

2018

# The challenge of detecting genotype-by-methylation interaction: GAW20

Mariza de Andrade

*Mayo Clinic*

E. Warwick Daw

*Washington University School of Medicine in St. Louis*

Aldi T. Kraja

*Washington University School of Medicine in St. Louis*

Virginia Fisher

*Boston University*

Lan Wang

*Boston University*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open\\_access\\_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

---

## Recommended Citation

de Andrade, Mariza; Daw, E. Warwick; Kraja, Aldi T.; Fisher, Virginia; Wang, Lan; Hu, Ke; Li, Jing; Romanescu, Razvan; Veenstra, Jenna; Sun, Rui; Weng, Haoyi; and Zhou, Wenda, "The challenge of detecting genotype-by-methylation interaction: GAW20." *BMC Genetics*.19,Suppl 1. 81. (2018).

[https://digitalcommons.wustl.edu/open\\_access\\_pubs/7172](https://digitalcommons.wustl.edu/open_access_pubs/7172)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact [engeszer@wustl.edu](mailto:engeszer@wustl.edu).

---

**Authors**

Mariza de Andrade, E. Warwick Daw, Aldi T. Kraja, Virginia Fisher, Lan Wang, Ke Hu, Jing Li, Razvan Romanescu, Jenna Veenstra, Rui Sun, Haoyi Weng, and Wenda Zhou

PROCEEDINGS

Open Access



# The challenge of detecting genotype-by-methylation interaction: GAW20

Mariza de Andrade<sup>1\*†</sup>, E. Warwick Daw<sup>2†</sup>, Aldi T. Kraja<sup>2†</sup>, Virginia Fisher<sup>3†</sup>, Lan Wang<sup>3†</sup>, Ke Hu<sup>4</sup>, Jing Li<sup>4</sup>, Razvan Romanescu<sup>5</sup>, Jenna Veenstra<sup>6,7</sup>, Rui Sun<sup>8,9†</sup>, Haoyi Weng<sup>8,9†</sup> and Wenda Zhou<sup>10</sup>

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

## Abstract

**Background:** GAW20 working group 5 brought together researchers who contributed 7 papers with the aim of evaluating methods to detect genetic by epigenetic interactions. GAW20 distributed real data from the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) study, including single-nucleotide polymorphism (SNP) markers, methylation (cytosine-phosphate-guanine [CpG]) markers, and phenotype information on up to 995 individuals. In addition, a simulated data set based on the real data was provided.

**Results:** The 7 contributed papers analyzed these data sets with a number of different statistical methods, including generalized linear mixed models, mediation analysis, machine learning, W-test, and sparsity-inducing regularized regression. These methods generally appeared to perform well. Several papers confirmed a number of causative SNPs in either the large number of simulation sets or the real data on chromosome 11. Findings were also reported for different SNPs, CpG sites, and SNP–CpG site interaction pairs.

**Conclusions:** In the simulation (200 replications), power appeared generally good for large interaction effects, but smaller effects will require larger studies or consortium collaboration for realizing a sufficient power.

**Keywords:** Methylation, Interaction, Genome wide association, Region based association, Candidate gene association, Mediation analysis, Multi-level Gaussian model, LASSO, Adaptive W-test, Regression and random forest trees

## Background

Genes do not act in isolation: the expression of gene-coded proteins is affected by many factors, including other genes, environment, and epigenetics. These interactions can complicate the identification of genes related to a particular trait or disease because some genes with an important effect through interaction can have a small marginal effect. The usefulness of interaction models for identifying causation based on disease and risk factors is demonstrated by Ruth Othman's 1996 seminal paper [1]. As a result of recognition of the importance of interactions, a variety of methods that incorporate it were

developed, especially in the case of gene–environment interaction. These methods include logistic [2] and generalized linear models [3], and machine learning [4].

This paper surveys the findings of the papers presented at GAW20 in group 5: Gene by Methylation Interactions. Even though there are issues common to all interaction approaches, each interaction type has its own specific features. In the case of gene by epigenetic interactions, all the mechanisms of epigenetics are not yet fully understood. However, it is generally accepted that methylation of a DNA strand causes loci on the methylated span to be turned “off.” The generally available measure of methylation is the percentage of copies of a DNA strand site that are methylated (generally in a single individual). The expectation is that individuals with higher levels of methylation at a site will have a lower level of gene expression for that site. Incorporating this expectation into a statistical analysis can simplify it, as

\* Correspondence: [mandrade@mayo.edu](mailto:mandrade@mayo.edu)

†Mariza de Andrade, E. Warwick Daw, Aldi T. Kraja, Virginia Fisher, Lan Wang, Rui Sun and Haoyi Weng contributed equally to this work.

