IDENTIFICATION OF GENE-GENE AND GENE-ENVIRONMENT INTERACTIONS WITHIN THE FIBRINOGEN GENE CLUSTER FOR FIBRINOGEN LEVELS IN THREE ETHNICALLY DIVERSE POPULATIONS *

JANINA M. JEFF†

Charles Bronfman Institute for Personalized Medicine , Icahn School of Medicine at Mount Sinai, 1468 Madison Ave.
New York, NY 10128, United States of America
Email: janina.jeff@mssm.edu

KRISTIN BROWN-GENTRY

Cigna-Health Spring, 500 Great Circle Drive
Nashville, TN 37228, United States of America
Email: kristin.gentry@healthspring.com

DANA C. CRAWFORD

Institute for Computational Biology, Department of Epidemiology and Biostatistics, Case Western Reserve University, Wolstein Research Building, 2103 Cornell Road, Suite 2527, Cleveland, OH 44106, USA
Email: dana.crawford@case.edu

Elevated levels of plasma fibrinogen are associated with clot formation in the absence of inflammation or injury and is a biomarker for arterial clotting, the leading cause of cardiovascular disease. Fibrinogen levels are heritable with >50% attributed to genetic factors, however little is known about possible genetic modifiers that might explain the missing heritability. The fibrinogen gene cluster is comprised of three genes (*FGA, FGB,* and *FGG)* that make up the fibrinogen polypeptide essential for fibrinogen production in the blood. Given the known interaction with these genes, we tested 25 variants in the fibrinogen gene cluster for gene x gene and gene x environment interactions in 620 non-Hispanic blacks, 1,385 non-Hispanic whites, and 664 Mexican Americans from a cross-sectional dataset enriched with environmental data, the Third National Health and Nutrition Examination Survey (NHANES III). Using a multiplicative approach, we added cross product terms (gene x gene or gene x environment) to a linear regression model and declared significance at $p < 0.05$. We identified 19 unique gene x gene and 13 unique gene x environment interactions that impact fibrinogen levels in at least one population at $p < 0.05$. Over 90% of the gene x gene interactions identified include a variant in the rate-limiting gene, *FGB* that is essential for the formation of the fibrinogen polypeptide. We also detected gene x environment interactions with fibrinogen variants and sex, smoking, and body mass index. These findings highlight the potential for the discovery of genetic modifiers for complex phenotypes in multiple populations and give a better understanding of the interaction between genes and/or the environment for fibrinogen levels. The need for more powerful and robust methods to identify genetic modifiers is still warranted.

† This work was supported in part by NIH U01 HG004798 and its ARRA supplements.

1. Introduction

Plasma fibrinogen is a key player in the pathogenesis of cardiovascular disease, specifically arterial thrombosis. In the coagulation pathway, fibrinogen is converted to fibrin, the main protein component of an arterial clot. [1; 2] Fibrinogen levels are highly heritable with reports of up to 50% of the trait variability attributable to genetic factors.[2] These estimates include genes within the fibrinogen gene cluster (*FGA*, *FGB*, and *FGG*) as well as other genes that regulate the production of fibrinogen.[3] However, collectively these loci do not account for the expected genetic component for fibrinogen levels. For complex traits such as fibrinogen, the search for highly penetrant disease loci that accurately predict fibrinogen levels has not been as successful compared to Mendelian diseases. Like many traits associated with cardiovascular disease, the role of the environment, several loci in the genome, or the interaction of amongst genes and/or with the environment can account for the heritability yet to be described for fibrinogen levels.

Fibrinogen synthesis is highly dependent on interactions between fibrinogen polypeptide pairs. Fibrinogen has three pairs of polypeptides Aα, Bβ, γ that are encoded by three genes: *FGA, FGB,* and *FGG*, respectively.[4] These polypeptides form disulfide bonds at the near their N-termini and once connected make up fibrinogen.[5; 6] Therefore it is highly plausible that interactions between genes within the fibrinogen gene cluster may affect fibrinogen levels. There have been reports in the literature on possible gene-gene interactions associated with fibrinogen levels from variants in the fibrinogen gene cluster and Factor XIII. [7] Missing from the literature is an association study of the possible gene x gene interactions between fibrinogen genes.

