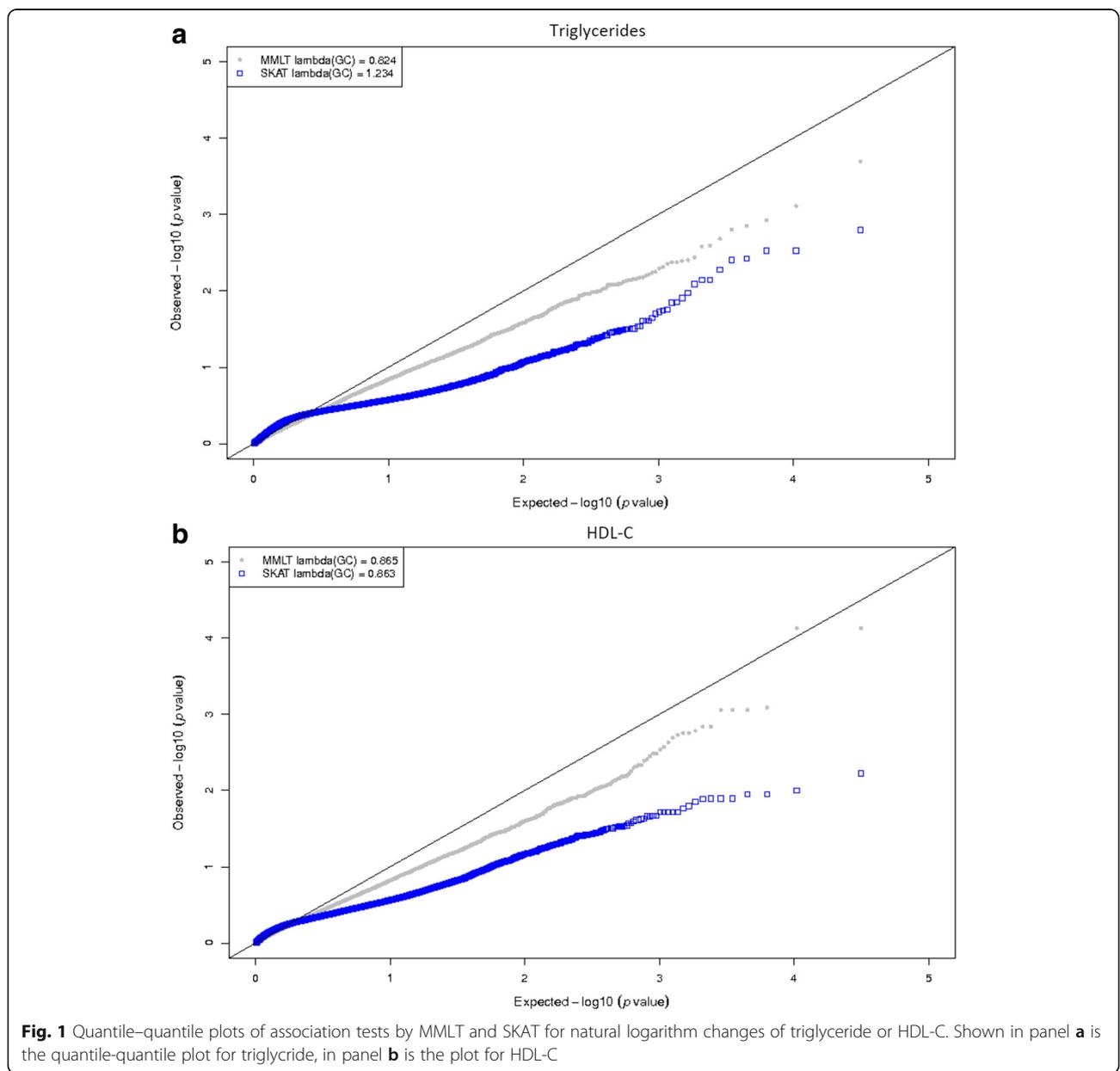
**Results**

Table 1 summarizes the characteristics of the 446 participants (mean age: 47.3 years). The number of women was slightly larger than the number of men. The triglyceride

level tended to decrease while HDL-C increased after 3 weeks of fenofibrate intervention.

Tables 2 and 3 show the top 10 genes associated with the natural logarithm changes of triglyceride and HDL-C, respectively. We did not find any genes significantly associated with triglyceride or HDL-C after correction for multiple testing. Interestingly, *CPT1A* was among one of the top triglyceride genes by SKAT, although it did not reach the predefined genome-wide significance.

Quantile-quantile plots were generated to identify whether the Type I error was inflated or deflated by median methylation level test (MMLT) and SKAT (Fig. 1). The plot of SKAT in natural logarithm changes in triglyceride show some inflation with a genomic control factor of 1.234, while



others were deflated. Further plotted by the numbers of CpG sites in the gene region (which was categorized by the quartiles of CpG sites) showed that for SKAT in natural logarithm changes in triglyceride, the plot seemed to be more inflated in the gene regions with more CpG sites (Fig. 2).