<sup>1</sup>Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, 200 First St. SW, Rochester, MN 55905, USA  
Full list of author information is available at the end of the article



compared to, for example, an analysis of an environmental factor for which the direction and mechanism of the effect are completely unknown. As our understanding of the relationships between epigenetics and genes, environment, and ancestral epigenetics improves, we will be able to further refine the statistical tests for gene by epigenetic interaction. The papers included in this group represent a wide range of approaches facilitated by the current understanding of gene by epigenetic interaction.

The data for GAW20 was provided by the Genetics of Lipid Lowering Drugs and Diet Network (GOLDN) project [5, 6]. GAW20 also provided simulated data with similar characteristics to the real data. While each group gave the data a different treatment, using the same underlying data facilitated comparison between the methods of the different groups. Most of the papers included in this group focused on methods to directly detect gene by epigenetic interaction effects, but some also explored related issues. The overall aim was to analyze the GAW20 data via different methods and compare their usefulness in the discovery of new/previously reported (with real data) or of expected findings (simulated data).

## Methods

The GAW20 group 5 applied methods to both available types of data: real and simulated [7]. The real data set consisted of phenotypes of high-density lipoprotein cholesterol, triglyceride (TG) levels, and metabolic syndrome diagnosis, before and after treatment with fenofibrate, genome-wide methylation pre- and posttreatment, and dense genome-wide single-nucleotide polymorphisms (SNPs) from the GOLDN study. The simulated data set, which was based on the real data, included 200 replicates of posttreatment TG and methylation levels. The subjects in the GOLDN study, who self-reported to be white, were recruited at the Minneapolis, MN, and Salt Lake City, UT Family Heart Study centers. The GOLDN study was designed with 2 arms: a high-fat meal challenge and fenofibrate treatment for 3 weeks to identify genetic determinants of lipid response to 2 interventions. A total of 1053 individuals met all eligibility requirements, while the data in use for lipids at visits 1 and 2, at the start of the trial, and at visits 3 and 4, at the end of the trial, that have genotype and methylation were smaller [8–12]. The GAW20 simulated data [7] followed the structure of real data, by simulating the effects of a lipid-lowering medication in interaction with methylation by affecting TG levels. A number of major variants, as well as a large number of background variants, at the genotype and methylation levels were causative of changing TG levels at visits 3 and 4. Three papers from the group studied changes on TG (Daw et al., Fisher et al., and Sun et al). Sun et al. differ from the other two in that they use percent change and dichotomized if the treatment was effective or not. Four

papers look at the average TG levels either  $TG_{pre}$  or  $TG_{post}$  (Daw et al., Romanescu et al., Veenstra et al., and Zhao and Lo). One paper examined cytosine-phosphate-guanine (CpG) site variation (Hu and Li).

Most papers examined real GOLDN (CpG site, SNP, and/or TG) data. One paper analyzed the simulated data in addition to real data (Sun et al), and one paper studied simulated data only (Daw et al.; Table 1). The methods employed for detecting epigenetic by genetic interactions/main effects included generalized linear model (GLM), mixed model, regression tree model, random forest [13], mediation interaction model [14], gene-based test [15], W-test [16], candidate gene approach [16] and LASSO (least absolute shrinkage and selection operator)-type regression [17] techniques. Additionally, Gaussian mixture models were applied to examine the relationship between methylation and genotype. For a full description of each method, see the appropriate paper.

## Results

Several methods were used to detect SNP by methylation interactions using either simulated or real data. Some of the real data papers conducted a genome-wide association study, whereas others focused on either candidate genes or chromosome 11 where a previous finding in the provided data was reported (see Table 1). Consequently, for the real data papers we focused on comparing results on chromosome 11 (Table 2). For assessing results in the simulated data, we can simply compare to the known answers.

