Using Genetic Burden Scores for Gene-by-Methylation Interaction Analysis on Metabolic Syndrome in African Americans

Jacquelyn Y. Taylor, PhD, PNP-BC, RN, FAHA, FAAN¹,
Erin B. Ware, PhD, MPH², Michelle L. Wright, PhD, RN³ ,
Jennifer A. Smith, PhD, MPH⁴, and Sharon L. R. Kardia, PhD⁵

Abstract

With the rapid advancement of omics-based research, particularly big data such as genome- and epigenome-wide association studies that include extensive environmental and clinical variables, data analytics have become increasingly complex. Researchers face significant challenges regarding how to analyze multifactorial data and make use of the findings for clinical translation. The purpose of this article is to provide a scientific exemplar for use of genetic burden scores as a data analysis method for studies with both genotype and DNA methylation data in which the goal is to evaluate associations with chronic conditions such as metabolic syndrome (MetS). This study included 739 African American men and women from the Genetic Epidemiology Network of Arteriopathy Study who met diagnostic criteria for MetS and had available genetic and epigenetic data. Genetic burden scores for evaluated genes were not significant after multiple testing corrections, but DNA methylation at 2 CpG sites (dihydroorotate dehydrogenase cg22381196 pFDR = .014; CTNNA3 cg00132141 pFDR = .043) was significantly associated with MetS after controlling for multiple comparisons. Interactions between the marginally significant CpG sites and burden scores, however, were not significant. More work is required in this area to identify intermediate biological pathways influenced by environmental, genetic, and epigenetic variation that may explain the high prevalence of MetS among African Americans. This study does serve, however, as an example of the use of the genetic burden score as an alternative data analysis approach for complex studies involving the analysis of genetic and epigenetic data simultaneously.

Keywords

genetic burden score, DNA methylation, $G \times E$ interaction, African American, metabolic syndrome, Gene $\times$ Methylation interaction

Advances in genomics and associated analytic technologies have led to renewed interest in personalized medicine and resulted in the New Initiative on Precision Medicine (Collins & Varmus, 2015). Within the initiative, one of the main targets for improving morbidity and mortality is to improve oncology treatment and outcomes since inherited genetic mutations can significantly contribute to cancer risk. However, the development of cancer is largely attributed to accumulated genetic damage over the life course, much of which is subsequent to environmental burden from exposures, diet, lifestyle, and such. Despite continued research after completion of the sequencing of the human genome in 2003, identification of genetic risk factors for many complex disease states has remained elusive or difficult to replicate across studies. Articles in the recent Primer in Genetics and Genomics series published in *Biological Research for Nursing* discussed some of the reasons for the difficulty of identifying specific variants associated with disease, including epigenetic

variation and gene-gene interactions (Fessele & Wright, 2018; Stanfill & Starlard-Davenport, 2018). Given that up to 70–90% of risk associated with the development of complex disease states and life expectancy are environmental in origin (Dwyer-Lindgren et al., 2017; Wild, 2012), it is crucial to also consider gene-environment ($G \times E$) interactions.

¹ New York University, New York, NY, USA

² Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

³ Emory University, Atlanta, Georgia

⁴ School of Public Health and Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

⁵ School of Public Health, University of Michigan, Ann Arbor, MI, USA

Corresponding Author:

Jacquelyn Y. Taylor, PhD, RN, Rory Meyers College of Nursing, New York University, 433 1st Avenue, New York, NY 10010, USA.

Email: jt139@nyu.edu

The concept of $G \times E$ interactions has been a topic of debate since the early twentieth century (Fisher, 1918, 1978) when scientists first posed the question of “nature versus nurture” (Tabery, 2014, 2015). Several statistical tools have been developed to help quantify the amount that environmental and/or genetic factors contribute to disease risk (Han & Chatterjee, 2018). Similar to the additive models described in gene–gene interactions (Stanfill & Starlard-Davenport, 2018), these $G \times E$ statistical tests measure the contributions from environmental and genetic factors in an additive or multiplicative fashion. This method allows investigators to evaluate the contribution of $G \times E$ relationships by measuring how much the individual effects and the interactions among effects contribute to the phenotype of interest. Interestingly, $G \times E$ interactions associated with risk of disease do not always overlap with genetic variants that contribute to increased risk without environmental influence. For example, Parnell and colleagues (2014) evaluated single nucleotide polymorphisms (SNPs) included in $G \times E$ interactions from nearly 400 studies investigating cardiovascular outcomes and determined that SNPs from significant $G \times E$ interactions tended to differ from the isolated SNPs that were associated with these cardiovascular outcomes.

