

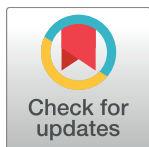
RESEARCH ARTICLE

A genome-wide trans-ethnic interaction study links the *PIGR-FCAMR* locus to coronary atherosclerosis via interactions between genetic variants and residential exposure to traffic

Cavin K. Ward-Caviness^{1,2,3}, Lucas M. Neas³, Colette Blach¹, Carol S. Haynes¹, Karen LaRocque-Abramson¹, Elizabeth Grass¹, Z. Elaine Dowdy¹, Robert B. Devlin³, David Diaz-Sanchez³, Wayne E. Cascio³, Marie Lynn Miranda⁴, Simon G. Gregory¹, Svati H. Shah^{1,5}, William E. Kraus^{1,5}, Elizabeth R. Hauser^{1,6,7*}

1 Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC, United States of America, **2** Institute of Epidemiology II, Helmholtz Zentrum München, Neuherberg, Germany, **3** National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, Chapel Hill, NC, United States of America, **4** National Center for Geospatial Medicine, Rice University, Houston, TX, United States of America, **5** Division of Cardiology, Duke University School of Medicine, Durham, NC, United States of America, **6** Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC, United States of America, **7** Cooperative Studies Program Epidemiology Center-Durham, Veterans Affairs Medical Center, Durham, NC, United States of America

* elizabeth.hauser@duke.edu



OPEN ACCESS

Citation: Ward-Caviness CK, Neas LM, Blach C, Haynes CS, LaRocque-Abramson K, Grass E, et al. (2017) A genome-wide trans-ethnic interaction study links the *PIGR-FCAMR* locus to coronary atherosclerosis via interactions between genetic variants and residential exposure to traffic. PLoS ONE 12(3): e0173880. <https://doi.org/10.1371/journal.pone.0173880>

Editor: Andreas Zirlik, Universitätsklinikum Freiburg, GERMANY

Received: September 25, 2016

Accepted: February 28, 2017

Published: March 29, 2017

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: Due to privacy restrictions, access to the residential data used in this article is restricted. The data (like all other CATHGEN data) may be requested from the Dr. William Kraus of the CATHGEN Steering Committee, and access is provided readily upon request. The link to get a form to submit a request is https://dmpi.duke.edu/files/cathgen_project_application.pdf.

Abstract

Air pollution is a worldwide contributor to cardiovascular disease mortality and morbidity. Traffic-related air pollution is a widespread environmental exposure and is associated with multiple cardiovascular outcomes such as coronary atherosclerosis, peripheral arterial disease, and myocardial infarction. Despite the recognition of the importance of both genetic and environmental exposures to the pathogenesis of cardiovascular disease, studies of how these two contributors operate jointly are rare. We performed a genome-wide interaction study (GWIS) to examine gene-traffic exposure interactions associated with coronary atherosclerosis. Using race-stratified cohorts of 538 African-Americans (AA) and 1562 European-Americans (EA) from a cardiac catheterization cohort (CATHGEN), we identify gene-by-traffic exposure interactions associated with the number of significantly diseased coronary vessels as a measure of chronic atherosclerosis. We found five suggestive ($P < 1 \times 10^{-5}$) interactions in the AA GWIS, of which two (rs1856746 and rs2791713) replicated in the EA cohort ($P < 0.05$). Both SNPs are in the *PIGR-FCAMR* locus and are eQTLs in lymphocytes. The protein products of both *PIGR* and *FCAMR* are implicated in inflammatory processes. In the EA GWIS, there were three suggestive interactions; none of these replicated in the AA GWIS. All three were intergenic; the most significant interaction was in a regulatory region associated with *SAMSN1*, a gene previously associated with atherosclerosis and B cell activation. In conclusion, we have uncovered several novel genes associated with coronary atherosclerosis in individuals chronically exposed to increased ambient concentrations

Funding: This work was supported by the following grants from the National Institutes of Health (nih.gov): HL095987, HL10621, and HL36587. This project was supported in part by an appointment to the Research Participation Program at the Office of Research and Development, U.S. Environmental Protection Agency (<http://orise.orau.gov/epa/applicants/current-research-opportunities.aspx>). This work was supported in part by a grant from the Health Effects Institute (healtheffects.org) award number R-82811201. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

of traffic air pollution. These genes point towards inflammatory pathways that may modify the effects of air pollution on cardiovascular disease risk.

Introduction

Ischemic heart disease and stroke are responsible for more deaths worldwide than any other singular communicable or non-communicable cause [1]. The hallmark of ischemic heart disease coronary artery disease (CAD) is atherosclerotic lesions forming in the coronary arteries, eventually rupturing and releasing prothrombotic factors. The final result is an occlusive thrombus limiting blood flow and leading to a myocardial infarction [2]. Large scale genome-wide association studies have repeatedly identified genetic variants associated with CAD [3,4]. Though CAD is highly heritable, known genetic variants explain relatively little of the heritability, an observation possibly explained by as yet unreported interactions between genetic variants and environmental exposures [5].

Air pollution is a heterogeneous mixture of suspended particles and volatile organic and inorganic compounds, and the most widespread environmental exposure associated with CAD [6–8], with an estimated 22% of cardiovascular disease morbidity and mortality attributable to air pollution [9]. Air pollution is generated by a variety of industrial, agricultural, and residential/traffic sources. Vehicular traffic is one of the most common sources of air pollution; it is generated by tailpipe emissions, re-suspended road dust, and evaporative emissions. Exposure to traffic-related air pollution constituents and roadway noise is strongly associated with CAD [10–12]. The distance between the primary residence and nearest major roadway is a commonly used measure of traffic-related air pollution exposure that is closely correlated with air pollution emissions from traffic [13,14]. Distance to major roadways is associated with coronary heart disease [15–17], atherosclerosis [18], peripheral artery disease (PAD) [19], obesity [20], and metabolic risk factors for cardiovascular disease [21].

Gene-environment interactions may play a large role in cardiovascular disease [22]. However, studies to discover interactions between genes and air pollution have been mostly restricted to targeted studies of detoxification or inflammation-pathway associated genes [23]. While these candidate gene studies have been informative, they lack the comprehensive assessment available in genome-wide interaction studies (GWIS). A recent GWIS for gene-air pollution interactions identified genetic variants in the bone morphogenetic protein gene family associated with PAD [24]. A smoking exposure GWIS identified several novel genetic variants associated with coronary artery calcification; these include genetic variants in *WWOX*, which are also associated with PAD via an interactions with smoking and traffic air pollution exposure [24,25].

To identify novel genes associated with coronary atherosclerosis in a traffic exposure GWIS, we used the CATHeterization GENetics (CATHGEN) cohort [26], which has proven to be a successful approach to study both cardiovascular disease and environmental exposures [21,24,27,28]. For our analyses we utilized a race-stratified study design, and used distance from the primary residence to the nearest major roadway as a measure of traffic-related air pollution exposure (traffic exposure).

Materials and methods

Study population

The CATHGEN cardiac catheterization cohort is composed of 9,334 individuals recruited through the Duke University cardiac catheterization laboratory from 2001–2011. Patients