In addition to possible gene-gene interactions, the interplay between genes and environmental risk factors can explain the inter-individual variability of fibrinogen levels. Risk factors such as smoking, age, lack of the exercise, sex, use of contraceptives, and body mass index (BMI) significantly contribute to the variability of fibrinogen levels amongst individuals.[8] However, after accounting for these risk factors, a significant amount of the variance in fibrinogen levels remains unknown.

One of the primary goals for the first phase of the Population Architecture using Genomics and Epidemiology (PAGE I) consortium was to identify genetic modifiers of common disease.[9] Likewise, the Epidemiologic Architecture of Genes Linked to Environment (EAGLE), a member of the PAGE I consortium, has access to a diverse, cross-sectional survey, the Third National Health and Nutrition Examination Survey (NHANES III). NHANES III is rich in environmental data to enable the identification of genetic modifiers. Here, we examine the effect of genetic modifiers (gene-gene and gene-environment) with plasma fibrinogen levels in three populations: non-Hispanic blacks, non-Hispanic whites, and Mexican Americans from NHANES III. Using 25 SNPs in the fibrinogen gene cluster, we tested for and identified gene-gene interactions within this cluster, as well as gene-environment interactions with these SNPs and known environmental risk factors across all three NHANES populations.

2. Methods

2.1. *Study population & genotyping*

The National Health and Nutritional Examination Surveys are cross-sectional surveys conducted across the United States by the National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). NHANES III was conducted between 1988-1990 (phase 1) and 1991-1994 (phase 2) [10; 11] as a complex survey that over-sampled minorities, the young, and the elderly. All NHANES have interviews that collect demographic, socioeconomic, dietary, and health-related data. Participant interviews, physical examinations, and laboratory measures were used to collect environmental factors for this study. Current smoking status was determined by the question "Do you smoke cigarettes now?" and by cotinine levels >15 ng/ml. Participants that answered "yes" and have cotinine levels >15ng/ml were classified as a current smoker. Environmental variables are described in Table 1.

All NHANES study participants undergo a detailed medical examination at a central location, the Mobile Examination Center (MEC). Body Mass Index (BMI) was measured directly in the MEC as height and weight. Biomarkers were also collected in the MEC including plasma fibrinogen on participants > 40 years using the Clauss clotting method. [12] Participants with fibrinogen levels greater than > 4.0 g/L were excluded from the analysis since extreme measurements can indicate acute inflammation or recent trauma. Plasma fibrinogen levels were normally distributed in all three study populations (Table 1).

Beginning with phase 2 of NHANES III, DNA samples were collected from study participants aged 12 years and older. DNA was extracted from crude cell lysates from lymphoblastoid cell lines in NHANES III as part of genetic NHANES.[13] A total of 25 SNPs were genotyped using the Illumina GoldenGate assay (as part of a custom 384-oligonucleotide pool assay (OPA)) by the Center for Inherited Disease Research (CIDR) through the National Heart Lung and Blood Institute's Resequencing and Genotyping Service. A total of 7,159 samples were genotyped, including 2,631 non-Hispanic whites, 2,108 non-Hispanic blacks and 2,073 Mexican Americans. TagSNPs were selected using LDSelect[14] for multiple populations at r^2 >0.80 for common variants (minor allele frequency (MAF) >5%) in three candidate genes (*FGA, FGB*, and *FGG*) based on data available for European Americans and African Americans in SeattleSNPs [15]. Quality control measurements were calculated locally using the Platform for the Analysis, Translation, and Organization of large-scale data.[16] We flagged SNPs that deviated from Hardy Weinberg Equilibrium expectations (p-value <0.001), MAF<0.05, and SNP call rates <95% for each subpopulation. In addition to these quality control metrics, we genotyped blinded duplicates as required by CDC, and all SNPs reported passed quality control metrics required by CDC. All genotype data reported here were deposited into the NHANES III Genetic database and are available for secondary analysis through the CDC. After all inclusion/exclusion criteria were applied our final study population was comprised of 620 Non-Hispanic Blacks, 1,385 Non-Hispanic Whites, and 664 Mexican Americans.