One method for analyzing the contribution of multiple genetic variants to disease risk while accounting for unknown environmental risk factors is to utilize genetic burden scores (Morgenthaler & Thilly, 2007). This method may be used with case–control study data and uses the presence of risk alleles in a region to identify association with disease risk. The method is robust in detecting associations for $\sim 80\%$ of heterozygous and dominant risk-conferring alleles while accounting for number of cases, size of the cohort, diagnostic accuracy, and multigenic confounding. The purpose of this article is to provide a scientific exemplar for use of genetic burden scores as a data analysis method for studies with both genetic and DNA methylation (DNAm) data, when the goal is to evaluate associations with chronic conditions such as metabolic syndrome (MetS).

Background

MetS is a combination of medical conditions, potentially including hypertension (HTN), diabetes, obesity, renal disease, and/or hyperlipidemia, that leads to increased risk of atherosclerosis and cardiovascular-related mortality. It is estimated that MetS affects more than 34% of the U.S. population and accounts for a 3-fold increase in risk of cardiovascular-related death (Ervin, 2009; Ford, Li, & Zhao, 2010; Mozaffarian et al., 2015). African Americans (AAs) have a higher prevalence of HTN (41.4%), obesity (women = 51%, men = 37%), diabetes (11.8%), and renal disease (19.9%) than their White counterparts and are at increased risk of the clinical sequela of MetS (Mozaffarian et al., 2015). Risk factors associated with MetS mirror those associated with HTN, diabetes, renal disease, dyslipidemia, and obesity. Patients with MetS tend to be older and heavier and to have a sedentary lifestyle. Gender, ethnicity, and

genetic precursors may also affect risk of MetS. Since MetS represents the combined effects of multiple medical diagnoses and overlapping risk factors, pinpointing the causative nature of the MetS phenotype has been difficult.

Mechanisms that lead to the development of MetS are likely complex and involve interactions among genetic, epigenetic, and environmental factors. Previous studies have identified SNPs that are associated with variable risk of developing MetS among AAs (Tekola-Ayele et al., 2015; Zubair et al., 2016). Some of the SNPs identified are pleiotropic, meaning that they are associated with more than one measure of MetS (e.g., triglyceride levels and blood pressure [BP]).

Materials and Methods

Sample

The Genetic Epidemiology Network of Arteriopathy (GENOA) is a multiphase, community-based prospective study of sibships with two or more siblings diagnosed with primary HTN before the age of 60; however, all additional siblings were invited to participate regardless of HTN status. Self-identified Black participants were sampled from the Jackson, MS, area. Phase I (1995–2000) included 1,854 AA participants from 683 sibships, with 1,482 returning for Phase II (2000–2005). Participants were interviewed to collect demographic data and medical history in the morning after an overnight fast (Daniels et al., 2004). Since many of the biomarkers that we are examining in the present study were only measured in Phase II, we only used Phase II data. Peripheral blood samples used for analysis of DNAm and serum lipid levels were taken at Phase II. The University of Mississippi Medical Center and the University of Michigan Institutional Review Boards approved the GENOA Study protocol.

Demographic, Anthropometric, and Clinical Measures

Clinical data for participants in the GENOA Study were available for 1,482 participants from Phase II. These data included age, sex, fasting plasma glucose level, diabetes medication status, self-report of doctor diagnosis of high blood sugar, systolic and diastolic BP (mm/Hg), current cigarette smoker (yes/no), serum high-density lipoprotein cholesterol (HDL-C) level (mmol/L), serum triglycerides (TG) level (mmol/L), waist-to-hip ratio, body mass index (BMI; kg/m^2), urinary albumin excretion ratio ($\mu\text{g}/\text{min}$), and albumin-to-creatinine ratio (mg/g). Serum TG and HDL-C levels were measured using standard enzymatic methods on a Hitachi 911 Chemistry Analyzer (Roche Diagnostics, Indianapolis, IN).

To determine the presence of the primary outcome of interest, MetS (yes/no), we used the World Health Organization's criteria for MetS (Simmons et al., 2010), which include the presence of at least one of four clinical traits: (1) diabetes mellitus diagnosis, (2) impaired glucose tolerance, (3) impaired fasting glucose, or (4) insulin resistance—along with at least two of four additional clinical traits: (1) HTN (BP $\geq 140/90$ mmHg), (2) dyslipidemia (TG level ≥ 1.695 mmol/L and HDL-C ≤ 0.9 mmol/L for males and 1.0 mmol/L for females), (3)

Table 1. Sample Characteristics for African Americans Participating in the Genetic Epidemiology Network of Arteriopathy Study.