NC CATHGEN Participants

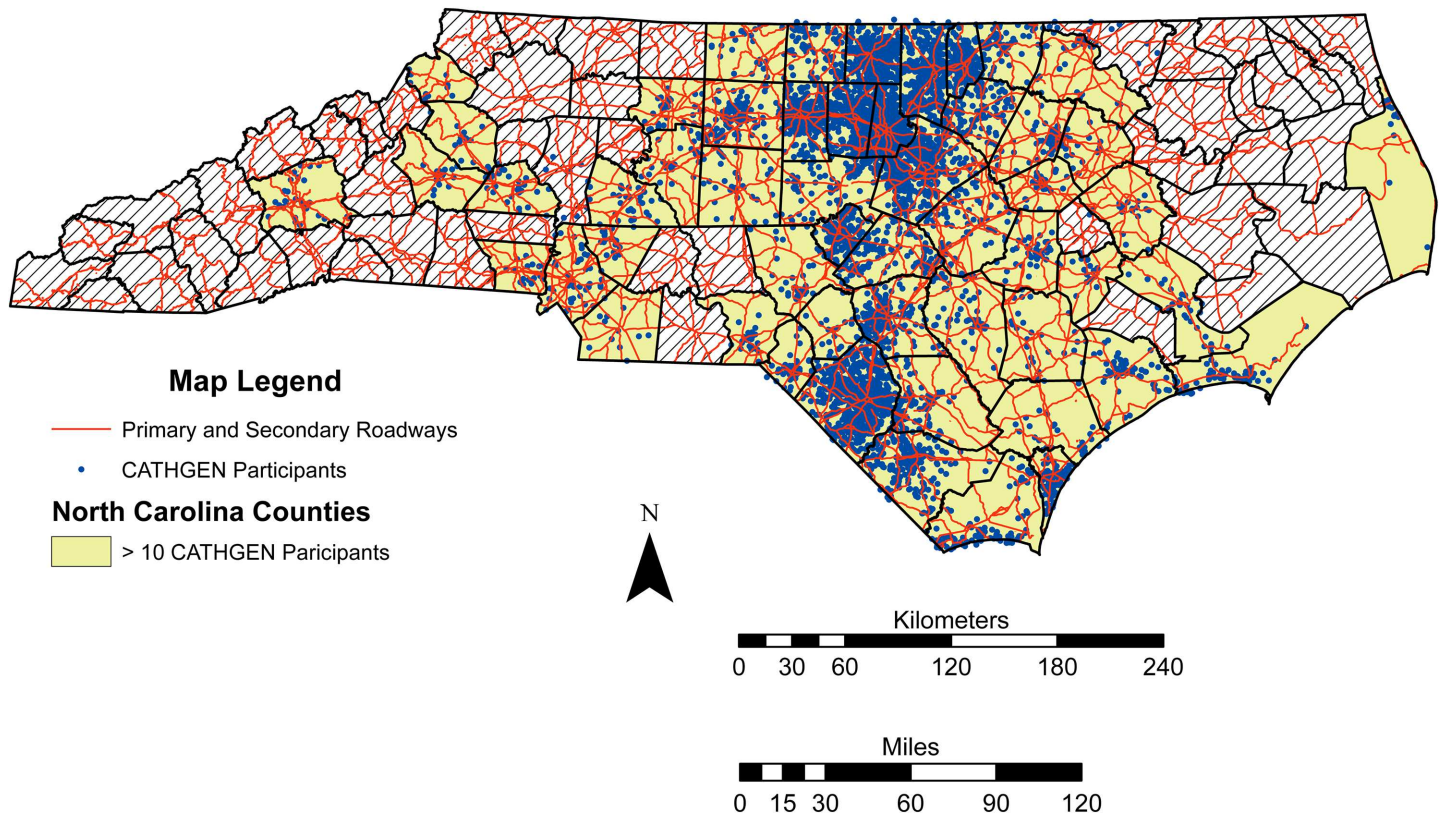


Fig 1. NC CATHGEN Participants. The distribution of CATHGEN participants overlaid with the primary and secondary roadway network (red lines) for North Carolina. To protect patient health information counties with 10 or fewer participants are suppressed and represented by hatch marks. All other counties have the CATHGEN participants locations (blue dots) randomized within a small area to preserve the overall structure and still protect patient health information.

<https://doi.org/10.1371/journal.pone.0173880.g001>

were administered a Health & Physical examination by a trained medical professional as part of the standard intake prior to the catheterization procedure. Demographic characteristics and peripheral blood were also collected prior to the procedure, and the patient’s Duke University medical records were linked to the CATHGEN database. A complete detailed description of CATHGEN is available [26]. Patient collection and data analyses associated with CATHGEN were approved by the Duke University Institutional Review Board.

Residential address information was available for 8,071 CATHGEN participants. To increase homogeneity, we restricted our study population to the 7,158 participants residing in North Carolina (Fig 1), of which 2,100 had genome-wide genotyping data. These 2,100 North Carolina residents with genome-wide genotyping comprised the study cohort.

Clinical information

All clinical information was obtained from a pre-procedure interview and Duke University medical records incorporated into the CATHGEN database via the PEDIGENE[®] system. Clinical variables were age, race, sex, smoking, BMI, presence of hypertension, hyperlipidemia, and type 2 diabetes (diabetes). Race was a self-reported variable used to stratify the cohort into self-identified African-Americans (AA) and European-Americans (EA). Hypertension,

Table 1. GWIS cohort demographic and clinical covariates.

GWIS Cohort Clinical Covariates	AA (N = 538)	EA (N = 1562)
Age (years)	56.3 (11.3)	61.2 (11.9)
BMI (kg/m ²)	32.1 (8.01)	29.7 (7.02)
Distance to Roadways (km)	0.78 (0.71)	0.94 (0.81)
Sex (male)	288 (53.5)	619 (39.6)
Smoking (0 = never, 1 = ever)	231 (42.9)	789 (50.5)
Type 2 Diabetes (0 = no, 1 = yes)	232 (43.1)	415 (26.6)
Hypertension (0 = no, 1 = yes)	434 (80.7)	1008 (64.5)
Hyperlipidemia (0 = no, 1 = yes)	291 (54.1)	953 (61.0)
Diseased Vessels (0)	294 (54.6)	639 (40.9)
Diseased Vessels (1)	96 (17.8)	325 (20.8)
Diseased Vessels (2)	58 (10.8)	274 (17.5)
Diseased Vessels (3)	90 (16.7)	324 (20.7)

Age, BMI, and traffic exposure (Distance to roadway) are given as the mean (SD). Binary and categorical variables are given as the N (%). BMI = body mass index; Diseased vessels = number of diseased coronary vessels.

<https://doi.org/10.1371/journal.pone.0173880.t001>

hyperlipidemia, and diabetes were all defined clinically by the operating team from pharmacologic treatment lists, and from clinical chemistries in the medical record or obtained prior to catheterization (Table 1). The coronary atherosclerosis outcome was defined as the number of diseased coronary vessels (number of diseased vessels, NUMDZV) at the catheterization. NUMDZV is a clinician annotated clinical variable that describes the number (0–3) of vessels with significant (> 75%) blockage while taking into account the location of the vessels as well as the left or right dominance of the patient. Disease in the left main carotid artery of greater than 50% is considered two-vessel disease, incrementing the NUMDZV variable by 2. Thus, NUMDZV is a measure of the degree of clinically significant coronary atherosclerosis.

Exposure assessment

Traffic-related air pollution exposure (traffic exposure) was indexed by the perpendicular distance between the patient’s residence at time of the catheterization and the nearest major roadway. Full details of the creation and use of this traffic exposure index in the CATHGEN cohort have been previously published [21] including in the context of a GWIS [24]. We defined major roadways as primary and secondary roadways. Primary roadways were major highways often distinguished by the presence of interchanges while secondary roadways were multi-lane intra and inter-city arterials. These roadways were defined by the North Carolina Department of Transportation [29], and the definition is consistent with the Master Address File/Topologically Integrated Geographic Encoding and Referencing Feature Class Code used by the U.S. Census Bureau [30]. The patient locations and North Carolina map of primary and secondary roadways were imported into ArcGIS [31] to calculate the perpendicular distance between primary residence and the nearest primary or secondary roadway. The mean distance was 0.78 km (SD = 0.71, inter-quartile range (IQR) = 0.94) in the AA subgroup and 0.94 km (SD = 0.82, IQR = 1.15) in the EA subgroup (Table 1).

Genotyping

A total of 3,512 CATHGEN participants were genotyped on the Illumina HumanOmni 1-Quad_v1-0_C array and 2,100 of them resided within North Carolina. Of these 2,100

genotyped individuals 1562 were European-American and 538 were African-American. These race-stratified cohorts and the quality control procedures have been previously described [24,32]. The quality control consisted of removing gender mismatches, variants with low (<98%) call frequency, individuals with a low (<98%) call rate, and related individuals. Analyses were restricted to variants with a minor allele frequency (MAF) greater than 0.05.