Table 1. Study population characteristics and hematological trait descriptive statistics for NHANES III participants. Un-weighted means (± standard deviations) or percentages are given for demographics and plasma fibrinogen by subpopulation for adults >40 years in age in phase 2 of NHANES III. Abbreviations: non-Hispanic black (NHB), non-Hispanic whites (NHW), Mexican Americans (MA)

Variable	Mean or %			Standard Deviation		
	NHB	NHW	MA	NHB	NHW	MA
Age (yrs)	40.7	53.4	41	±16	±20	±17
Female (%)	58	60	50	-	-	-
Current Smokers (%)	37	26	24	-	-	-
Body Mass Index (kg/m^2)	28.2	26.6	27.7	±6.67	±5.56	±5.42
Plasma Fibrinogen (g/L)	2.95	2.93	2.97	±0.51	±0.51	±0.50

2.2 *Statistical Methods*

All analyses were performed using the Statistical Analysis Software (SAS v.9.2; SAS Institute, Cary, NC) either locally or via the Analytic Data Research by Email (ANDRE) portal of the CDC Research Data Center (RDC) in Hyattsville, MD. Using multivariate linear regression, cross product terms (all pair-wise gene-gene or gene-environment terms for 25 fibrinogen variants) were added to the regression models for plasma fibrinogen. All SNPs were coded additively and all models were adjusted for the main effect of the SNP (both the GxG and GxE models) and environmental factor (only the GxE models). In addition to adjusting for main effects, all analyses were adjusted for known covariates age, BMI, sex, and smoking status. Age and BMI were coded as continuous variables whereas sex and smoking status were coded as binary variables in all models. Any results or variables with counts less than 5 were 'suppressed' and unavailable for reporting by the CDC. All p-values are presented uncorrected for multiple testing, and we considered an interaction significant at $p<0.05$. We chose a liberal significance threshold for several reasons: 1.) we only tested 25 variants in a candidate gene setting, 2.) several variants are correlated yielding redundant results, and 3.) we have limited power (due to sample size and limited loci) to detect moderate (or small) effects.

3. Results

3.1 Gene-gene interactions

We identified a total of 18 unique significant gene-gene (GxG) interactions out of 828 tests performed that effect fibrinogen levels in at least one NHANES III population across all three NHANES III populations (Table 2). We define a significant GxG interaction here as having significant interaction term ($p<0.05$) from two different genes in the fibrinogen gene cluster.

Of the GxG interactions we detected, 50% were identified in non-Hispanic whites, all of which included a *FGB* SNP in the interaction term (Table 2). In fact, all of the significant interaction terms included either *FGB* rs2227395 or *FGB* rs4220 in non-Hispanic whites (Table 2). In our previous work, we demonstrated that these SNPs are in strong linkage disequilibrium ($r^2 = 0.86$) and that both SNPs had significant main effects. [17] After accounting for this correlation, only seven unique gene-gene interactions reached the significance threshold in non-Hispanic whites. Of the seven unique interaction terms, three were associated with increased fibrinogen levels and four were associated with decreased levels of fibrinogen.

