Role of Rare and Low-Frequency Variants in Gene-Alcohol Interactions on Plasma Lipid Levels

Running title: *Wang et al.; Rare Genetic Variants and Alcohol on Lipids*

Zhe Wang, PhD[1]; Han Chen, PhD[1,2]; Traci M. Bartz, MS[3]; Lawrence F. Bielak, PhD [4]; Daniel I. Chasman, PhD[5,6]; Mary F. Feitosa, PhD[7]; Nora Franceschini, MD, MPH[8]; Xiuqing Guo, PhD[9]; Elise Lim, MS[10]; Raymond Noordam, PhD[11]; Melissa A. Richard, PhD[12]; Heming Wang, PhD[6,13]; Brian Cade, PhD[6,13]; L. Adrienne Cupples, PhD[10,14]; Paul S. de Vries, PhD[1]; Franco Giulanini, PhD[5]; Jiwon Lee, MS[6,13]; Rozenn N. Lemaitre, PhD[15]; Lisa W. Martin, MD[16]; Alex P. Reiner, MD[17]; Stephen S. Rich, PhD[18]; Pamela J. Schreiner, PhD[19]; Stephen Sidney, MD, MPH[20]; Colleen M. Sitlani, PhD[15]; Jennifer A. Smith, PhD[4,21]; Ko Willems van Dijk, PhD[22,23]; Jie Yao, MS[9]; Wei Zhao, PhD[4]; Myriam Fornage, PhD[1,12]; Sharon L.R. Kardia, PhD[4]; Charles Kooperberg, PhD[17]; Ching-Ti Liu, PhD[10]; Dennis O. Mook-Kanamori, PhD[24,25]; Michael A. Province, PhD[7]; Bruce M. Psaty, MD, PhD[26,27]; Susan Redline, MD, MPH[6,13]; Paul M. Ridker, MD, MPH[6,13]; Jerome I. Rotter, MD[9]; Eric Boerwinkle, PhD[1,28]; Alanna C. Morrison, PhD[1]; On behalf of the CHARGE Gene-Lifestyle Interactions Working Group

[1]Human Genetics Center, Dept of Epidemiology, Human Genetics & Environmental Sciences, School of Public Health, [2]Center for Precision Health, School of Public Health & School of Biomedical Informatics, 12Brown Foundation Inst of Molecular Medicine, Univ of Texas Health Science Ctr at Houston, Houston, TX; [3]Cardiovascular Health Research Unit, Depts of Biostatistics & Medicine, [15]Cardiovascular Health Research Unit, Dept of Medicine, [26]Cardiovascular Health Research Unit, Depts of Medicine, Epidemiology & Health Services, Univ of Washington, Seattle, WA; [4]School of Public Health, Dept of Epidemiology, [21]Inst for Social Research, Survey Research Ctr, Univ of Michigan, Ann Arbor, MI; [5]Division of Preventive Medicine, [13]Division of Sleep & Circadian Disorders, Dept of Medicine, Brigham and Women's Hospital; [6]Harvard Medical School, Boston, MA; [7]Division of Statistical Genomics, Dept of Genetics, Washington Univ School of Medicine, St. Louis, MO; [8]Dept of Epidemiology, Gillings School of Global Public Health, Univ of North Carolina, Chapel Hill, NC; [9]The Inst for Translational Genomics & Population Sciences, Dept of Pediatrics, The Lundquist Institute at Harbor-UCLA Medical Ctr, Torrance, CA; [10]Biostatistics Dept, Boston Univ School of Public Health, Boston, MA; [11]Section of Gerontology & Geriatrics, Dept of Internal Medicine, [22]Dept of Human Genetics, [23]Division of Endocrinology, Dept of Internal Medicine, [24]Dept of Clinical Epidemiology, [25]Dept of Public Health and Primary Care, Leiden Univ Medical Ctr, Leiden, Netherlands; [14]NHLBI Framingham Heart Study, Framingham, MA; [16]George Washington Univ School of Medicine & Health Sciences, Washington, DC; [17]Division of Public Health Sciences, Fred Hutchinson Cancer Research Ctr, Seattle, WA; [18]Ctr for Public Health Genomics, Dept of Public Health Sciences, Univ of Virginia, Charlottesville, VA; [19]Epidemiology & Community Health, School of Public Health, Univ of Minnesota, Minneapolis, MN; [20]Division of Research, Kaiser Permanente of Northern California, Oakland, CA; [27]Kaiser Permanente Washington Health Research Inst, Seattle, WA; [28]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX

Correspondence:

Alanna C. Morrison, PhD

UTHealth School of Public Health

RAS E447, 1200 Pressler St

Houston, TX, 77030

Tel: 713-500-9913

Email: alanna.c.morrison@uth.tmc.edu

Journal Subject Terms: Lipids and Cholesterol; Genetic, Association Studies; Biomarkers

Abstract:

Background - Alcohol intake influences plasma lipid levels and such effects may be moderated by genetic variants. We aimed to characterize the role of aggregated rare and low-frequency protein coding variants in gene by alcohol consumption interactions associated with fasting plasma lipid levels.