Statistical methods

For the race-stratified GWIS, to properly account for the ordinal nature of the number of diseased vessels outcome, a cumulative link model as implemented in the *ordinal* package [33] in R [34] was used. This model is equivalent to a proportional odds model. We adjusted for age, sex, BMI, hypertension, smoking, hyperlipidemia, and diabetes in all analyses. Additionally, Eigenstrat-calculated principal components [35] were used to adjust for racial substructure within each of the race-specific GWIS. As in previous analyses four principal components were used for the EA subgroup; two principal components were sufficient to adjust for racial substructure in the AA subgroup [24,32]. The number of principal components was based on the adjustment necessary to remove racial substructure in previous race-stratified, genome-wide analyses in CATHGEN [24,32]. Results from the cumulative link model are reported in terms of an odds ratio which signifies the increased risk due to incrementing the number of diseased vessels by one. For the interaction term, the traffic exposure was scaled to the interquartile range. All P-values represent a test of the null hypothesis that the regression coefficient for the interaction term equals 0.

A trans-ethnic meta-analysis was performed in the METAL software package [36] using a P-value-based meta-analysis accounting for effect direction, allele and weighted by sample size. As there were multiple instances of different minor alleles for the race-stratified cohorts, it was essential to properly account for the effect allele and differential sample sizes to insure consistent meta-analysis results.

Given the often low power of interaction studies [37,38], in addition to examining any genome-wide significant associations ($P < 5 \times 10^{-8}$) we used a suggestive cutoff of $P < 1 \times 10^{-5}$ in a single subgroup for the initial analyses. This suggestive cutoff has been used in previous GWIS [24,25] to determine interaction associations worthy of being examined for replication in independent cohorts. We used $P < 0.05$ as the cutoff for replication in our race-stratified cohorts. For the trans-ethnic meta-analysis, $P < 1 \times 10^{-5}$ was used to indicate suggestive association in the full dataset.

Functional relevance of each single nucleotide polymorphism (SNP) involved in a suggestive interaction with traffic exposure was assessed using multiple publically available databases. To identify variants with potential regulatory or epigenetic implications (e.g., CpG sites), the sequence surrounding each SNP was examined using the NCBI dbSNP database [39]. We identified potential regulatory intergenic variants by integrating data on open chromatin regions, defined as DNaseI hypersensitivity sites, based on data from the ENCODE project [40]. In addition, recently published results on allele-specific DNA openness [41] were used to identify variants in regions related to transcription factor binding and/or nucleosome positioning. Finally, we investigated the evidence that suggestive variants may regulate gene expression by integrating data on tissue-specific expression quantitative trait loci (eQTLs) via Genotype Tissue Expression (GTEx) database Release V6 [42,43].

Results

The study population for this analysis consisted of 2,100 individuals, residing in North Carolina. The geographic distribution of the study cohort is given in Fig 1 while the clinical

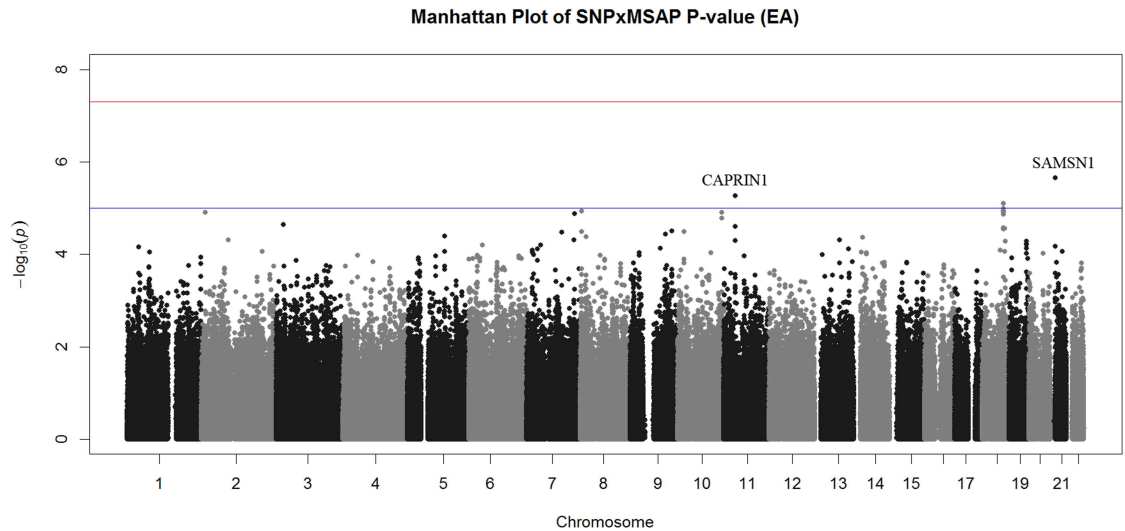


Fig 2. Manhattan Plot of Interaction P-values for EA GWIS. Manhattan plot of interaction P-values for the EA GWIS. Though no interactions achieved genome-wide significance (red line) interactions on chromosome 11 (*CAPRIN1*), 18, and 21 (*SAMS1* open chromatin region) exceeded the suggestive significance threshold (blue line).

<https://doi.org/10.1371/journal.pone.0173880.g002>

covariates are given in [Table 1](#). A total of 756,588 SNPs with a MAF > 0.05 converged in the AA model and a total of 674,795 SNPs with MAF > 0.05 a converged in the EA model. The IQR for the traffic exposure variable for the AA and EA GWIS was 0.94 km and 1.15 km respectively.

EA results

Three interactions exceeded the threshold for suggestive evidence of association ($P < 1 \times 10^{-5}$) ([Fig 2](#), [Table 2A](#)). All of the suggestive EA interactions were in intergenic regions. The most significant association was rs2822693 ($P = 2.2 \times 10^{-6}$), an intergenic SNP on chromosome 21

Table 2. GWIS interactions with $P < 1 \times 10^{-5}$.

a. EA GWIS interaction $P < 1 \times 10^{-5}$						
SNP	Locus	AA OR (CI)	AA P	EA OR	EA P	Minor Allele (AA/EA)
rs12285326	<i>SAMS1</i> (DHS)	1.42 (0.77, 1.72)	0.11	0.61 (0.32, 0.62)	5.4×10^{-6}	A/A
rs2822693	<i>CAPRIN1</i> (Intergenic)	1.15 (0.92, 2.19)	0.49	0.45 (0.49, 0.76)	2.3×10^{-6}	A/A
rs1108775	Intergenic	1.00 (0.72, 1.39)	0.99	1.55 (1.28, 1.87)	7.9×10^{-6}	G/A
b. AA GWIS interaction $P < 1 \times 10^{-5}$						
SNP	Locus	AA OR (CI)	AA P	EA OR	EA P	Minor Allele (AA/EA)
rs1856746	<i>FCAMR</i> (Intron)	2.96 (1.89, 4.64)	2.3×10^{-6}	0.83 (0.68, 0.99)	0.048	A/G
rs2791713	<i>FCAMR-PIGR</i> (Intergenic)	3.00 (1.90, 4.73)	2.4×10^{-6}	0.80 (0.66, 0.97)	0.02	G/A
rs291096	<i>PIGR</i> (V183V)	2.97 (1.88, 4.67)	2.7×10^{-6}	0.85 (0.77, 1.02)	0.09	A/G
rs17366136	Intergenic	2.47 (1.66, 3.68)	8.1×10^{-6}	0.86 (0.71, 1.04)	0.12	A/A
rs11012265	Intergenic	0.46 (0.33, 0.65)	9.1×10^{-6}	1.01 (0.83, 1.21)	0.95	A/A

We observed three suggestive interactions in the EA GWIS (**a**) and five suggestive interactions in the AA GWIS (**b**) two of which replicated: rs1856746 and rs2791713. Rs1108775, rs17366136 and rs11012265 were all located in intergenic regions not near any annotated gene. Genotypes were coded according to the number of copies of the minor allele. BP = base pair location; CI = 95% confidence interval; DHS = DNase I hypersensitivity site; OR = odds ratio for interaction from cumulative link model; SNP = SNP involved in the SNP-traffic exposure interaction