In non-Hispanic blacks a total of six unique GxG interactions were associated with fibrinogen levels. Three GxG interactions were associated with decreased fibrinogen levels and three were associated with increased fibrinogen levels (Table 2). Missense variant rs6050 (Thr312Ala) in the *FGA* gene is a known association with fibrinogen levels in Europeans. In previous work, we were unable to detect this association with non-Hispanic blacks [17]; however, the interaction term rs6050 x rs2227395 (*FGG*) was significantly associated with decreased fibrinogen levels in non-Hispanic blacks ($\beta = 0.15$, p-value = 0.017, Table 2). Significant interactions terms with *FGB* variant rs2227395 were consistently observed with SNPs in the *FGA* gene (Table 2, last column). Based on our previous single SNP analysis, none of the SNPs identified in these interaction terms had a significant main effect (Table 2, column 3). [17] Overall there were four associations we observed across non-Hispanic whites and non-Hispanic blacks, all of which include rate the rate-limiting gene, *FGB* (Table 2).

In Mexican Americans, we identified six unique interaction terms at the $p < 0.05$ level. Based on the genetic similarity between non-Hispanic whites and Mexican Americans in this population, the associations including rs2227395 and rs4220, and may be correlated in this Mexican American population. After accounting for the correlation between SNPs, three interactions are associated with increased fibrinogen levels and two are associated with decreased levels (Table 2).

Table 2. Significant fibrinogen cluster gene-gene interactions in NHANES III for plasma fibrinogen. Using multivariate linear regression, we added cross product terms (all pair-wise SNP-SNP) to the regression models for plasma fibrinogen. Below are the association results for each SNP individually as well as the association results for the interaction term. Abbreviations: SE = standard error.

Gene-Gene Interactions	SNP-SNP Interactions	SNP 1 Effect		SNP 2 Effect		Interaction Term	
		Beta (SE)	P	Beta (SE)	P	Beta (SE)	P
Non-Hispanic Blacks							
FGB x *FGG*	rs1800791 x rs1800792	-0.08 (0.04)	0.07	-0.01 (0.04)	0.71	0.20 (0.08)	0.02
FGA x *FGB*	rs2070006 x rs2227395	0.01 (0.03)	0.85	<0.00 1.00 (0.03)	0.94	0.14 (0.07)	0.04
FGA x *FGB*	rs2070009 x rs2227395	<0.001 (0.03)	0.94	<0.00 1.00 (0.03)	0.94	-0.16 (0.07)	0.02
FGA x *FGG*	rs2070022 x rs1049636	-0.08 (0.04)	0.06	0.06 (0.03)	0.07	-0.18 (0.08)	0.03
FGB x *FGG*	rs2227395 x rs2066861	1.00 (0.03)	0.94	0.05 (0.03)	0.12	0.17 (0.07)	0.01
FGA x *FGB*	rs6050 x rs2227395	0.02 (0.03)	0.52	<0.00 1.00 (0.03)	0.94	-0.15 (0.06)	0.02
Non-Hispanic Whites							
FGB x *FGB*	rs1800791 x rs2227395	<0.001 (0.03)	0.89	-0.08 (0.02)	6.7E-4	0.15 (0.06)	0.02
FGA x *FGB*	rs2070006 x rs2227395	0.02 (0.02)	0.20	-0.08 (0.02)	6.7E-4	-0.09 (0.03)	0.005
FGA x *FGB*	rs2070006 x rs4220	0.02 (0.02)	0.20	-0.09 (0.02)	2.1E-3	-0.09 (0.04)	0.01
FGA x *FGB*	rs2070009 x rs2227395	-0.03 (0.02)	0.16	-0.08 (0.02)	6.7E-4	0.09 (0.04)	0.005
FGA x *FGB*	rs2070009 x rs4220	-0.03 (0.02)	0.16	-0.09 (0.02)	2.1E-3	0.09 (0.04)	0.01
FGA x *FGB*	rs2070011 x rs2227395	0.03 (0.02)	0.14	-0.08 (0.02)	6.7E-4	-0.08 (0.04)	0.03
FGA x *FGB*	rs2070011 x rs4220	0.03 (0.02)	0.14	-0.09 (0.02)	2.1E-3	-0.09 (0.04)	0.01
FGA x *FGB*	rs2070022 x rs2227395	-0.03 (0.02)	0.23	-0.08 (0.02)	6.7E-4	0.18 (0.06)	0.002