In Daw et al. [13], GLMs, mixed models, regression trees, and random forest models were applied to 200 replications of simulated data. PROC GLM and PROC MIXED of SAS were used for the linear models, taking into account the family structure as a random effect. Four models were used: model 1.a included the intercept, the main effects, CpG sites and SNPs, and their interaction effect on TG post-medication-treatment plus a vector of covariates (age, age<sup>2</sup>, sex, and the average of log[TG<sub>1</sub>] and log[TG<sub>2</sub>] before treatment); model 1.b included only the interaction effect on TG post-medication-treatment plus the vector of covariates; model 2.a included as the response variable the change of TG (posttreatment-pretreatment) including the main effects, SNP and the methylation difference between visits 4 and 2, their interaction and covariates (age, age<sup>2</sup>, sex); model 2.b was a reduced version of model 2.a that included only interactions. The models 1.a and 1.b were fitted using PROC GLM and PROC MIXED. Models 2.a and 2.b were fitted using only PROC MIXED. Regression trees were applied using SAS (v. 9.4) and SAS Data Mining (v. 14.1), as well as random forest in Python from the SciKit-Learn package, to predict outcome variables by averaging the predictions of a large number of uncorrelated regression trees. The authors tested for sensitivity using causative SNP-CpG

**Table 1** Paper, data used, and methods implemented for capturing genetic, epigenetic, and associations with phenotypes

No	Paper	Data used	GAW20 simulations Answers known	Methods	Specific data and modeling	Response variable
1	Daw et al	Simulated	YES	GLM Linear mixed models Regression trees Random forest	Simulated data (200 replications) Simulated effect sites + noise	$\text{ave}(\log(TG_{\text{post}}))$ $\text{ave}(\log(TG_{\text{post}})) - \text{ave}(\log(TG_{\text{pre}}))$
2	Fisher et al	Real data	N/A	Mediation interaction model	GWAS EWAS pairs with $P_1 < 10^{-3}$ and $P_2 < 0.05$ The effect of change in methylation on change in TG differs by genotype Power: simulation with 500 replications	$\log(TG_{\text{post}}) - \log(TG_{\text{pre}})$
3	Hu and Li	Real data	N/A	Gaussian mixture-model clustering Gaphunter a thresholding-based method	CpG x SNP (GW) multimodal CpG sites Association of CpG identified with overlapping SNP Association with nearby SNPs	Pretreatment CpG Posttreatment CpG
4	Romanescu et al	Real data	N/A	Regional-based test	Single CpG association (CH11) Regional SNP PC association (CH11) Regional CpG PC association (CH11) Regional SNP + regional CpG association (CH11)	$\text{ave}(\ln(TG_{\text{pre}}))$
5	Sun et al	Both	YES	Adaptive W-test with 2000 repetitions to get Type I error	In both analysis main and interaction associations Simulation: candidate pairs and noise pairs Real data: (CH11)	$\text{delta}(TG\%) = 1$ , if $\text{delta}(TG\%) > 30\%$ = 1, otherwise 0 $\text{delta}(TG\%) = (TG_{\text{pre}} - TG_{\text{post}})/TG_{\text{pre}}$
6	Veenstra et al	Real data	N/A	LmeKin of coxme, lm in R	(1) SNP association + (2) CpG association (1) + (2) + CpG x SNP	Ave(TG1 + TG2) baseline HDL
7	Zhou and Lo	Real data	N/A	Group LASSO-regression, kinship matrix, stability selection	CpG association (GW) SNP association (GW) (1) + (2) + CpG x SNP (CH11)	Ave(TG)

Abbreviations: CH11 chromosome 11, CpG cytosine-phosphate-guanine, EWAS epigenome-wide association study, GLM generalized linear model, GW genome-wide, GWAS genome-wide association study, HDL high-density lipoprotein cholesterol, PC principal components; pre, pretreatment; post, posttreatment, SNP single-nucleotide polymorphism, TG triglyceride