Variable	Full Sample, <i>N</i> = 739 <i>M</i> (SD)	Male, <i>n</i> = 203 <i>M</i> (SD)	Female, <i>n</i> = 536 <i>M</i> (SD)
Age, years	66.95 (7.43)	66.62 (7.47)	67.85 (7.26)
Smoker, yes	77 (10.4)	37 (18.2)	40 (7.5)
Systolic blood pressure, mmHg	154.12 (21.68)	154.89 (21.7)	152.08 (21.53)
Diastolic blood pressure, mmHg	91.93 (11.59)	91.13 (11.5)	94.03 (11.58)
HDL-C, mmol/L	1.52 (0.47)	1.6 (0.48)	1.3 (0.38)
Triglycerides, mmol/L	1.32 (0.62)	1.32 (0.62)	1.32 (0.65)
BMI, kg/m ²	31.44 (6.46)	32.3 (6.79)	29.16 (4.82)
Waist-to-hip ratio	0.86 (0.07)	0.88 (0.06)	0.94 (0.05)
Urinary albumin excretion ratio, µg/min	64.97 (305.11)	67.92 (343.87)	57.13 (162.84)
Albumin-to-creatinine ratio, mg/g	62.33 (316.71)	66.53 (356)	51.18 (174.01)
	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Diabetes, yes	311 (42.1)	90 (44.3)	221 (41.2)
Impaired fasting glucose	84 (11.4)	26 (12.8)	58 (10.8)
Hypertension ^a	605 (81.9)	166 (81.8)	439 (81.9)
Dyslipidemia ^b	34 (4.6)	10 (4.9)	24 (4.5)
Central obesity ^c	632 (85.5)	173 (85.2)	459 (85.6)
Microalbuminuria ^d	204 (27.6)	63 (31.0)	141 (26.3)
Metabolic syndrome	263 (35.6)	79 (38.9)	184 (34.3)

Note. BMI = body mass index; HDL-C = high-density lipoprotein cholesterol.

^aHypertension was defined as blood pressure $\geq 140/90$ mmHg. ^bDyslipidemia was defined as triglycerides level ≥ 1.695 mmol/L and HDL-C level ≤ 0.9 mmol/L for males and ≤ 1.0 mmol/L for females. ^cCentral obesity was defined as waist: hip ratio ≥ 0.90 for males and ≥ 0.85 for females, or BMI ≥ 30 kg/m². ^dMicroalbuminuria was defined as urinary albumin excretion ratio ≥ 20 µg/min or albumin: creatinine ratio ≥ 30 mg/g.

central obesity (waist-to-hip ratio ≥ 0.90 for males and 0.85 for females) or BMI ≥ 30 kg/m², or (4) microalbuminuria (urinary albumin excretion ratio ≥ 20 µg/min or albumin: creatinine ratio ≥ 30 mg/g). Of the 1,482 AA participants in Phase II with clinical data available, 1,467 had complete data on demographic, anthropometric, and clinical measures.

Genotype Measures

A total of 1,599 GENOA participants had genotyping data available. Genotyping was completed on the Illumina Human 1M-Duo BeadChip or the Affymetrix Genome-Wide Human SNP Array 6.0. Quality control for individuals included removing samples if they had a missing call rate ≥ 0.05 or were an outlier ≥ 6 standard deviations from the mean of the first 10 genome-wide principal components from genotype data. SNPs with a missing call rate ≥ 0.05 were removed. Imputation to the 1,000 Genomes project Phase I integrated variant set release (v3) in National Center for Biotechnology Information build 37 (hg19) coordinates (released in March 2012) for the Affymetrix- and Illumina-genotyped samples were performed separately. For each type, samples were prephased using the Segmented HAPlotype Estimation & Imputation Tool (SHA-PEIT), Version v2.r, using HapMap Phase II b37. Imputation was performed using IMPUTE, Version 2. Since the two genotyping platforms contain only a small number of overlapping SNPs ($\sim 200,000$), association analyses were performed using only imputed data with INFO quality scores greater than 0.5. Four genetic principal components were estimated and used in analyses to control for population stratification within the AA

genetic sample using independent SNPs with minor-allele frequencies (MAFs) $> 5\%$. Of the 1,467 participants with complete data for demographic, anthropometric, and clinical measures, 1,267 also had genetic data available.

Methylation Measures

Detailed description of the DNAm processing has been described elsewhere (Wright, Ware, Smith, Kardia, & Taylor, 2018). Briefly, DNA was extracted from peripheral blood leukocytes and processed using the EZ DNA Methylation Gold Kit (Zymo Research, Orange, CA) and the Illumina[®] Infinium HumanMethylation27 BeadChips and Illumina BeadXpress reader at the Mayo Clinic Advanced Technology Center. Quality control to remove failed samples and initial raw data processing was completed using the *lumi* package (Du, Kibbe, & Lin, 2008) in R (free programming language and statistical software environment). A total of 739 samples passed all quality control metrics and had complete clinical and genotyping data available for analyses in the present study. Only participants who reported fasting for more than 10 hr prior to sample collection were included in final analyses.