FGA x *FGB*	rs2070022 x rs4220	-0.03 (0.02)	0.23	-0.09 (0.02)	2.1E-3	0.14 (0.06)	0.02
FGB x *FGG*	rs2227395 x rs2066861	-0.08 (0.02)	6.7 E-4	0.05 (0.02)	0.01	-0.13 (0.04)	0.002
FGB x *FGG*	rs4220 x rs2066861	-0.09 (0.02)	2.1 E-3	-0.03 (0.02)	0.23	-0.08 (0.04)	0.05
FGA x *FGB*	rs6050 x rs2227395	-0.04 (0.02)	0.04	-0.08 (0.02)	6.7E-4	0.13 (0.04)	0.001
FGA x *FGB*	rs6050 x rs4220	-0.04 (0.02)	0.04	-0.09 (0.02)	2.1E-3	0.09 (0.04)	0.02
Mexican Americans							
FGA x *FGB*	rs2070008 x rs1800791	-0.02 (0.04)	0.57	-0.03 (0.03)	0.26	0.18 (0.08)	0.022
FGA x *FGG*	rs2070033 x rs1049636	-0.16 (0.15)	0.26	0.04 (0.03)	0.08	0.59 (0.30)	0.051
FGA x *FGB*	rs2070033 x rs1800791	-0.16 (0.15)	0.26	-0.03 (0.03)	0.26	-0.66 (0.33)	0.048
FGB x *FGG*	rs2227395 x rs1049636	-0.02 (0.04)	0.58	0.04 (0.03)	0.08	0.17 (0.06)	0.006
FGB x *FGG*	rs2227395 x rs1800792	-0.02 (0.04)	0.58	-0.03 (0.03)	0.36	-0.13 (0.06)	0.025
FGB x *FGG*	rs4220 x rs1049636	-0.04 (0.04)	0.36	0.04 (0.03)	0.08	0.14 (0.07)	0.035
FGB x *FGG*	rs4220 x rs1800792	-0.04 (0.04)	0.36	-0.03 (0.03)	0.36	-0.12 (0.06)	0.032

3.2 Gene-environment Interactions

We tested for gene-environment (GxE) interactions with fibrinogen cluster variants and known environmental factors: sex, age, smoking status, and BMI, which were all associated with elevated fibrinogen levels.[8] We tested each environmental variable separately for an association with plasma fibrinogen. We detected 13 significant SNP environment interactions at $p < 0.05$ out of 299 totals test performed.

Over 60% of significant GxE interaction terms were observed in non-Hispanic blacks. Interaction terms with rs2227434 (*FGB*, Pro337Ser) were consistently associated with fibrinogen levels and represented more than 37% (3/8) of the significant associations in non-Hispanic blacks (Table 3). However, this SNP is rare (MAF < 0.05) in all NHANES III populations and did not have significant main effects; thus, these associations are likely false positives (Table 3). We observed four significant GxE interactions in non-Hispanic whites and one in Mexican Americans.

We identified six interaction terms with any given fibrinogen SNP and sex, after excluding rare SNPs rs2227434 and rs6063 (Table 3). All interaction terms identified in non-Hispanic whites were with sex and three were associated with increased levels of fibrinogen (Table 3). There was only one SNP x sex interaction significantly associated with increased fibrinogen levels in non-

Hispanic blacks, rs6058 x sex (Table 3). There were no significant interaction terms with age after excluding rare SNPs that did not have significant main effects, as previously mentioned.

Three SNP x BMI interactions were associated with fibrinogen levels in non-Hispanic blacks or Mexican Americans. The interaction term rs6050 x BMI was significantly associated with increased fibrinogen levels (Table 3). Non-synonymous SNP rs6050 (Thr331Ala) is located in the *FGA* gene and was associated with decreased levels of serum fibrinogen in non-Hispanic whites but not non-Hispanic blacks based on previous single SNP analyses.[17] Another interaction term with BMI, rs2070006 x BMI, was significantly associated with decreased fibrinogen levels (Table 3). Rs2070006 is also located in the *FGA* gene; however this SNP was not significant in any NHANES III population for any single SNP model.[17] There was one GxE interaction term significant in Mexican Americans, rs2006879 x BMI, which was significantly associated with decreased fibrinogen levels, Table 3.