<https://doi.org/10.1371/journal.pone.0173880.t002>

SAMSN1 Locus – EA

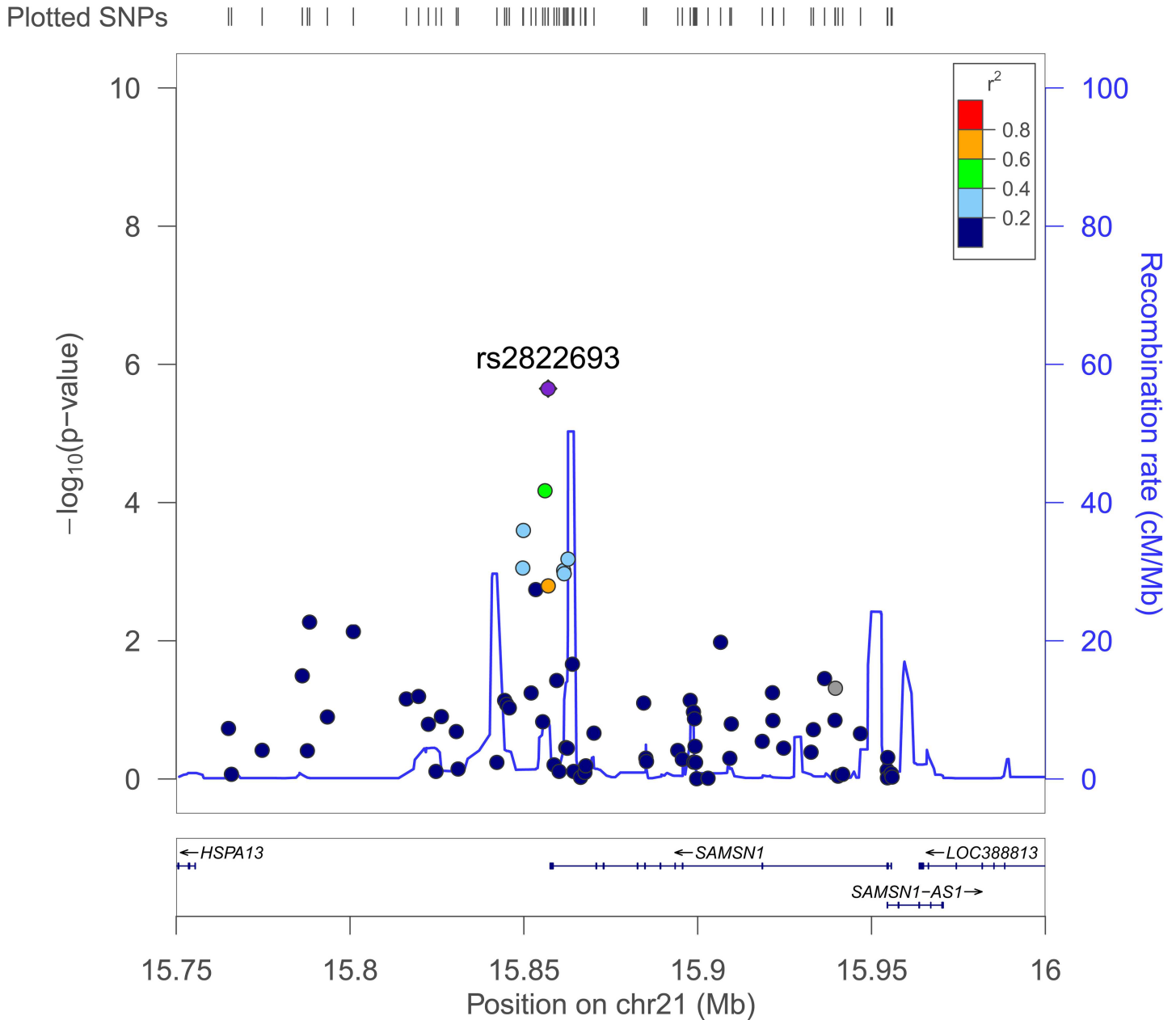


Fig 3. LocusZoom Plot of Open Chromatin Region Downstream of *SAMSN1* for EA GWIS. Overall the LD across the region is modest to low supporting the hypothesis that the interaction association signal is coming from rs2822693 or an untyped variant in high LD with it, likely restricted to the region between the two recombination peaks. For plotting 1000 Genomes data was used (population = EUR) in genome build hg19.

<https://doi.org/10.1371/journal.pone.0173880.g003>

located near *SAMSN1*. This region contained relatively few typed variants and the most significant variants were those that were in LD with rs2822693 (Fig 3). Rs2822693 was located in a DNase I hypersensitivity site present in hematopoietic stem cells. *SAMSN1* expression was highest in hematopoietic stem cells and the DNase I hypersensitivity site containing rs2822693 was highly correlated with *SAMSN1* expression (dnase.genome.duke.edu [40]). Rs12285326

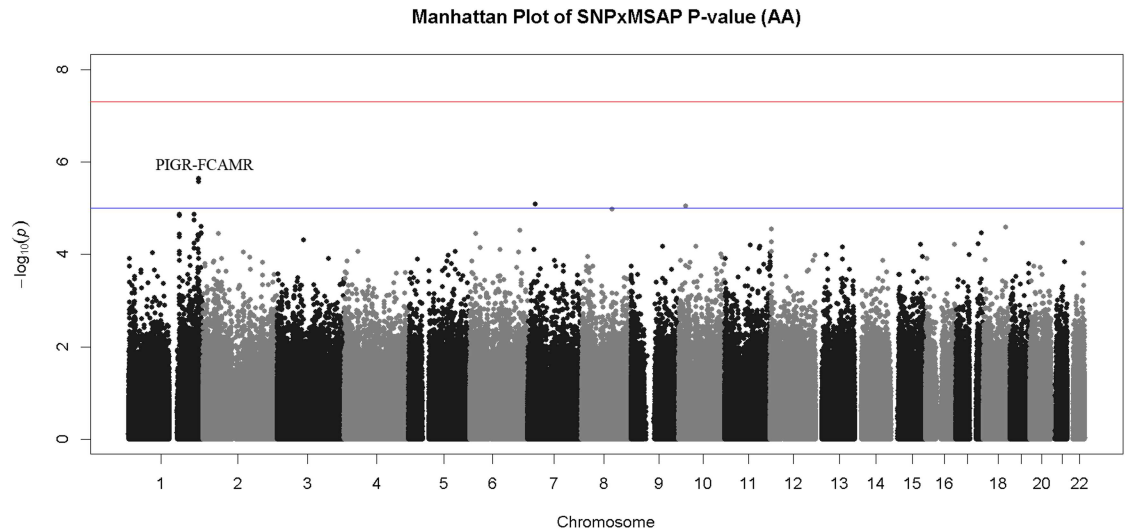


Fig 4. Manhattan Plot of Interaction P-values for AA GWIS. Though no interactions exceeded the genome-wide significance threshold a cluster of SNP-by-TRAP interactions on chromosome 1 (*FCAMR-PIGR* locus) exceeded the threshold for suggestive evidence of association ($P < 1 \times 10^{-5}$, blue line).

<https://doi.org/10.1371/journal.pone.0173880.g004>

was located near *CAPRINI* on chromosome 21 however was not in a DHS site or an eQTL for this gene.

AA results

A total of five interactions had a $P < 1 \times 10^{-5}$ in the AA GWIS (**Table 2B**, **Fig 4**). The *FCAMR-PIGR* locus on chromosome 1 accounted for the three most significant interactions in the AA GWIS: rs1856746 ($P = 2.3 \times 10^{-6}$) is in an intron of *FCAMR*; rs2791713 ($P = 2.4 \times 10^{-6}$) is located in the intergenic region between *FCAMR* and *PIGR*; and rs291096 is a synonymous variant in *PIGR* ($P = 2.7 \times 10^{-6}$, V183V). Both of the remaining suggestive interactions were with intergenic variants (**Table 2B**). Rs1856746 and rs2791713 both replicated in the EA GWIS. The minor allele was flipped between the EA and AA cohorts for rs1856746 and rs2791713. The coding variant rs291096 had an interaction $P = 0.09$ in the EA cohort, narrowly missing the threshold for replication. None of the suggestive interactions from the EA cohort replicated (**Table 2**).

Interactions in the *FCAMR-PIGR* locus

We examine all genotyped SNPs in the *FCAMR-PIGR* locus, defined as chromosome 1 base pairs 207101866–207144307 (genome build GRCh37). There were five variants in addition to rs1856746 in *FCAMR*, one of which, rs11119925, had a nominally significant interaction P in the AA and EA GWIS (AA $P = 0.019$, EA $P = 0.034$, **Table A(a) in S1 File**). There were 11 variants in *PIGR* represented in our GWIS; In addition to the synonymous variant rs291096, five additional SNPs had interactions with $P < 0.05$ in the AA GWIS: two were intronic variants, two synonymous variants, and one a missense variant (rs29110, A \rightarrow V, AA $P = 1.2 \times 10^{-4}$, **Table A(b) in S1 File**). The LD in African ancestry individuals amongst the coding variants in the *PIGR-FCAMR* locus was low ($r^2 < 0.40$); however, there was substantial LD ($r^2 > 0.60$) between the most significant *PIGR* and *FCAMR* variants (**Fig 5**). The observed LD for *PIGR* and *FCAMR* variants was reduced in European ancestry individuals but still substantial (**Fig 6**).