Gene Selection

We identified nine genes from two studies among AA populations that were significantly associated ($p < 5 \times 10^{-8}$) with either MetS (Tekola-Ayele et al., 2015) or lipid traits (HDL-C, low-density lipoprotein cholesterol [LDL-C], or TG; Zubair et al., 2016; see Supplemental Table 1). Only five of the genes

(*CTNNA3* from Tekola-Ayele et al., 2015, and *LPL*, *APOA5*, *LCAT*, and *DHODH* from Zubair et al., 2016) contain DNAm sites on the 27 k chip. Gene regions were defined using University of California Santa Cruz (UCSC) Genome Browser genome browser start and stop position ± 5 kb for build 37. All imputed SNPs with imputation quality > 0.5 were included.

Data availability for the GENOA Study is as follows: GENOA genotype and phenotype data are available through the Database of Genotypes and Phenotypes (accession number phs001238.v1.p1). DNAm data are available with a data use agreement upon reasonable request to the study investigators by contacting S. Kardia (skardia@umich.edu) or J. Smith (smjenn@umich.edu).

Statistical Methods

We examined MetS as our outcome of interest using a generalized linear mixed-effects model for the binary outcome (R package lme4, glmer function; Bates, Machler, Bolker, & Walker, 2015). We examined only gene regions from index SNPs identified in previous lipid and MetS genome-wide association studies in AA that had corresponding DNAm sites. For each gene region, we created a cohort allelic sums test (Morgenthaler & Thilly, 2007), or burden score, for the rare variants in the region:

$$C_i = \begin{cases} 1 & \text{if } \sum_{j=1}^p g_{ij} > 0 \\ 0 & \text{if } \sum_{j=1}^p g_{ij} = 0 \end{cases},$$

where i is the individual, j is the variant, and p is the total number of variants in the region. A genetic burden score of $C_i = 0$ is assigned when there are no minor alleles in the gene region, and $C_i = 1$ is assigned otherwise. This model assumes that the presence of any rare variant increases disease risk. To focus on rare variants, we further limit p by allowing only variants with MAFs less than a certain threshold (MAF $< 5\%$) to enter the calculation. For each genetic-burden test, models were adjusted by age, sex, and the top four ancestry-specific genetic principal components to control for population stratification, as well as sibship modeled as a random effect.

We also examined the association between each of the 10 DNAm sites and MetS (Supplemental Table 2). M values, calculated as the $\log_2$ ratio of the intensities of methylation probe versus unmethylated probe, were calculated for each DNAm site. Positive M values mean that more molecules are methylated than unmethylated, while negative M values mean the opposite (Du et al., 2010). We used the M value of DNAm levels because the β value has a bounded range from 0 to 1 that violates the Gaussian distribution assumption. We first adjusted each DNAm site for peripheral blood cell heterogeneity using the Houseman correction method, technical covariates (i.e., DNAm chip and position), and smoking status (ever/never; Houseman et al., 2012; Houseman, Molitor, & Marsit,

Table 2. Association of Genetic Burden Score With Metabolic Syndrome in African Americans Participating in the Genetic Epidemiology Network of Arteriopathy Study.

Gene	Chr	Odds	95% CI	p	p_{FDR}
<i>LCAT</i>	16	0.81	[0.58, 1.12]	.20 ^a	.77
<i>LPL</i>	8	1.13	[0.76, 1.69]	.55	.77
<i>DHODH</i>	16	1.08	[0.77, 1.50]	.67	.77
<i>CTNNA3</i> ^b	10	1.06	[0.76, 1.48]	.72	.77
<i>APOA5</i>	11	1.05	[0.76, 1.46]	.77	.77

Note. $N = 739$. Genetic burden score is calculated for a gene region as the number of risk alleles for variants with minor-allele frequency (MAF) $< 5\%$ dichotomized as none/any. Gene regions were defined using the UCSC genome browser start and stop position ± 5 kb for build 37. All imputed single-nucleotide polymorphisms with imputation quality > 0.5 were included. Models were adjusted for age, sex, and top four genetic principal components and accounted for sibship structure as a random effect. Chr = chromosome, $p_{FDR} = p$ value corrected for false discovery rate.

^aIndicates p values $< .2$ threshold for inclusion in interaction. ^bBecause the *CTNNA3* gene had no variation in the burden score at MAF $< 5\%$, we examined the genetic variants at MAF $< 1\%$: $n = 445$, 60% of the sample had at least one minor variant at an MAF of 1%.

Table 3. Association of DNA Methylation With Metabolic Syndrome in African Americans Participating in the Genetic Epidemiology Network of Arteriopathy Study.