There were two SNP x smoking interaction terms significantly associated with decreased serum fibrinogen, rs2070033 and rs2070008, in non-Hispanic blacks (Table 3). Both SNPs are located in the *FGA* gene and were not associated with fibrinogen levels in previous the single SNP test of association.[17]

Table 3. Gene-environment interactions with fibrinogen variants and age, sex, BMI, and smoking status. Using linear regression we added the multiplicative terms (SNP x environmental factor) to the regression model while simultaneously adjusting for covariates: age, sex, BMI, and smoking status. Below are the association results for the single SNP or environmental factor tests of association as well as the association results for the interaction term. Suppressed = results not released by CDC because counts were < 5 (see 'Statistical Methods').

Interaction	Gene	SNP Effect		Environment Effect		Interaction Effect	
		Beta (SE)	P	Beta (SE)	P	Beta (SE)	P
Non-Hispanic Blacks							
rs2227434 x age	*FGB*	0.26 (0.49)	0.59	0.008 (0.002)	3.76E-07	0.004 (<0.01)	3.76E-07
rs2227434 x BMI	*FGB*	0.26 (0.49)	0.59	0.02 (0.003)	4.32E-06	0.01 (<0.01)	4.32E-06
rs6050 x BMI	*FGA*	0.01 (0.03)	0.52	0.01 (0.003)	5.87E-06	0.01 (0.05)	8.10E-03
rs2227434 x sex	*FGB*	0.26 (0.49)	0.59	0.09 (0.04)	0.02	0.05 (0.02)	0.02
rs6058 x sex	*FGB*	0.06 (0.05)	0.18	0.10 (0.04)	0.01	0.22 (0.10)	0.03
rs2070033 x smoking	*FGA*	0.07 (0.05)	0.16	-0.06 (0.04)	0.14	-0.22 (0.10)	0.03
rs2070006 x BMI	*FGA*	0.005 (0.03)	0.85	0.02 (0.003)	6.94E-06	-0.01 (<0.01)	0.03
rs2070008 x smoking	*FGA*	-0.01 (0.04)	0.75	-0.05 (0.04)	-0.16	-0.19 (0.01)	0.04
Non-Hispanic Whites							
rs2070017 x sex	*FGA*	0.24 (0.22)	0.28	0.07 (0.03)	0.01	0.03	0.01
rs2066861 x sex	*FGG*	0.05 (0.02)	0.01	0.06 (0.03)	0.02	-0.08	0.05
rs2070009 x sex	*FGA*	-0.03 (0.02)	0.16	0.06 (0.03)	0.02	0.07	0.05
rs6050 x sex	*FGA*	-0.04 (0.02)	0.04	0.06 (0.03)	0.04	0.08	0.05
Mexican Americans							
rs2066879 x BMI	*FGG*	-0.37 (0.24)	0.12	0.02 (0.004)	3.1E-5	-0.28	0.03

4. Discussion

In the present study, we performed an exhaustive analysis to detect gene-gene and gene-environment interactions that may affect fibrinogen levels in NHANES III. Identifying genetic modifiers of fibrinogen levels is important since the genetic contribution of variable fibrinogen

levels at the population level is poorly understood. Additionally, current heritability estimates do not include possible gene-environment interactions. Here using a multiplicative approach, we tested for statistical epistasis or the deviation from linearity in the presence and absence of main effects. We tested for interactions amongst 25 tagSNPs in the fibrinogen gene cluster (gene-gene) as well as tested each SNP with environmental risk factors: age, sex, BMI and smoking status (gene-environment). We observed 19 gene-gene interactions and 13 gene-environment interactions with an empirical p-value < 0.05.