Meta-analysis

In a sample size weighted meta-analysis combining the EA and AA cohorts, the most significant interaction was with rs10830090 on chromosome 10, located in an intron of *RARS2*, a mitochondrial arginine t-RNA synthase (Table B in S1 File). Mutations in *RARS2* cause a neurodegenerative disorder, pontocerebellar hypoplasia type 6 [44]. Mitochondrial function is linked to both air pollution and cardiovascular disease as ultra-fine particulates can cause

FCAMR-PIGR Locus – AA

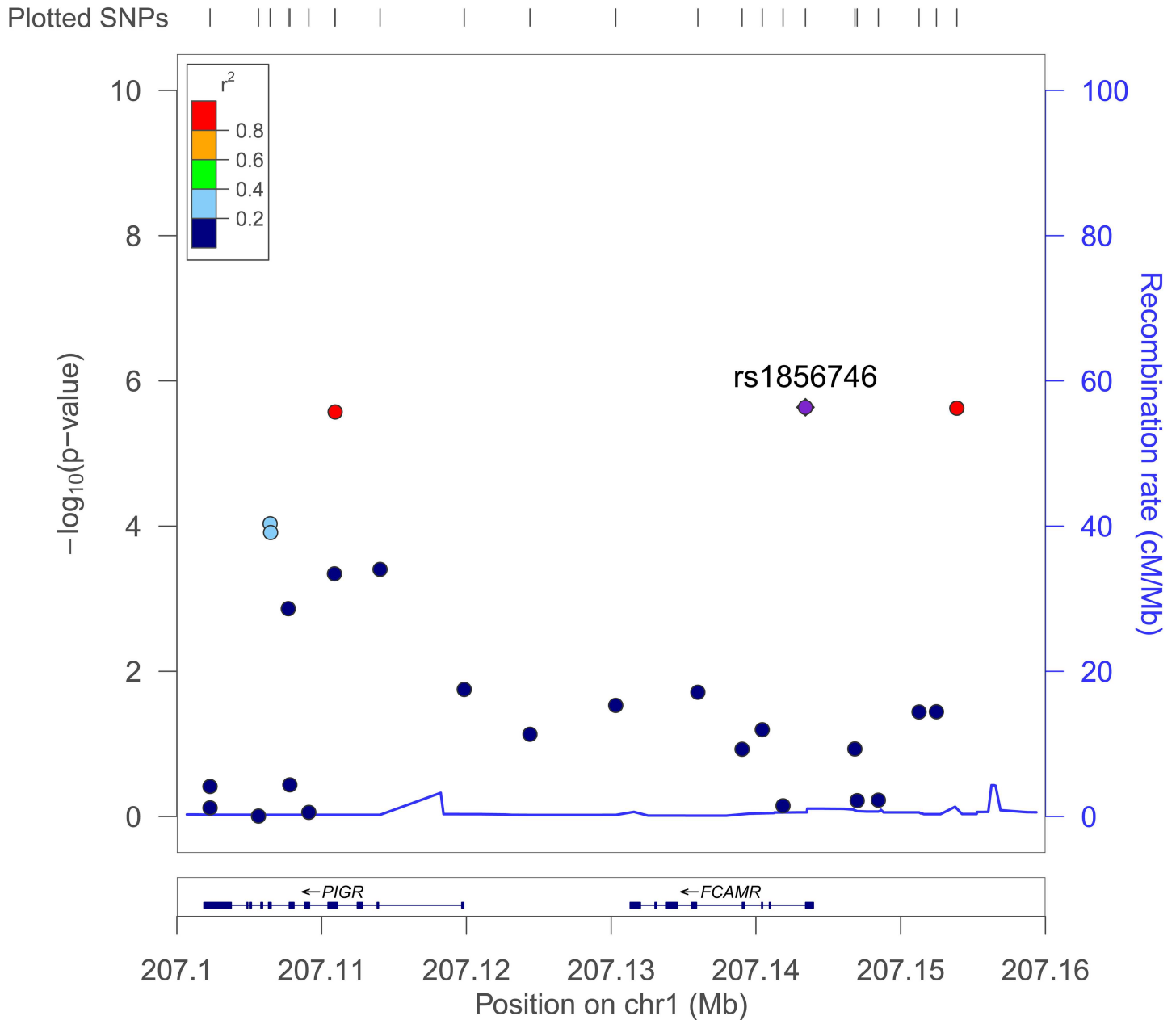


Fig 5. LocusZoom Plot of *FCAMR-PIGR* locus for African-Americans. LD is calculated relative to the most significant variant rs1856746. LD across the three most significant variants within the region is substantial however all other variants in the region show little correlation with rs291096 despite the relatively low recombination rate across the region. For plotting 1000 Genomes data is used (population AFR) in genome build hg19.

<https://doi.org/10.1371/journal.pone.0173880.g005>

mitochondrial damage [45] linked to cardiovascular disease [46]. All of the other suggestive meta-analysis interactions were intergenic.

Integration of suggestive associations with multi-tissue eQTL data

Investigation of the suggestive AA and EA GWIS variants in data from the GTEx consortium [42] revealed that rs291096, rs1856746, and rs2791713—the three *FCAMR-PIGR* locus variants—are