Chr	DNAm CpG Site	Associated Gene	Odds	95% CI	p	p_{FDR}
16	cg22381196	<i>DHODH</i>	0.76	[0.63, 0.90]	.002 ^{a,b}	.022
10	cg00132141	<i>CTNNA3</i>	1.28	[1.06, 1.54]	.009 ^{a,b}	.045
16	cg07817698	<i>DHODH</i>	1.15	[0.97, 1.36]	.118 ^a	.307
16	cg01489608	<i>LCAT</i>	1.13	[0.97, 1.33]	.123 ^a	.307
11	cg25682080	<i>APOA5</i>	1.13	[0.95, 1.35]	.157 ^a	.314
8	cg22108175	<i>LPL</i>	0.92	[0.77, 1.09]	.320	.533
8	cg08918749	<i>LPL</i>	0.93	[0.79, 1.11]	.432	.617
10	cg09538287	<i>CTNNA3</i>	0.96	[0.82, 1.12]	.589	.737
11	cg02157083	<i>APOA5</i>	0.97	[0.82, 1.15]	.715	.789
16	cg26924825	<i>LCAT</i>	0.98	[0.83, 1.15]	.789	.789

Note. $N = 739$. Generalized linear mixed effects were used to account for sibships. Models were adjusted for age and sex (except the sex-stratified models). The associated gene was provided by the Illumina 27 K chip annotation file. Results are sorted by p value. Chr = chromosome; DNAm = DNA methylation adjusted for smoking (ever/never), batch effects, and cell distribution using the Houseman method; MetS = metabolic syndrome; $p_{FDR} = p$ value corrected for false discovery rate.

^a p value $< .2$ threshold for inclusion in interaction analysis. ^bSignificant association with MetS after FDR correction for multiple testing at $\alpha_{FDR} = .05$.

2014). We then used generalized linear mixed-effect models to analyze each adjusted DNAm site.

To assess interaction effects, we carried forward genetic burden and DNAm sites with uncorrected p values $< .2$. We chose, a priori, to carry forward sites with a less stringent p value because a range of interaction effect sizes can be detected even when marginal effects are not detectable (Thomas, 2010). Statistical analysis code for this manuscript is available on GitHub under GNU General Public License v2.0 (<https://github.com/erinbware/GeneBurdenMetS>).

Table 4. Genetic Burden $\times$ DNAm Interaction on Metabolic Syndrome in African Americans Participating in the Genetic Epidemiology Network of Arteriopathy Study, $N = 739$.

	Model 1			Model 2			Model 3			Model 4		
	Odds	95% CI	<i>p</i>	Odds	95% CI	<i>p</i>	Odds	95% CI	<i>p</i>	Odds	95% CI	<i>p</i>
<i>LCAT</i>	0.81	[0.58, 1.12]	.20				1.30	[0.92, 1.79]	0.145	1.28	[0.92, 1.79]	.143
cg01489608				1.13	[0.97, 1.33]	.123	1.15	[0.97, 1.35]	0.102	1.14	[0.97, 1.35]	.671
<i>LCAT</i> $\times$ cg01489608										1.30	[0.88, 1.70]	.229

Note. $N = 739$. Generalized linear mixed effects were used to account for sibships. DNAm = DNA methylation adjusted for smoking (ever/never), batch effect, and cell distribution using the Houseman method; PC = genetic principal component. Model 1: Metabolic syndrome = PC1 + PC2 + PC3 + PC4 + Age + Sex + Burden. Model 2: Metabolic syndrome = Age + Sex + DNAm. Model 3: Metabolic syndrome = PC1 + PC2 + PC3 + PC4 + Age + Sex + Burden + DNAm. Model 4: Metabolic syndrome = PC1 + PC2 + PC3 + PC4 + Age + Sex + Burden + DNAm + Burden $\times$ DNAm.

Results

The sample comprised self-identified AA men and women. Demographic data are available in Table 1. Our participants were predominately female, older, and obese, and more than one third had MetS (Meyer, Yoon, & Kaufmann, 2013). We have combined males and females within the same cohort when presenting our results, controlling for sex statistically as a covariate, per models described above. Neither sex ($p > .21$) nor age ($p > .06$) was significant predictors of the presence of MetS in any of our models. We did not identify any significant associations between the five gene regions and MetS (Table 2). We did find that two CpG sites (*DHODH* cg22381196 $p_{FDR}[\text{false discovery rate}] = .014$; *CTNNA3* cg00132141 $p_{FDR} = .043$) were significantly associated with MetS after controlling for multiple comparisons (Table 3). Also, one gene region (*LCAT*) and five DNAm sites (*DHODH* cg22381196; *CTNNA3* cg00132141; *DHODH* cg07817698; *LCAT* cg01489608; *APOA5* cg25682080) had nominal p values $< .2$; however, we carried only the *LCAT*-cg01489608 combination forward for interaction analyses and detected no significant interaction effects (Table 4).