**Table 2** Chromosome 11 findings from real data analyses

No	Approximate position (MBs)	Gene(s)	Papers	Legend No	Paper
1	0.6	<i>C11orf63</i>	5	1	Daw et al
2	1.9	<i>TNNT3</i>	5	2	Fisher et al
3	5.6	<i>TRIM5,TRIM6-TRIM34,TRIM3</i>	5	3	Hu and Li
4	15.7	<i>rs900865</i>	5	4	Romanescu et al
5	19.4	<i>NAV2</i>	5	5	Sun et al
6	26.5	<i>rs4347345</i>	5	6	Veenstra et al
7	30.4	<i>MPPED2</i>	5	7	Zhou and Lo
8	40.1	<i>LLRC4C</i>	4		
9	57.3	<i>SLC43A1</i>	4		
10	61.6	<i>FADS1, FADS2, FADS3</i>	6		
11	68.6	<i>CPT1A<sup>a</sup></i>	4,7		
12	77.1	<i>CUCY2E</i>	5		
13	102.8	<i>LOC100128088</i>	5		
14	114.1	<i>ZBTB16</i>	5		
15	116.7	<i>APOA5, APOA1, BUD13, AP006216.5<sup>a</sup></i>	4,6,7		
16	133.4	<i>B3GAT1</i>	5		

<sup>a</sup>Genes on chromosome 11 that were identified in more than one paper from the GAW group5

site pairs and noncausative SNP-CpG site pairs to estimate the empirical null. They identified reasonable power to detect the main causative loci, depending on sample size and strength of the effects at the SNP and CpG sites.

Sun et al. [16] worked with both simulated data and real data. They analyzed only the 84th replication of GAW20 simulated data. This replicate was recommended as the one to use for comparing results with other teams when not working with all 200 replicates. Their *W*-test [18] checks for pairwise epistasis based on the null hypothesis that the joint distribution of a set of SNPs has no difference in the cases versus controls. They tested the gene-methylation epistasis in subjects' posttreatment using an additive model. Individuals' treatment response was dichotomized based on change from pre-TG to post-TG levels. A change greater than 30% [19] was defined as effective; a lesser change was defined as ineffective. For the *W*-test, methylation was dichotomized into high and low groups. Thus, the SNP-CpG site pair forms a genetic combination of 6 categories. In addition, Sun et al. examined logistic regression models with both dichotomized and continuous CpG sites [16]. In their application of 3 tests to 1 simulated replicate, they examined 10 sites: 5 "causative" and 5 "noise." None of the "noise" sites were significant for any of the tests. Three of the 5 "causative" sites were significant for all 3 tests, and 1 "causative" site was marginally significant for *W*-test and for the 2 logistic regression tests. The fifth site was not detected using any of the 3 methods. In real data they focused on chromosome 11, and applied

only their *W*-test. Their top 15 hits had *p* values from  $7.5 \times 10^{-6}$  to  $7.0 \times 10^{-5}$ . None of these 15 were reported by other papers in our group.

Mediation analysis was implemented by Fisher et al. [14], who investigated change in pretreatment to post-treatment TG levels as outcome by considering the hypothesis that genotype affects change in TG level by means of its effects on methylation. Thus, the effect of change in methylation on change in TG differs by genotype. For their analysis, they used a real data sample of 406 individuals with complete genotype, methylation, TG levels, and covariate data, performing estimation with the *coxme* R package [14]. They conducted a genome study, but none of their reported hits were on chromosome 11. They also had no hits in common with Veenstra et al., who also conducted genome-wide analysis. Fisher et al.'s power calculations suggest that the available data were underpowered for their method.

Hu and Li investigated the relation between genotype and methylation levels pretreatment and posttreatment via Gaussian mixture-model clustering. First, they identified CpG sites with multimodal methylation level distributions (mmCpGs). Subsequently, they identified SNPs that segregate with mmCpG modes to identify genes that may play a role in methylation levels. The authors identified 3785 and 3847 mmCpGs in pretreatment and posttreatment data sets [20]. Of these, 453 directly overlap with SNPs. Some show strong correlation between genotype and methylation levels, which suggests a relation between

epigenetics and genetics that may be modeled in the statistical analysis.

Romanescu et al. applied a region-based test to identify CpG site regions associated with TG levels. They applied 4 statistical models: a single-site test for CpG site association and 3 regional tests. In regions defined by GENCODE annotation, they conducted a principal component analysis for SNP genotypes and for methylation levels. Then they conducted association tests incorporating (a) the SNP-principal component, (b) the CpG sites principal component, and (c) both. The authors used chromosome 11 real data. Their mixed linear models account for family structure, while incorporating the method proposed by Gauderman [21] using principal components on SNPs and CpG site regions [15]. They identified 4 regions on chromosome 11, including 2 that were identified in other studies: the *CPT1A* and the *APOA5* regions.