FCAMR-PIGR Locus – EA

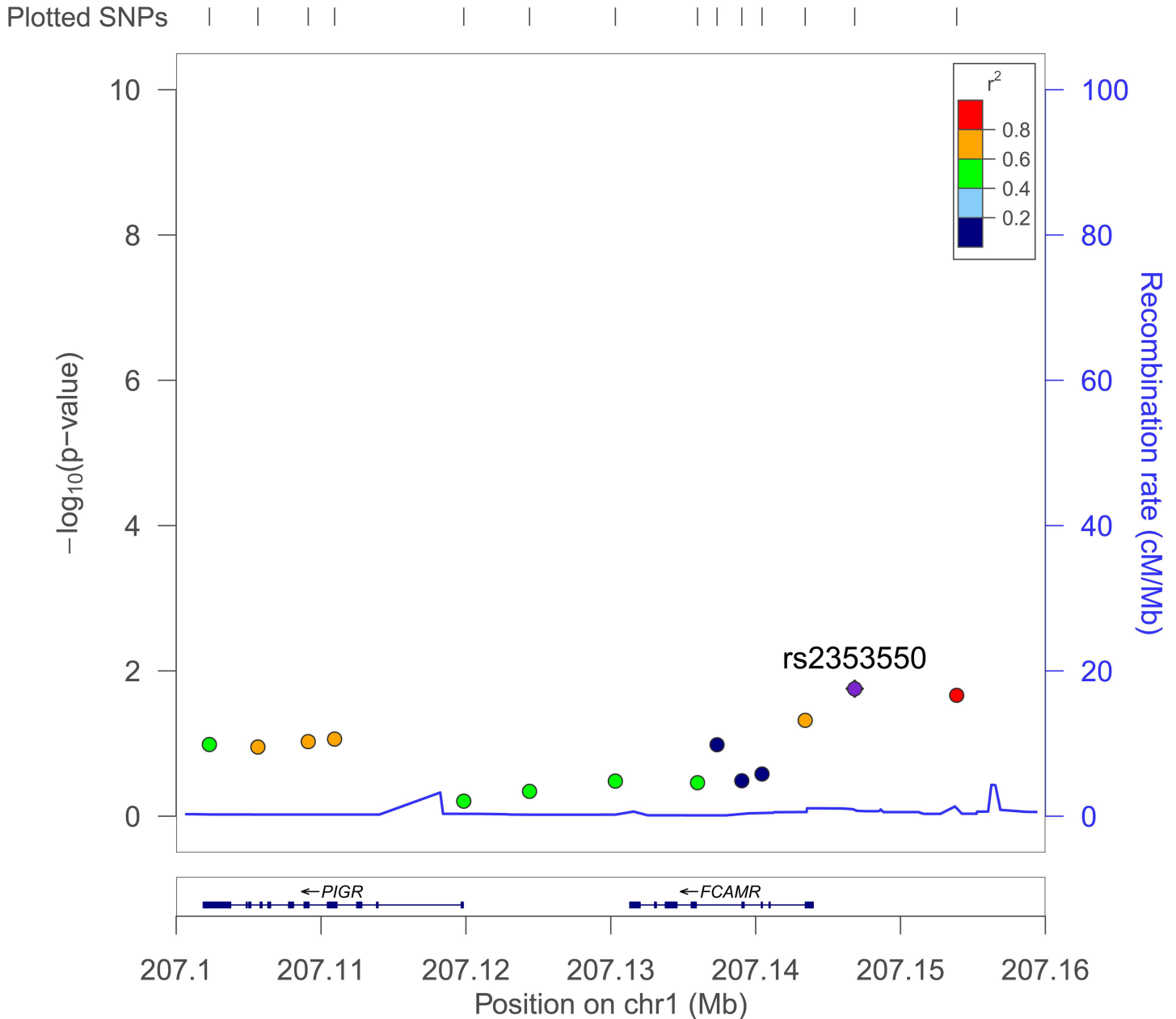


Fig 6. LocusZoom Plot of *FCAMR-PIGR* locus for European-Americans. Similar to what was observed in the African-Americans (Fig 5) substantial LD is observed in the *FCAMR-PIGR* region particularly between the replicated variant (rs185674) and variants in *PIGR*. For plotting 1000 Genomes data is used (population = EUR) in genome build hg19.

<https://doi.org/10.1371/journal.pone.0173880.g006>

eQTLs for *RP11-564A8.8* in lymphocytes. Rs17366136 is an eQTL for *AC073072.5* in tibial nerve and artery tissue (**Table C in S1 File**). From the meta-analysis, rs6894083 was an eQTL for *MARCH3* in skeletal muscle; it was also an eQTL for *C5orf63* in tibial nerve and aorta (**Table D in S1 File**).

Discussion

In this trans-ethnic meta-analysis of gene-by-traffic exposure interactions, we have uncovered multiple suggestive interactions; two of which replicated in an independent cohort. This study continues the success of recent genome-wide scans for environmental interactions associated with chronic disease. Genetic variants in the bone morphogenic protein family of genes associated were found to be associated with peripheral arterial disease [24] via a gene-by-traffic exposure GWIS. A smoking exposure GWIS identified novel genetic variants associated with coronary artery calcification, including genetic variants in *WVWX* that were also associated with peripheral arterial disease in a traffic air pollution exposure GWIS [24,25]. Here we found that the three most significant interactions in the AA GWIS were all in the *FCAMR-PIGR* locus. The two most significant interactions replicated in the EA GWIS, while the third was a synonymous coding variant within *PIGR*, which narrowly missed our replication threshold. In the EA GWIS the most significant interaction was located in *SAMSNI*. Though this interaction did not replicate in the AA GWIS, it was in a DNA hypersensitivity site in hematopoietic stem cells; further, the genetic variant correlated with the expression of *SAMSNI*.

FCAMR-PIGR associations

We observed multiple associations between variants in the *FCAMR-PIGR* locus on chromosome 1. A coding variant in *PIGR* was associated with the number of diseased coronary vessels in our interaction model (rs291096, $P = 2.3 \times 10^{-6}$). The association signal from this exonic variant combined with the replication of the two most significant variants from the AA GWIS—both of which were in the *FCAMR-PIGR* locus—prompting us to further investigate this locus. Given the proximity of the two genes, their similar functions, and LD structure (**Figs 5 and 6**) we cannot completely disregard that potentially both *PIGR* and *FCAMR* are involved in mediating the effects of traffic-related air pollution on coronary atherosclerosis.

The protein products of both *FCAMR* and *PIGR* are participants in immune response. *FCAMR* is a receptor for the Fc fragments of immunoglobulin, and is upregulated by *IL1A*, an inflammatory cytokine [47]. *PIGR* is a poly-Ig receptor mediating the transport of polymeric immunoglobulin molecules [48], and is primarily found in the mucosal epithelium [49]. Both *PIGR* and *FCAMR* recognize IgA and IgM [49,50], and play an important role in the immune response in mucosal cells. *FCAMR* is widely expressed in the kidney, intestine, heart, and lung while *PIGR* is primarily found in the mucosal epithelium [49]. Both *PIGR* and *FCAMR* are implicated in IgA nephropathy, a condition where IgA immune complexes are deposited in the glomerular mesangium [47,50]. All three *FCAMR-PIGR* variants from the suggestive AA GWIS interactions were associated with the expression of *RP11-564A8.8* in lymphocytes. Although *RP11-564A8.8* is a pseudogene of unknown function, these variants potentially play an important functional role in immune cells. Though *FCAMR* has yet to be directly associated with air pollution, *PIGR* plasma levels are greater in smokers than in non-smokers, confirming that plasma *PIGR* is regulated by inhaled pollutants [51]. Finally, in a recent metabolomics quantitative trait locus analysis in this dataset a variant in *PIGR* was associated with a cluster of long-chain dicarboxyl acylcarnities (LCDA) metabolites [32]. LCDA metabolites are predictive of cardiovascular disease events [27,52].

SAMSN1

In the EA cohort, we found suggestive evidence that rs2822693 is associated with the number of diseased coronary vessels ($P = 2.2 \times 10^{-6}$) via an interaction with traffic exposure. Rs2822693 is located in an open chromatin region downstream of *SAMSN1*. Open chromatin regions signify nucleosome free regions of DNA, often correlated with the binding of regulatory factors. The region containing rs2822693 was open in hematopoietic cells, and the gene expression of *SAMSN1* was highly correlated with this hypersensitivity site [40].

SAMSN1 (also known as *HACSI*) encodes a 441 amino-acid protein with SAM and SH3 domains that indicate adaptor or scaffolding functions. Expression of *SAMSN1* is detected in several tissues including the brain, lung, heart, and hematopoietic stem cells [53]. *SAMSN1* expression is up-regulated in B cell activation signaling cascades [54], as well as in the peripheral blood mononuclear cells [55] and atherosclerotic lesions of femoral arteries [56] of PAD cases.

Meta-analysis

In a sample size weighted meta-analysis combining the EA and AA cohorts, the most significant interaction was with rs10830090, located in an intron of *RARS2*, a mitochondrial arginine t-RNA synthase. Particulate air pollution is known to cause mitochondrial damage [45] which can lead to cardiovascular disease [46], making *RARS2* an interesting candidate gene for further analyses. As this was a sample size weighted meta-analysis the most significant interactions from the AA GWIS do not have suggestive associations as their P-values were often lower in GWIS of the larger EA population.

Uniqueness, strengths and limitations

This study is the first genome-wide interaction study of air pollution and coronary atherosclerosis. Although there are other studies of gene-air pollution interactions in cardiovascular disease [23], this is the first genome-wide examination. A strength of the study design is the use of independent race-stratified cohorts to replicate interactions; this helped to overcome the limitation of our limited sample size in each of our race-stratified cohorts by allowing us to validate observed associations in an independent sample drawn from the same population. It also allows us to observe the ethnicity driven heterogeneity in any observed associations, though this should be extended by testing these associations in ethnicities beyond EA and AA. The detailed clinical phenotyping and precise quantification of the extent of coronary artery pathology is a strength not typically available in genetic epidemiologic studies. Additionally, the distance to primary roadways is a well-validated proxy for traffic air pollution exposure that has been used in several studies [15,19,21,24,57]; it is strongly correlated with air pollution in our study area [14] and allowed us to evaluate air pollution exposure on all individuals with address information. We used the number of diseased coronary vessels as a measure of coronary atherosclerosis. This measure allowed us to evaluate clinically significant atherosclerosis as assessed by a physician across multiple coronary vessels. A strength of our measure is that it allows for assessment of clinically significant atherosclerosis across multiple vessels, however a limitation is that it does not assess subclinical atherosclerosis that may still be impacted by these associations.