Discussion

Although no gene regions were independently significantly associated with MetS after multiple testing correction, we did find two CpG sites that were significantly associated with MetS: *DHODH* cg22381196 and *CTNNA3* cg00132141.

Dihydroorotate dehydrogenase (*DHODH*) is a protein-coding gene located on chromosome 16 and functional within the mitochondria of cells. The gene produces the enzyme *DHODH*, which contributes to the production of pyrimidines is needed for DNA and RNA synthesis. Researchers have identified mutations within this gene as causal factors for postaxial acrofacial dysostosis/Miller syndrome, an autosomal recessive inherited condition (Ng et al., 2010). Individuals with Miller syndrome suffer from bone abnormalities in facial structure and limbs and may also (less frequently) have issues with cardiovascular and renal abnormalities. *DHODH* has also been associated with various cancers because of its role in nucleotide metabolism (Liu et al., 2017). Furthermore, epigenetic and genetic variants within *DHODH* have been associated with

various clinical traits frequently observed among AAs with MetS. Sun et al. (2013) found that hypomethylation of *DHODH* cg22381196 was significantly associated with inflammatory biomarkers such as C-reactive protein in studies of AAs. Additionally, Zubair et al. (2016) reported that SNP variations within *DHODH* were associated with LDL-C levels among AAs. Because inflammation plays an important role in determining levels of LDL-C (a key component in MetS), it is possible that changes in this CpG site influence metabolic and inflammatory pathways associated with MetS in this population.

Catenin α 3 (*CTNNA3*) is a protein-coding gene located on chromosome 10. The protein for which it codes is responsible for cell-cell adhesion. Mutations in this gene result in cardiovascular disorders such as arrhythmogenic right ventricular dysplasia familial 13 and arrhythmogenic right ventricular cardiomyopathy (Corrado, Link, & Calkins, 2017). Tekola-Ayele et al. (2015) showed that *CTNNA3*'s association with MetS is specific to individuals with African ancestry, with alterations in the gene including both a risk and a protective factor for MetS. The MetS trait that *CTNNA3* is most notably associated with is waist circumference, but it has also been associated with insulin resistance in MetS (Brown & Walker, 2016). In fact, Brown and Walker also successfully utilized a genetic burden score in this gene to examine the genetics of insulin resistance and MetS. Literature specifically related to the CpG site cg00132141 is sparse, however; therefore, replication is needed before conclusions may be drawn regarding the influence of DNAm at this site on MetS.

Implications for Future Research

Future work could extend the use of genetic burden scores for determining genetic associations for complex chronic conditions by exploring other statistical packages such as famSKAT (Chen, Meigs, & Dupuis, 2012) and/or GRAMMAR+ transformation in the R package GenABEL (Amin, van Duijn, & Aulchenko, 2007; Aulchenko, de Koning, & Haley, 2007; Aulchenko, Ripke, Isaacs, & van Duijn, 2007) for analyses. These packages allow gene-level inference using genetic data from family studies. Gene-level inference may help identify

intermediate biological pathways that might explain the high prevalence of MetS among AAs.

Limitations

For the present study, we used preexisting data from the GENOA cohort; therefore, we did not have control on how and what data were collected. Although the sample size was small ($N = 739$) for an omics study, the power was sufficient to detect associations between DNAm and MetS in the small number of genes analyzed. There were more female than male participants in the study, but we controlled for sex in the analysis. For this data set, we used the 27 K DNAm array; however, the data from the Illumina EPIC DNAm array (850 K CpG sites) are now available in GENOA, which will provide wider coverage of the epigenome.

Replication is important for validation of genetic and epigenetic effects. We will, therefore, seek other consortia with similar samples for future work. In the present study, we measured DNAm at only one point in time. Future studies should include the investigation of the effect of baseline DNAm and also repeat DNAm analysis to assess change over time and its influence on MetS traits.

Conclusions

Mechanisms that lead to the development of MetS are complex and likely involve interactions among genetic, epigenetic, and environmental factors. Studies on epigenetic-by-risk factor and gene-by-epigenetic factor interactions on MetS and related traits are only beginning to emerge but represent an important avenue of research into these mechanisms. Because of the complexity of analyzing genetic and epigenetic data simultaneously, the genetic burden score may be a useful alternative to single SNP data analysis in such studies.

Although alterations in key genes regulating MetS have been associated with individual traits that comprise MetS, no published studies have integrated the genetic and epigenetic risk factors that influence MetS traits among AAs. If these factors can be identified early in the disease process, opportunities will exist for the implementation of prevention strategies to reduce expression of MetS traits in AAs, a group that is disproportionately affected by this syndrome and related complications but is underrepresented in research studies. More work in this area is required to identify intermediate biological pathways influenced by genes and epigenetic markers that can help explain the high prevalence of MetS traits in AAs. Future work could also highlight therapeutic targets for subsequent interventional and translational studies for clinical prevention and treatment of MetS.