One criticism of exhaustively testing for interactions is the biological interpretation of the results. The fibrinogen gene cluster is an ideal candidate for identifying gene-gene interactions since each gene is dependent on the other to produce active fibrin in the blood.[5; 6] In this gene-gene analysis, we detected several interactions between different genes in the fibrinogen cluster (Table 2, last column). More importantly, we identified several interactions with rs2227395, which is in the *FGB* gene, also known as the rate-limiting gene in fibrinogen production. It is important to note that this SNP did not have a significant main effect in the single SNP test of association (Table 2, first column).[17] Interestingly, all gene-gene interactions we identified were observed in the absence of main effects (Table 2, first column). It is conceivable that SNPs such as rs2227395, where we did not observe a significant main effect, are interacting with other fibrinogen SNPs and ultimately affecting fibrinogen levels, an effect that would not be detected at the single SNP level.

Environmental variables are well-known risk factors of elevated fibrinogen levels.[8] We identified several gene-environment interactions that were associated with fibrinogen levels across multiple populations (Table 3). We observed consistent associations with interaction terms containing rs2227434 or rs6063 and environmental factors age, sex and BMI in non-Hispanic blacks (Table 3). More often than not, consistent associations among different SNPs are indicative of high correlation or linkage disequilibrium between SNPs. We attempted to calculate linkage disequilibrium between these SNPs in our study population but failed to obtain accurate results due to low genotype counts for both SNPs. These SNPs were selected based on frequency and biological relevance for two populations as previously mentioned. However, due to poor genotyping efficiency, we have low genotype counts for these SNPs. We have added this statement to the discussion section. We then attempted to use reference HapMap population YRI to calculate linkage disequilibrium between these SNPs but like our study population, genotype frequencies were too low for an accurate measurement.[18] As previously mentioned, both SNPs are rare in non-Hispanic blacks with minor allele frequencies less 0.001 and we only had 5% power to detect these interactions in our study population. After evaluating the main effect of the environmental factors we noticed that the association results for these factors and the interaction term with rs2227434 or rs6063 were identical as well (Table 3). Given the low minor allele frequencies, it is likely these interaction terms are false positives and actually represent the environmental factors alone. For these reasons we caution the interpretation of these results. In contrast, we identified three SNP x BMI interactions both of which had significant environmental main effects as well as two SNP x smoking interactions and five SNP x sex interactions.

While NHANES III contains over 33,000 study participants, our study population sample size was limited for several reasons. The Centers for Disease Control and Prevention (CDC) only collected DNA samples from a subset of the entire dataset, which includes 7,159 participants from phase 2 of NHANES III. DNA was not collected from participants who reported having hemophilia or chemotherapy within four weeks of collection. [19] Our outcome variable for this study was limited to plasma fibrinogen, which was only collected on a subset of NHANES participants. Plasma fibrinogen levels in NHANES III was only measured for participants > 40 years of age[19], which drastically reduced our sample size. Due to sample size, on average we were only 70% powered to detect effects that explain less < 0.5% of the trait variability at $\alpha = 0.05$ and minor allele frequency of at least 0.05. One caveat of exhaustively testing all pair-wise interactions is the increase in type I error or false positives. It is important to note for this study that we did not correct for multiple testing, which increases the likelihood of reporting false positives, thereby reducing the confidence of our reported findings.

Despite the small sample sizes, we identified several gene-gene and gene-environment associations at a liberal significance threshold of $p<0.05$. To date this is the first study that examines the role genetic modifiers has on plasma fibrinogen levels in admixed populations such as non-Hispanic blacks and Mexican Americans. Despite the statistical evidence of epistasis, we cannot confirm the effect these interactions have on fibrinogen levels from a biological standpoint and future analyses are warranted. This work serves as first step in uncovering genetic epistasis for complex diseases such as cardiovascular disease that is mediated to an extent by fibrinogen levels. Furthermore, since the majority of our findings are in admixed populations this works highlights the importance and need for conducting genetic association studies in these high-risk populations for discovery.