The primary limitation of this study is that we do not have additional replication cohorts. For some of the SNPs that achieved a suggestive level of significance we were able to replicate the association in an independent cohort of a different ethnicity drawn from the same source population, which strengthens our confidence in the associations. Of course, replication in additional independent populations is required. Until then these current results require careful

interpretation. A secondary limitation is not being able to assess the specific traffic air pollution components contributing to the association and evaluate their association with air pollution sources beyond traffic. Associations with traffic exposure zones may be able to refine the associations, but would require detailed maps of land usage and traffic patterns across the state, which are currently only available for a limited number of counties [21,58]. Future studies should assess whether these genes only interact with traffic-generated air pollution or if sources such as biomass burning or wood smoke generate similar associations. In this study we are also limited to a single geographic region. To better evaluate generalizability, these associations should be validated in other regions where local traffic emissions may differ from what we observed.

Conclusion

We use a race-stratified genome-wide interaction study design with meta-analysis to investigate the role of genetic variants in modifying the association between traffic related air pollution and coronary atherosclerosis. Using this approach, we identified several novel candidate genes that may link air pollution and coronary atherosclerosis. Candidate gene interaction studies have added much to our knowledge of air pollution and coronary atherosclerosis [23]; our findings add to these studies by implicating additional inflammatory response genes (*FCAMR* and *PIGR*). We also found an EA-specific interaction in a regulatory region associated with *SAMSNI*, a gene previously associated with vascular disease. These results demonstrate the importance of considering both genetic and environmental factors in assessments of cardiovascular disease risk while highlighting the need for further systems biology research in this field.

Supporting information

S1 File. Contains Supplemental Tables A-D. (DOCX)

Acknowledgments

We would like to acknowledge and thank all CATHGEN staff and participants for making this work possible. The research described in this article has been reviewed by the National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Agency, nor does the mention of trade names of commercial products constitute endorsement or recommendation for use.

Author Contributions

Conceptualization: CKWC WEK ERH.

Data curation: CB CSH ZED.

Formal analysis: CKWC KLA EG.

Funding acquisition: SGG RBD DDS WEC MLM WEK ERH.

Methodology: CKWC LMN WEK ERH.

Project administration: CB CSH ZED RBD DDS WEC MLM SGG SHS WEK ERH.

Resources: MLM SGG SHS WEK ERH.

Software: CKWC.

Supervision: LMN SGG MLM SHS WEK ERH.

Writing – original draft: CKWC WEC MLM WEK ERH.

Writing – review & editing: CKWC WEK ERH LMN.

References

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. (2013) Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* 380: 2095–2128.
2. Tabas I (2011) Pulling down the plug on atherosclerosis: finding the culprit in your heart. *Nature medicine* 17: 791–793. <https://doi.org/10.1038/nm0711-791> PMID: 21738159
3. The CARDIoGRAMplusC4D Consortium (2015) A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 47: 1121–1130. <https://doi.org/10.1038/ng.3396> PMID: 26343387
4. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. (2013) Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* 45: 25–33. <https://doi.org/10.1038/ng.2480> PMID: 23202125
5. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* 109: 1193–1198.
6. Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, et al. (2010) Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement From the American Heart Association. *Circulation* 121: 2331–2378. <https://doi.org/10.1161/CIR.0b013e3181dbee1> PMID: 20458016
7. Cesaroni G, Forastiere F, Stafoggia M, Andersen ZJ, Badaloni C, Beelen R, et al. (2014) Long term exposure to ambient air pollution and incidence of acute coronary events: prospective cohort study and meta-analysis in 11 European cohorts from the ESCAPE Project. *BMJ* 348.
8. Kaufman JD, Adar SD, Barr RG, Budoff M, Burke GL, Curl CL, et al. (2016) Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. *The Lancet*.
9. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. (2013) A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet* 380: 2224–2260.
10. Beckerman BS, Jerrett M, Finkelstein M, Kanaroglou P, Brook JR, Arain MA, et al. (2012) The Association Between Chronic Exposure to Traffic-Related Air Pollution and Ischemic Heart Disease. *Journal of Toxicology and Environmental Health, Part A* 75: 402–411.
11. Gan WQ, Davies HW, Koehoorn M, Brauer M (2012) Association of Long-term Exposure to Community Noise and Traffic-related Air Pollution With Coronary Heart Disease Mortality. *American Journal of Epidemiology* 175: 898–906. <https://doi.org/10.1093/aje/kwr424> PMID: 22491084
12. Rosenlund M, Picciotto S, Forastiere F, Stafoggia M, Perucci CA (2008) Traffic-related air pollution in relation to incidence and prognosis of coronary heart disease. *Epidemiology* 19: 121–128. <https://doi.org/10.1097/EDE.0b013e31815c1921> PMID: 18091421
13. Zhu Y, Hinds WC, Kim S, Sioutas C (2002) Concentration and size distribution of ultrafine particles near a major highway. *J Air & Waste manage Assoc* 52: 1032–1042.
14. Hagler GSW, Baldauf RW, Thoma ED, Long TR, Snow RF, Kinsey JS, et al. (2009) Ultrafine particles near a major roadway in Raleigh, North Carolina: Downwind attenuation and correlation with traffic-related pollutants. *Atmospheric Environment* 43: 1229–1234.
15. Hoffmann B, Moebus S, Dragano N, Möhlenkamp S, Memmesheimer M, Erbel R, et al. (2009) Residential traffic exposure and coronary heart disease: results from the Heinz Nixdorf Recall Study. *Biomarkers* 14: 74–78. <https://doi.org/10.1080/13547500902965096> PMID: 19604064
16. Hoffmann B, Moebus S, Möhlenkamp S, Stang A, Lehmann N, Dragano N, et al. (2007) Residential Exposure to Traffic Is Associated With Coronary Atherosclerosis. *Circulation* 116: 489–496. <https://doi.org/10.1161/CIRCULATIONAHA.107.693622> PMID: 17638927
17. Ghosh R, Lurmann F, Perez L, Penfold B, Brandt S, Wilson J, et al. (2016) Near-Roadway Air Pollution and Coronary Heart Disease: Burden of Disease and Potential Impact of a Greenhouse Gas Reduction