Acknowledgments

Support for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the National Heart, Lung and Blood Institute (HL054457, HL100185, HL 087660, HL119443, and HL133221); additional support for this project was provided by the National Institute of Nursing Research (NINR; R01NR013520) and the National Institute on Aging (NIA; P30AG012846).

Author Contributions

J. Taylor contributed to conception, design, acquisition, and interpretation; drafted the manuscript; critically revised the manuscript; gave final approval; and agreed to be accountable for all aspects of work ensuring integrity and accuracy. E. Ware contributed to design, acquisition, analysis, and interpretation; drafted the manuscript; critically revised the manuscript; gave final approval; and agreed to be accountable for all aspects of work ensuring integrity and accuracy. M. Wright contributed to interpretation, drafted the manuscript, critically revised the manuscript, gave final approval, and agreed to be accountable for all aspects of work ensuring integrity and accuracy. J. Smith contributed to design, acquisition, analysis, and interpretation; critically revised the manuscript; gave final approval; and agreed to be accountable for all aspects of work ensuring integrity and accuracy. S. Kardia contributed to design and acquisition, critically revised the manuscript, gave final approval and agreed to be accountable for all aspects of work ensuring integrity and accuracy.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Funding for the Genetic Epidemiology Network of Arteriopathy (GENOA) was provided by the National Heart, Lung and Blood Institute (HL054457, HL100185, HL 087660, HL119443, and HL133221); additional funding for this project was provided by the National Institute of Nursing Research (NINR; R01NR013520) and the National Institute on Aging (NIA; P30AG012846).

ORCID iD

Michelle L. Wright <https://orcid.org/0000-0002-9348-8740>

Supplemental Material

Supplemental material for this article is available online.

References

- Amin, N., Van Duijn, C. M., & Aulchenko, Y. S. (2007). A genomic background based method for association analysis in related individuals. *PLoS One*, 2, e1274.
- Aulchenko, Y. S., de Koning, D. J., & Haley, C. (2007). Genomewide rapid association using mixed model and regression: A fast and simple method for genome-wide pedigree-based quantitative trait loci association analysis. *Genetics*, 177, 577–585.
- Aulchenko, Y. S., Ripke, S., Isaacs, A., & van Duijn, C. M. (2007). GenABEL: An R package for genome-wide association analysis. *Bioinformatics*, 23, 1294–1296.
- Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01
- Brown, A. E., & Walker, M. (2016). Genetics of insulin resistance and the metabolic syndrome. *Current Cardiology Reports*, 18, 75. doi: 10.1007/s11886-016-0755-4
- Chen, H., Meigs, J. B., & Dupuis, J. (2012). Sequence kernel association test for quantitative traits in family samples. *Genetic Epidemiology*, 37, 196–204.

- Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, 372, 793–795. doi:10.1056/NEJMp1500523
- Corrado, D., Link, M. S., & Calkins, H. (2017). Arrhythmogenic right ventricular cardiomyopathy. *New England Journal of Medicine*, 376, 61–72. doi:10.1056/NEJMra1509267
- Daniels, P. R., Kardia, S. L., Hanis, C. L., Brown, C. A., Hutchinson, R., Boerwinkle, E., & Turner, S. T. (2004). Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *American Journal of Medicine*, 116, 676–681. doi:10.1016/j.amjmed.2003.12.032
- Du, P., Kibbe, W. A., & Lin, S. M. (2008). lumi: A pipeline for processing Illumina microarray. *Bioinformatics*, 24, 1547–1548.
- Du, P., Zhang, X., Huang, C., Jafari, N., Kibbe, W. A., Hou, L., & Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, 11, 587. doi:10.1186/1471-2105-11-587
- Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., Mackenbach, J. P., van Lenthe, F. J., ... Murray, C. J. L. (2017). Inequalities in life expectancy among US counties, 1980 to 2014. *JAMA Internal Medicine*, 177, 1003. doi:10.1001/jamainternmed.2017.0918
- Ervin, R. B. (2009, May 5). *Prevalence of metabolic syndrome among adults 20 years of age and over, by sex, age, race and ethnicity, and body mass index: United States, 2003–2006 (National Health Statistics Reports, no. 13)*. Hyattsville, MD: National Center for Health Statistics.
- Fessele, K. L., & Wright, F. (2018). Primer in genetics and genomics, Article 6: Basics of epigenetic control. *Biological Research For Nursing*, 20, 103–110. doi:10.1177/1099800417742967
- Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399–433.
- Fisher, R. A. (1978). *The life of a scientist*. New York, NY: Wiley.
- Ford, E. S., Li, C., & Zhao, G. (2010). Prevalence and correlates of metabolic syndrome based on a harmonious definition among adults in the US. *Journal of Diabetes*, 2, 180–193. doi:10.1111/j.1753-0407.2010.00078.x
- Han, S. S., & Chatterjee, N. (2018). Review of statistical methods for gene–environment interaction analysis. *Current Epidemiology Reports*, 5, 39–45. doi:10.1007/s40471-018-0135-2
- Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., ... Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, 13, 86. doi:10.1186/1471-2105-13-86
- Houseman, E. A., Molitor, J., & Marsit, C. J. (2014). Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics (Oxford, England)*, 30, 1431–1439.
- Liu, L., Dong, Z., Lei, Q., Yang, J., Hu, H., Li, Q., ... Cui, H. (2017). Inactivation/deficiency of DHODH induces cell cycle arrest and programmed cell death in melanoma. *Oncotarget*, 8, 112354–112370. doi:10.18632/oncotarget.19379
- Meyer, P. A., Yoon, P. W., & Kaufmann, R. B. (2013). CDC health disparities and inequalities report—United States, 2013. *MMWR Supplements*, 62, 1–187.
- Morgenthaler, S., & Thilly, W. G. (2007). A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). *Mutation Research—Fundamental and Molecular Mechanisms of Mutagenesis*, 615, 28–56. doi:10.1016/j.mrfmmm.2006.09.003
- Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., ... Turner, M. B. (2015). Heart disease and stroke statistics-2015 update : A report from the American Heart Association. *Circulation*, 131. doi:10.1161/CIR.0000000000000152
- Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., ... Bamshad, M. J. (2010). Exome sequencing identifies the cause of a Mendelian disorder. *Nature Genetics*, 42, 30–35. doi:10.1038/ng.499
- Parnell, L. D., Blokker, B. A., Dashti, H. S., Nesbeth, P.-D., Cooper, B. E., Ma, Y., ... Ordovás, J. M. (2014). CardioGxE, a catalog of gene–environment interactions for cardiometabolic traits. *BioData Mining*, 7, 21. doi:10.1186/1756-0381-7-21
- Simmons, R. K., Alberti, K. G. M. M., Gale, E. A. M., Colagiuri, S., Tuomilehto, J., Qiao, Q., ... Roglic, G. (2010). The metabolic syndrome: Useful concept or clinical tool? Report of a WHO Expert Consultation. *Diabetologia*, 53, 600–605. doi:10.1007/s00125-009-1620-4
- Stanfill, A. G., & Starlard-Davenport, A. (2018). Primer in genetics and genomics, Article 7—Multifactorial concepts: Gene–gene interactions. *Biological Research For Nursing*, 20, 359–364. doi:10.1177/1099800418761098
- Sun, Y. V., Lazarus, A., Smith, J. A., Chuang, Y.-H., Zhao, W., Turner, S. T., & Kardia, S. L. R. (2013). Gene-specific DNA methylation association with serum levels of C-reactive protein in African Americans. *PLoS One*, 8, e73480. doi:10.1371/journal.pone.0073480
- Tabery, J. (2014). *Beyond versus: The struggle to understand the interaction of nature and nurture*. Cambridge: MIT Press.
- Tabery, J. (2015). Debating interaction: The history, and an explanation. *International Journal of Epidemiology*, 44, 1117–1123. doi:10.1093/ije/dyv053
- Tekola-Ayele, F., Doumatey, A. P., Shriner, D., Bentley, A. R., Chen, G., Zhou, J., ... Rotimi, C. N. (2015). Genome-wide association study identifies African-ancestry specific variants for metabolic syndrome. *Molecular Genetics and Metabolism*, 116, 305–313. doi:10.1016/j.ymgme.2015.10.008
- Thomas, D. (2010). Gene–environment-wide association studies: Emerging approaches. *Nature Reviews Genetics*, 11, 259–272. doi:10.1038/nrg2764
- Wild, C. P. (2012). The exposome: From concept to utility. *International Journal of Epidemiology*, 41, 24–32. doi:10.1093/ije/dyr236
- Wright, M. L., Ware, E. B., Smith, J. A., Kardia, S. L. R., & Taylor, J. Y. (2018). Joint influence of SNPs and DNA methylation on lipids in African Americans from hypertensive sibships. *Biological Research For Nursing*, 20, 161–167. doi:10.1177/1099800417752246
- Zubair, N., Graff, M., Ambite, J. L., Bush, W. S., Kichaev, G., Lu, Y., ... Carty, C. L. (2016). Fine-mapping of lipid regions in global populations discovers ethnic-specific signals and refines previously identified lipid loci. *Human Molecular Genetics*, 25, 5500–5512. doi:10.1093/hmg/ddw358