Veenstra et al. employed a candidate gene and a genome-wide approach to identify genetic and epigenetic effects on baseline TG. To remove any confounding effects, they adjusted for 6 covariates: age, sex, smoking, metabolic syndrome diagnosis (msdx), fasting time at baseline, and high-density lipoprotein cholesterol. They used the *lmekin* function from the *coxme* R package [22]. Furthermore, they analyzed the residual TG, after adjusting for these covariates, by modeling main effects of SNPs, CpG sites, and their interactions for approximately 700,000 pairs using an F-test. Additionally, they carried out the same procedure on 18 gene regions containing 423 unique SNP-CpG sites. Their genome-wide analysis identified no results on chromosome 11. In contrast, their candidate gene analysis, which included 7 genes in 2 regions on chromosome 11, found them as significant in both regions. One of the regions, *APOA5*, was also found in other papers of this group.

Zhou and Lo applied LASSO (least absolute shrinkage and selection operator)-type regression models to identify SNP-CpG site interactions [17]. The methods included the group LASSO approach for categorical variables and interactions [23]. These coauthors analyzed whole-genome CpG site association and whole-genome SNP association, as well as interaction modeling with main effects on chromosome 11 only. Zhou and Lo implemented a stability selection method, which quantified the uncertainty in variable selection, thus controlling the family-wise error rate. They identified 2 regions on chromosome 11 that were also identified by other papers in this group, including *CPT1A*.

## Discussion

The contributing papers in the Gene by Methylation Interaction group employed 10 major methods to capture such interaction. Their results suggest that a variety

of methods can be employed to productively analyze combined genetic and epigenetic data. The choice of response variable was mostly  $\text{pre}(\log(\text{TG}))$ ; 2 papers looked at absolute change in  $\log(\text{TG})$ , with 1 paper also examining  $\text{post}(\log(\text{TG}))$  and 1 dichotomizing percent change in TG. In addition, 1 study used CpG sites as a response variable. Papers 4, 6, and 7 all analyzed baseline TG and obtained a set of overlapping results on chromosome 11 (see Table 1). One of these results was *CPT1A*, which was previously reported by the GOLDN study [6]. This overlap of results lends credence to the accuracy of the results. Furthermore, the 2 regions of overlap, *CPT1A* and the genes in the *APOA5* region, are biologically involved in lipid metabolism. The other papers that analyzed real TG data (2 and 5; see Table 1), analyzed measures in TG between pretreatment and posttreatment. This difference in phenotype may explain why the 2 changes in real TG papers do not replicate the findings of the TG-baseline papers. Furthermore, these 2 last papers use different measures of change in TG, which might explain the lack of replication between the two. Although we have summarized the reported results in the real data papers, each paper used slightly different criteria for reporting results.

For evaluating power, the simulated data, in which the answers are known, are more suited than the real data. The GAW20 GOLDN real data and GAW20 simulated data were similar in sample size, less than 900 individuals in baseline and less than 700 individuals in visits 3 and 4, thus making for samples that limit the power of discovery of causative variants. The Daw et al. paper shows that in the GAW20 simulated data there was a reasonable power to detect causative loci, but detection of background causative rare alleles with very small contributions is difficult with this sample size. Fisher et al. report the need for a sample size of approximately 19.5 K to be able to detect significant indirect effects. Thus, it is reasonable to expect that methods to detect gene by methylation interaction will yield better results in many large studies.

The analyses used employed various types of linear models. The Daw et al., Fisher et al., Sun et al., Veenstra et al., and Zhou and Lo papers all include models with SNP by methylation interaction, while the paper by Romanescu et al. includes a gene-region model that comprises effects for the principle components of both SNP genotype and methylation within a single gene-region. These models appear reasonable, given the current understanding of genetics and epigenetics. However, the paper of Hu and Li found some correlations between measured CpG site levels and SNP genotypes, which may indicate additional relationships that should be incorporated as our understanding improves. The methods applied in these papers have clear utility, but improvements may be possible as new

relationships between genetic and epigenetic data are discovered and incorporated into the models.