- Strategy in Southern California. *Environmental Health Perspectives* 124: 193–200. <https://doi.org/10.1289/ehp.1408865> PMID: 26149207
18. Wang Y, Wellenius GA, Hickson DA, Gjelsvik A, Eaton CB, Wyatt SB, et al. (2016) Residential Proximity to Traffic-Related Pollution and Atherosclerosis in 4 Vascular Beds Among African-American Adults: Results From the Jackson Heart Study. *American Journal of Epidemiology*.
 19. Hoffmann B, Moebus S, Kröger K, Stang A, Möhlenkamp S, Dragano N, et al. (2009) Residential Exposure to Urban Air Pollution, Ankle–Brachial Index, and Peripheral Arterial Disease. *Epidemiology* 20: 280–288. <https://doi.org/10.1097/EDE.0b013e3181961ac2> PMID: 19194299
 20. Li W, Dorans KS, Wilker EH, Rice MB, Schwartz J, Coull BA, et al. (2016) Residential proximity to major roadways, fine particulate matter, and adiposity: The framingham heart study. *Obesity*.
 21. Ward-Caviness CK, Kraus WE, Blach C, Haynes CS, Dowdy E, Miranda ML, et al. (2015) Association of roadway proximity with fasting plasma glucose and metabolic risk factors for cardiovascular disease in a cross-sectional study of cardiac catheterization patients. *Environmental health perspectives* 123: 1007. <https://doi.org/10.1289/ehp.1306980> PMID: 25807578
 22. Lanktree MB, Hegele RA (2009) Gene-gene and gene-environment interactions: new insights into the prevention, detection and management of coronary artery disease. *Genome Med* 1: 28. <https://doi.org/10.1186/gm28> PMID: 19341499
 23. Zanobetti A, Baccarelli A, Schwartz J (2011) Gene–Air Pollution Interaction and Cardiovascular Disease: A Review. *Progress in Cardiovascular Diseases* 53: 344–352. <https://doi.org/10.1016/j.pcad.2011.01.001> PMID: 21414469
 24. Ward-Caviness CK, Neas LM, Blach C, Haynes CS, LaRocque-Abramson K, Grass E, et al. (2016) Genetic Variants in the Bone Morphogenic Protein Gene Family Modify the Association between Residential Exposure to Traffic and Peripheral Arterial Disease. *PloS one* 11: e0152670. <https://doi.org/10.1371/journal.pone.0152670> PMID: 27082954
 25. Polfus LM, Smith JA, Shimmin LC, Bielak LF, Morrison AC, Kardina SLR, et al. (2013) Genome-Wide Association Study of Gene by Smoking Interactions in Coronary Artery Calcification. *PLoS One* 8: e74642. <https://doi.org/10.1371/journal.pone.0074642> PMID: 24098343
 26. Kraus WE, Granger CB, Sketch MH Jr, Donahue MP, Ginsburg GS, Hauser ER, et al. (2015) A guide for a cardiovascular genomics biorepository: the CATHGEN experience. *Journal of cardiovascular translational research* 8: 449–457. <https://doi.org/10.1007/s12265-015-9648-y> PMID: 26271459
 27. Shah SH, Bain JR, Muehlbauer MJ, Stevens RD, Crosslin DR, Haynes C, et al. (2010) Association of a Peripheral Blood Metabolic Profile With Coronary Artery Disease and Risk of Subsequent Cardiovascular Events. *Circulation: Cardiovascular Genetics* 3: 207–214.
 28. McGuinn LA, Ward-Caviness CK, Neas LM, Schneider A, Diaz-Sanchez D, Cascio WE, et al. (2016) Association between satellite-based estimates of long-term PM 2.5 exposure and coronary artery disease. *Environmental research* 145: 9–17. <https://doi.org/10.1016/j.envres.2015.10.026> PMID: 26613345
 29. NCDOT (North Carolina Department of Transportation) (2015) GIS Data Layers. Statewide Primary & Secondary Road Routes. <https://connect.ncdot.gov/resources/gis/Pages/GIS-Data-Layers.aspx>. North Carolina Department of Transportation.
 30. U.S. Census Bureau (2013) TIGER/Line® Shapefiles and TIGER/Line® Files. 2013 TIGER/Line Shapefiles Technical Documentation. <https://www.census.gov/geo/maps-data/data/tiger-line.html>. U.S. Census Bureau.
 31. ESRI (2011) ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.
 32. Kraus WE, Muoio DM, Stevens R, Craig D, Bain JR, Grass E, et al. (2015) Metabolomic Quantitative Trait Loci (mQTL) Mapping Implicates the Ubiquitin Proteasome System in Cardiovascular Disease Pathogenesis. *PLoS Genet* 11: e1005553. <https://doi.org/10.1371/journal.pgen.1005553> PMID: 26540294
 33. Christensen RHB (2010) ordinal—regression models for ordinal data. R package version 22.
 34. R Development Core Team (2012) R: A Language and Environment for Statistical Computing. Vienna, Austria.
 35. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904–909. <https://doi.org/10.1038/ng1847> PMID: 16862161
 36. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190–2191. <https://doi.org/10.1093/bioinformatics/btq340> PMID: 20616382
 37. Hunter DJ (2005) Gene–environment interactions in human diseases. *Nature Reviews Genetics* 6: 287–298. <https://doi.org/10.1038/nrg1578> PMID: 15803198

38. Murcay CE, Lewinger JP, Gauderman WJ (2009) Gene-environment interaction in genome-wide association studies. *American journal of epidemiology* 169: 219–226. <https://doi.org/10.1093/aje/kwn353> PMID: 19022827
39. Sherry S, Ward M, Kholodov M, Phan L, Smigielski E, Sirotkin K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* 29: 308–311. PMID: 11125122
40. Sheffield NC, Thurman RE, Song L, Safi A, Stamatoyannopoulos JA, Lenhard B, et al. (2013) Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome research* 23: 777–788. <https://doi.org/10.1101/gr.152140.112> PMID: 23482648
41. Maurano MT, Haugen E, Sandstrom R, Vierstra J, Shafer A, Kaul R, et al. (2015) Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat Genet* 47: 1393–1401. <https://doi.org/10.1038/ng.3432> PMID: 26502339
42. The GTEx Consortium (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* 348: 648–660. <https://doi.org/10.1126/science.1262110> PMID: 25954001
43. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. (2015) The human transcriptome across tissues and individuals. *Science* 348: 660–665. <https://doi.org/10.1126/science.aaa0355> PMID: 25954002
44. Cassandrini D, Cilio MR, Bianchi M, Doimo M, Balestri M, Tessa A, et al. (2013) Pontocerebellar hypoplasia type 6 caused by mutations in RARS2: definition of the clinical spectrum and molecular findings in five patients. *J Inher Metab Dis* 36: 43–53. <https://doi.org/10.1007/s10545-012-9487-9> PMID: 22569581
45. Li N, Sioutas C, Cho A, Schmitz D, Misra C, Sempf J, et al. (2003) Ultrafine particulate pollutants induce oxidative stress and mitochondrial damage. *Environmental Health Perspectives* 111: 455–460. PMID: 12676598
46. Ballinger SW (2005) Mitochondrial dysfunction in cardiovascular disease. *Free Radic Biol Med* 38: 1278–1295. <https://doi.org/10.1016/j.freeradbiomed.2005.02.014> PMID: 15855047
47. McDonald KJ, Cameron AJ, Allen JM, Jardine AG (2002) Expression of Fc $\alpha\mu$ receptor by human mesangial cells: a candidate receptor for immune complex deposition in IgA nephropathy. *Biochemical and biophysical research communications* 290: 438–442. <https://doi.org/10.1006/bbrc.2001.6218> PMID: 11779189
48. Hood L, Kronenberg M, Hunkapiller T (1985) T cell antigen receptors and the immunoglobulin supergene family. *Cell* 40: 225–229. PMID: 3917857
49. Wines BD, Hogarth PM (2006) IgA receptors in health and disease. *Tissue antigens* 68: 103–114. <https://doi.org/10.1111/j.1399-0039.2006.00613.x> PMID: 16866880
50. Barratt J, Feehally J, Smith AC. *Pathogenesis of IgA nephropathy*; 2004. Elsevier. pp. 197–217.
51. Ohlmeier S, Mazur W, Linja-aho A, Louhelainen N, Rönty M, Toljamo T, et al. (2011) Sputum proteomics identifies elevated PIGR levels in smokers and mild-to-moderate COPD. *Journal of proteome research* 11: 599–608. <https://doi.org/10.1021/pr2006395> PMID: 22053820
52. Shah SH, Sun J-L, Stevens RD, Bain JR, Muehlbauer MJ, Pieper KS, et al. (2012) Baseline metabolic profiles predict cardiovascular events in patients at risk for coronary artery disease. *American Heart Journal* 163: 844–850.e841. <https://doi.org/10.1016/j.ahj.2012.02.005> PMID: 22607863
53. Claudio JO, Zhu YX, Benn SJ, Shukla AH, McGlade CJ, Falcioni N, et al. (2001) HACS1 encodes a novel SH3-SAM adaptor protein differentially expressed in normal and malignant hematopoietic cells. *Oncogene* 20: 5373. <https://doi.org/10.1038/sj.onc.1204698> PMID: 11536050
54. Zhu YX, Benn S, Li ZH, Wei E, Masih-Khan E, Trieu Y, et al. (2004) The SH3-SAM adaptor HACS1 is up-regulated in B cell activation signaling cascades. *The Journal of experimental medicine* 200: 737–747. <https://doi.org/10.1084/jem.20031816> PMID: 15381729
55. Masud R, Shameer K, Dhar A, Ding K, Kullo IJ (2012) Gene expression profiling of peripheral blood mononuclear cells in the setting of peripheral arterial disease. *J Clinical Bioinformatics* 2.
56. Fu S, Zhao H, Shi J, Abzhanov A, Crawford K, Ohno-Machado L, et al. (2008) Peripheral arterial occlusive disease: global gene expression analyses suggest a major role for immune and inflammatory responses. *BMC Genomics* 9: 369. <https://doi.org/10.1186/1471-2164-9-369> PMID: 18673543
57. Hoffmann B, Moebus S, Kröger K, Stang A, Erbel R, Jöckel K. (2007) Long-Term Residential Exposure to Traffic and Peripheral Arterial Disease in the Heinz Nixdorf Recall Study. *Epidemiology* 18: S75.
58. Shaibal M, Luther S, Halley B, Casson S, Lucas N, Sue K, et al. (2015) Comparison of modeled traffic exposure zones using on-road air pollution measurements. *Atmospheric Pollution Research* 6: 82–87.