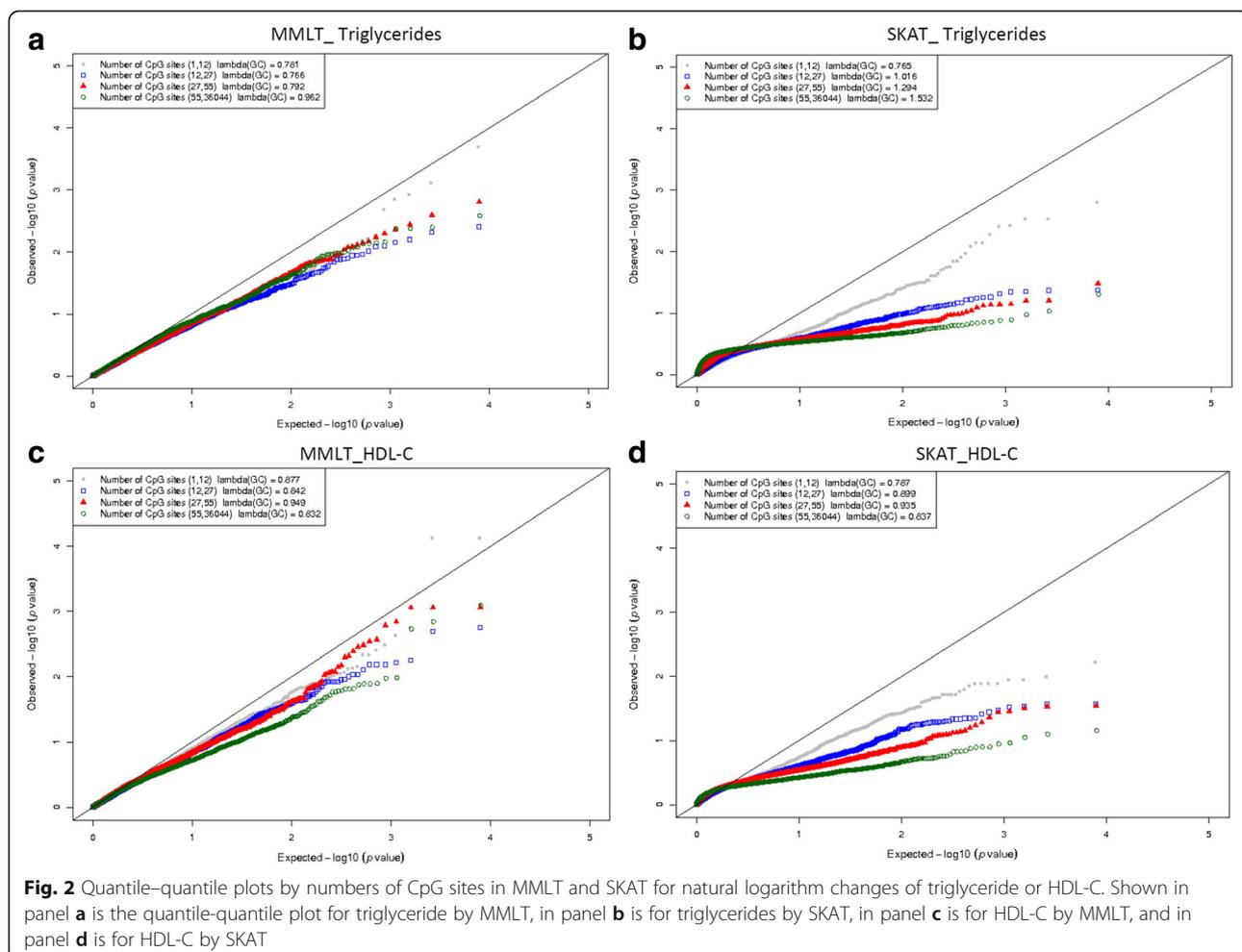
Sensitivity analysis showed that no genes were significantly associated with natural logarithm changes of triglycerides or HDL-C (see Additional file 1: Table S1 and Table S2), but with relatively lower Type I error than the unweighted SKAT method (see Additional file 1: Figures S1 and S2).

## Discussion

In the present study, we used 2 region-based association tests to examine epigenetic modifications associated with changes in lipid levels induced by fenofibrate in a family-based population. No genes were significantly associated with the change in triglyceride or HDL-C after 3 weeks of intervention, similar to the single CpG test.

Two region-based methods were evaluated in our study. The Type I error seemed deflated in testing by MMLT across lipids changes. When the number of CpG sites increased, the Type I error slightly increased but still remained deflated. Our results suggest that MMLT needs a relatively large number of changing CpG sites to capture the overall methylation profile of gene. The method has less power if only a small number of CpG sites changed as it only calculated the median level. In contrast, the Type I error inflated along with the number of CpG sites in the gene when tested by SKAT, suggesting that noises might be introduced by including too many irrelevant CpG sites. Moreover, the Type I error decreased after assigning more weight to the changed CpG sites, which also indicated unchanged CpG sites might create “noise” and dilute the effect of changed CpG sites.