REFERENCES

1. Collet, J.P., Soria, J., Mirshahi, M., Hirsch, M., Dagonnet, F.B., Caen, J., and Soria, C. (1993/10/15). Dusart syndrome: a new concept of the relationship between fibrin clot architecture and fibrin clot degradability: hypofibrinolysis related to an abnormal clot structure. Blood 82, 2462-2469.
2. Voetsch, B., and Loscalzo.J. (2003). Genetic Determinants of Arterial Thrombosis. Arterioscler Thromb Vasc Biol 24, 216-229.
3. de Maat, M.P. (2005/3). Genetic variation in the fibrinogen gene cluster. ThrombHaemost 93, 401-402.
4. Mosesson, M.W. (1998). Fibrinogen structure and fibrin clot assembly. SeminThrombHemost 24, 169-174.
5. Zhang, J.Z., and Redman, C.M. (1994/1/7). Role of interchain disulfide bonds on the assembly and secretion of human fibrinogen. JBiolChem 269, 652-658.
6. Zhang, J.Z., and Redman, C. (1996/11/22). Fibrinogen assembly and secretion. Role of intrachain disulfide loops. JBiolChem 271, 30083-30088.
7. Lim, B.C., Ariens, R.A., Carter, A.M., Weisel, J.W., and Grant, P.J. (2003/4/26). Genetic regulation of fibrin structure and function: complex gene-environment interactions may modulate vascular risk. Lancet 361, 1424-1431.

8. de Maat, M.P. (2001). Effects of diet, drugs, and genes on plasma fibrinogen levels. AnnNYAcadSci 936, 509-521.
9. Matise, T.C., Ambite, J.L., Buyske, S., Carlson, C.S., Cole, S.A., Crawford, D.C., Haiman, C.A., Heiss, G., Kooperberg, C., Marchand, L.L., et al. (2011/10/1). The Next PAGE in understanding complex traits: design for the analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. AmJ Epidemiol 174, 849-859.
10. CDC. (1996). Centers for Disease Control and Prevention Third National Health and Nutrition Examination Survey, 1988-94, Plan and Operations Procedures Manual. In., pp -.
11. CDC. (2004). Centers for Disease Control and Prevention Plan and Operation of the Third National Health and Nutrition Examination Survey, 1988-94. In. (Bethesda, MD), pp -.
12. Daum, P., Estergreen, J., and Wener.M. (2000). Laboratory Procedure Manual:Rate of Clot Formation on the STA-Compact. In., pp 1-12.
13. Steinberg, K.K., Sanderlin, K.C., Ou, C.Y., Hannon, W.H., McQuillan, G.M., and Sampson, E.J. (1997). DNA banking in epidemiologic studies. EpidemiolRev 19, 156-162.
14. Carlson, C.S., Eberle, M.A., Rieder, M.J., Yi, Q., Kruglyak, L., and Nickerson, D.A. (2004/1). Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. AmJ HumGenet 74, 106-120.
15. Crawford, D.C., Akey, D.T., and Nickerson, D.A. (2005). The patterns of natural variation in human genes. AnnuRevGenomics HumGenet 6, 287-312.
16. Grady, B.J., Torstenson, E., Dudek, S.M., Giles, J., Sexton, D., and Ritchie, M.D. (2010). Finding unique filter sets in plato: a precursor to efficient interaction analysis in gwas data. PacSympBiocomput, 315-326.
17. Jeff, J.M., Brown-Gentry, K., and Crawford, D.C. (2012). Replication and characterisation of genetic variants in the fibrinogen gene cluster with plasma fibrinogen levels and haematological traits in the Third National Health and Nutrition Examination Survey. Thrombosis and haemostasis 107, 458-467.
18. Project, I.H. (2003/12/18). The International HapMap Project. Nature 426, 789-796.
19. W, E., Lewis, B., and Koncikowski, S. (1996). Laboratory Procedures Used for the Third National Health and Nutrition Examination Survey (NHANES III), 1988-1994. In., pp 537-546.