## Conclusions

The SNP–methylation interaction group provides very creative, interesting approaches for application to the GAW20 data sets, using either the GOLDN or simulated data. Several papers found a number of causative SNPs using the simulation set and the previous results on chromosome 11 were confirmed. Other papers identified different SNPs, CpG sites, and SNP–CpG interaction sites. Power appeared generally good for large interaction effects for these methods in a sample similar in size to the one provided (approximately 500 to 1000 individuals), but smaller effects appear to require larger studies or consortiums to have sufficient power.

## Abbreviations

CpG: Cytosine-phosphate-guanine; GAW: Genetic Analysis Workshop; GLM: Generalized linear model; GOLDN: Genetics of Lipid Lowering Drugs and Diet Network; LASSO: Least absolute shrinkage and selection operator; mmCpGs: CpGs with multimodal methylation level distributions; msdx: Metabolic syndrome diagnosis; SNP: Single-nucleotide polymorphism; TG: Triglyceride

## Funding

Publication of the proceedings of Genetic Analysis Workshop 20 was supported by National Institutes of Health grant 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 Genetics* Volume 19 Supplement 1, 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://bmccgenet.biomedcentral.com/articles/supplements/volume-19-supplement-1>.

## Authors' contributions

MdA wrote the first draft of the paper with the contribution of all the authors. EWJ and ATK reviewed the paper and included additional information about the original data and results. This paper was contributed equally for all the authors, without them this paper would not be written. All authors have read and approved the 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>Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, 200 First St. SW, Rochester, MN 55905, USA.

<sup>2</sup>Division of Statistical Genomics, Department of Genetics, Center for

Genome Sciences and Systems Biology, Washington University School of Medicine, 660 Euclid Ave, Saint Louis, MO 63110, USA. <sup>3</sup>Department of Biostatistics, Boston University School of Public Health, Boston, 715 Albany St, Boston, MA 02118, USA. <sup>4</sup>Department of Electrical Engineering and Computer Science, Case Western Reserve University, 10900 Euclid Ave, Cleveland, OH 44106, USA. <sup>5</sup>Lunenfeld-Tanenbaum Research Institute, Sinai Health System, University of Toronto, 600 University Ave, Toronto, ON M5G 1X5, Canada. <sup>6</sup>Department of Biology, Dordt College, 498 4th Ave. NE, Sioux Center, IA 51250, USA. <sup>7</sup>Department of Mathematics and Statistics, Dordt College, 498 4th Ave. NE, Sioux Center, IA 51250, USA. <sup>8</sup>Division of Biostatistics, Centre for Clinical Research and Biostatistics, JC School of Public Health and Primary Care, the Chinese University of Hong Kong, Shatin, N.T, Hong Kong, SAR, China. <sup>9</sup>CUHK Shenzhen Research Institute, Shenzhen, China. <sup>10</sup>Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York, NY 10027, USA.