Although neither method identified gene-level significant methylation profiles associated with lipid changes, SKAT identified *CPT1A* as a potentially interesting gene associated with the changes of triglyceride. *CPT1A* is a gene responded to a key enzyme in the carnitine-dependent transport across



the mitochondrial inner membrane and its deficiency results in a decreased rate of fatty acid beta-oxidation. The MMLT method has lower power compared to SKAT when the directions of effects are different [11]. For example, a gene could contain some CpG sites that were positively associated with the drug response, and some CpG sites that were negatively associated with the drug response. The overall signal could be attenuated because the 2 groups of CpG sites have opposite effects. In contrast, SKAT is particularly powerful for identifying significant gene sets in such a case, because it takes into account of the direction of effects in the model [12].

We acknowledge several limitations of our study. Given the limited access to the GOLDN study, we might have missed some important factors related to batch effects or potential confounders. In addition, 3 weeks of intervention might not be long enough to observe significant change in DNA methylation. Additional observations are needed to observe methylation trajectory [13].

## Conclusions

We found limited evidence for region-based analysis in identifying methylation genes associated with drug response. Longer and multipoint observations could provide additional insights into the effects of pharmaceutical interventions.

## Additional file

**Additional file 1: Table S1.** Top 10 genes associated with change in triglycerides. **Table S2.** Top 10 genes associated with change in HDLC. **Figure S1.** Quantile-quantile plots of association tests by MMLT and weighted SKAT for natural logarithm changes of triglyceride or HDLC. The weight of SKAT was the algorithm of minor allele frequency. **Figure S2.** Quantile-quantile plots by numbers of CpGs in MMLT and weighted SKAT for for natural logarithm changes of triglyceride or HDLC. (PDF 600 kb)

## Acknowledgements

We acknowledge Samantha Lent and Achilleas Pitsillides for conducting parts of the methylation quality control and for support of programming improvements.

## Funding

Publication of this article was supported by NIH R01 GM031575.

## 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 license 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

HL conceived the study design and analysis plan. BW conducted the analysis and drafted the manuscript. HL and AD made critical reviews for the manuscript. All authors have 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>Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, Boston, MA 02118, USA. <sup>2</sup>National Heart, Lung, and Blood Institute's and Boston University's Framingham Heart Study, 73 Mount Wayte Avenue, Framingham, MA 01702, USA. <sup>3</sup>Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, 72 E Concord St, B-616, Boston, MA 02118, USA.

Published: 17 September 2018

## References

1. Wojczynski MK, Gao G, Borecki I, Hopkins PN, Parnell L, Lai C-Q, Ordovas JM, Chung BH, Arnett DK. ApoB genetic variants modify the response to fenofibrate: a GOLDN study. *J Lipid Res.* 2010;52(11):520–6.
2. Das M, Irvin MR, Sha J, Aslibekyan S, Hidalgo B, Perry RT, Zhi D, Tiwari HK, Absher D, Ordovas JM, 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.
3. Irvin MR, Zhi D, Joehanes R, Mendelson M, Aslibekyan S, Claas SA, Thibeault 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.
4. Lee S, Abecasis GR, Boehnke M, Lin X. Rare-variant association analysis: study designs and statistical tests. *Am J Hum Genet.* 2014;95(1):5–23.
5. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* 2013;8(2):203–9.
6. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalisation method for correcting probe design bias in Illumina Infinium 450k DNA methylation data. *Bioinformatics.* 2013;29(2):189–96.
7. Fleischer T, Frigessi A, Johnson KC, Edvardsen H, Touleimat N, Klajic J, Riis ML, Haakensen VD, Wärnberg F, Naume B, et al. Genome-wide DNA methylation profiles in progression to in situ and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol.* 2014;15(8):435.
8. Therneau T: *coxme: Mixed effects cox models.* R package version 2.2–5. <https://CRAN.R-project.org/package=coxme>. 2015.
9. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet.* 2011;89(1):82–93.
10. Voorman AA, Brody J, Chen H, Lumley T, Davis B: *seqMeta: meta-analysis of region-based tests of rare DNA variants.* R package version 1.6.7. <https://CRAN.R-project.org/package=seqMeta>. 2016.
11. Tian S, Bertelsmann K, Yu L, Sun S. DNA methylation heterogeneity patterns in breast cancer cell lines. *Cancer Inform.* 2016;15 (Suppl 4):1–9.
12. Zhang Q, Zhao Y, Zhang R, Wei Y, Yi H, Shao F, Chen F. A comparative study of five association tests based on CpG set for epigenome-wide association studies. *PLoS One.* 2016;11(6):e0156895.
13. Kananen L, Marttila S, Nevalainen T, Kummola L, Junttila I, Mononen N, Kähönen M, Raitakari OT, Hervonen A, Jylhä M, et al. The trajectory of the blood DNA methylome ageing rate is largely set before adulthood: evidence from two longitudinal studies. *Age (Dordr).* 2016;38(3):65.