Published: 17 September 2018

## References

- Ottman R. Gene-environment interaction: definitions and study designs. *Prev Med.* 1996;25:764–70.
- Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered.* 2007; 63:111–9.
- Lin X, Lee S, Christiani DC. Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics.* 2013;14:667–81.
- Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics.* 2003;19:376–82.
- Aslibekyan S, Goodarzi MO, Frazier-Wood AC, Yan X, Irvin MR, Kim E, Tiwari HK, Guo X, Straka RJ, Taylor KD, et al. Variants identified in a GWAS meta-analysis for blood lipids are associated with the lipid response to fenofibrate. *PLoS One.* 2012;7:e48663.
- 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;130:565–72.
- Aslibekyan S, Almasy L, Province MA, Absher DM, Arnett DK. Data for GAW20: genome-wide DNA Seq variation and epigenome-wide DNA methylation before and after fenofibrate treatment in a family study of metabolic phenotypes. *BMC Proc.* 2018;12(Suppl 9). <https://doi.org/10.1186/s12919-018-0114-0>.
- Yusuf N, Hidalgo B, Irvin MR, Sha D, Zhi D, Tiwari HK, Absher D, Arnett DK, Aslibekyan SW. An epigenome-wide association study of inflammatory response to fenofibrate in the genetics of lipid lowering drugs and diet network. *Pharmacogenomics.* 2017;18:1333–41.
- Sayols-Baixeras S, Irvin MR, Arnett DK, Elosua R, Aslibekyan SW. Epigenetics of lipid phenotypes. *Curr Cardiovasc Risk Rep.* 2016;10(10)
- Aslibekyan S, Irvin MR, Hidalgo B, Perry RT, Jevrajah EJ, Garcia E, Shalaurova I, Hopkins PN, Province MA, Tiwari HK, et al. Genome- and CD4+ T-cell methylome-wide association study of circulating trimethylamine-N-oxide in the genetics of lipid lowering drugs and diet network (GOLDN). *J Nutr Intermed Metab.* 2017;8:1–7.
- Aslibekyan S, Do AN, Xu H, Li S, Irvin MR, Zhi D, Tiwari HK, Absher DM, Shuldiner AR, Zhang T, et al. CPT1A methylation is associated with plasma adiponectin. *Nutr Metab Cardiovasc Dis.* 2017;27:225–33.
- Lai CQ, Wojczynski MK, Parnell LD, Hidalgo BA, Irvin MR, Aslibekyan S, Province MA, Absher DM, Arnett DK, Ordovas JM. Epigenome-wide association study of triglyceride postprandial responses to a high-fat dietary challenge. *J Lipid Res.* 2016;57:2200–7.
- Daw EW, Hicks J, Lenzini P, Lin SJ, Wang J, Williams PA, An P, Province MA, Kraja AT. Methods for detecting methylation by SNP interaction in GAW20 SIMULATION. *BMC Proc.* 2018;12(Suppl 9). <https://doi.org/10.1186/s12919-018-0140-y>.
- Fisher V, Wang L, Deng X, Sarnowski C, Cupples LA, Ching-Ti L. Do changes in DNA methylation mediate or interact with SNP variation? A pharmacoeconomic analysis. *BMC Genet.* 2018;19(Suppl 1). <https://doi.org/10.1186/s12863-018-0635-6>.
- Romanescu R, Espin-Garcia O, Ma J, Bull SB. Integrating epigenetic, genetic, and phenotypic data to uncover gene-region associations with triglycerides in the GOLDN lipid-lowering drug study. *BMC Proc.* 2018;12(Suppl 9). <https://doi.org/10.1186/s12919-018-0142-9>.



16. Sun R, Weng H, Men R, Xia X, Chong KC, Wu WKK, Zee BC, Wang MH. Gene-methylation epistatic analyses via the W-test identifies enriched signals of neuronal genes in patients undergoing lipid-control treatment. *BMC Proc.* 2018;12(Suppl 9). <https://doi.org/10.1186/s12919-018-0143-8>.
17. Zhou W, Lo SH. Analysis of genotype by methylation interactions through sparsity-inducing regularized regression. *BMC Proc.* 2018;12(Suppl 9). <https://doi.org/10.1186/s12919-018-0145-6>.
18. Wang MH, Sun R, Guo J, Weng H, Lee J, Hu I, Sham PC, Zee BC. A fast and powerful W-test for pairwise epistasis testing. *Nucleic Acids Res.* 2016;44(21):10526.
19. Balfour JA, Mctavish D, Heel RC. Fenofibrate-a review of its pharmacodynamic and pharmacokinetic properties and therapeutic use in dyslipidemia. *Drugs.* 1990;40(2):260–90.
20. Hu K, Li J. Detection and analysis of CpG sites with multimodal DNA methylation level distributions and their relationships with SNPs. *BMC Proc.* 2018;12(Suppl 9). <https://doi.org/10.1186/s12919-018-0141-x>.
21. Gauderman WJ, Murcray C, Gilliland F, Conti DV. Testing association between disease and multiple SNPs in a candidate gene. *Genet Epidemiol.* 2007;31:383–95.
22. Veenstra J, Kalsbeek A, Koster K, Ryder K, Bos A, Huisman J, et al. Epigenome wide association study of SNP-CpG interactions on changes in triglyceride levels after pharmaceutical intervention: a GAW20 analysis. *BMC Proc.* 2018;12(Suppl 9). <https://doi.org/10.1186/s12919-018-0144-7>.
23. Dudoit S, Schaffer JP, Boldrick JC. Multiple hypothesis testing in microarray experiments. *Stat Sci.* 2003;18:71–103.

**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)