Methods - In the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, fasting plasma triglycerides (TG), and high- and low-density lipoprotein cholesterol (HDL-C and LDL-C) were measured in 34,153 individuals with European ancestry from five discovery studies and 32,277 individuals from six replication studies. Rare and low-frequency functional protein coding variants (minor allele frequency ≤ 5%) measured by an exome array were aggregated by genes and evaluated by a gene-environment interaction (G×E) test and a joint test of genetic main and G×E interaction effects. Two dichotomous self-reported alcohol consumption variables, current drinker, defined as any recurrent drinking behavior, and regular drinker, defined as the subset of current drinkers who consume at least two drinks per week, were considered.

Results - We discovered and replicated 21 gene-lipid associations at 13 known lipid loci through the joint test. Eight loci (*PCSK9, LPA, LPL, LIPG, ANGPTL4, APOB, APOC3 and CD300LG*) remained significant after conditioning on the common index single nucleotide polymorphism (SNP) identified by previous genome-wide association studies, suggesting an independent role for rare and low-frequency variants at these loci. One significant gene-alcohol interaction on TG in a novel locus was significantly discovered (p-value = 6.65×10^{-6} for the interaction test) and replicated at nominal significance level (p-value = 0.013) in *SMC5*.

Conclusions - In conclusion, this study applied new gene-based statistical approaches and suggested that rare and low-frequency genetic variants interacted with alcohol consumption on lipid levels.

Key words: Genome Wide Association Study; lipids; alcohol; gene-environment interactions, rare variant test

Nonstandard Abbreviations and Acronyms

HDL-C	high-density lipoprotein cholesterol
LDL-C	low-density lipoprotein cholesterol
TG	triglyceride
CVD	cardiovascular disease
GWAS	genome-wide association study
SNP	single nucleotide polymorphism
MAF	minor allele frequency
G×E	gene-by-environment
MAC	minor allele count

Plasma lipid profiles, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglyceride (TG) levels have been well characterized for their roles in the development and prevention of cardiovascular disease (CVD) [1, 2]. Genome-wide association studies (GWAS) and advanced DNA sequence technology have uncovered more than two hundred genetic loci influencing lipid levels [3-8], and these common (minor allele frequency [MAF] >5%) single nucleotide polymorphisms (SNPs) often reside in non-coding regions of the genome. In addition to the evidence that genetic factors affect plasma lipid profiles, environmental factors influence lipid levels as well. Epidemiologic studies have demonstrated an association between moderate alcohol consumption and improved lipid profile, including higher HDL-C levels, HDL particle concentration, and HDL-C subfractions [9, 10]. However, the association evidence between alcohol use and LDL-C or TG levels is inconsistent. Some studies reported positive associations while others reported negative associations [11-15].

Studying gene-by-environment (G×E) interactions is important, as it extends our knowledge of the genetic architecture of complex traits and improves our understanding of the underlying mechanisms of common diseases for novel and known loci [16-18]. Several large-scale genome-wide G×E studies have successfully identified novel common variants accounting for the environmental effects such as alcohol consumption and smoking status on lipid levels and other CVD related traits [19-22]. These studies have successfully identified common variant loci that were not detected in main effects GWAS. However, unlike well-established G×E interaction tests for common variants [23, 24], methods for detecting rare variant G×E interactions are emerging. Recently developed novel approaches for testing rare variant G×E interaction effects include a joint test that allows for simultaneous testing of the genetic main effect and interaction

effect as well as the ability to assess gene-based G×E interactions for both related and unrelated individuals [25].

Accounting for the effect of alcohol consumption in defining the genetic architecture of lipid levels may not only provide valuable insights into relationship between alcohol consumption and lipids, but also may help refine association signals at previously identified GWAS loci or identify new loci. This study is the first to incorporate G×E interaction in modeling rare and low-frequency variant genetic and alcohol effects on plasma lipid levels.

Methods

This study includes 66,430 men and women between 18-80 years of age from 11 European-ancestry population studies that are part of the CHARGE Gene-Lifestyle Interactions Working Group [18] (Supplemental Figure 1). Each study obtained informed consent from participants and approval from the appropriate institutional review boards. Additional detail for these studies and full Methods are available in the Data Supplement of the article. Data from consortia were accessed subject to the applicable data-sharing agreements. Summary data are available to other researcher on reasonable request to the corresponding authors.

Results

Descriptive statistics for up to 34,153 participants of the five discovery and 32,277 participants of the six replication studies are summarized in **Table 1** and Supplemental Table 1. On average, two thirds of the study participants were current drinkers and 39.5 percent were regular drinkers. The proportion of current and regular drinkers was greater in the discovery studies as compared to the replication studies.

We performed gene-based analyses for each lipid/alcohol consumption combination using: 1) a G×E test that considers the genetic main effects as random effects, and 2) a joint analysis of the genetic main and the G×E interaction effects in each study participating in the discovery phase. Significant genes from meta-analysis of the discovery studies were pursued for replication. Overall, meta-analyses showed highly consistent results across current drinker and regular drinker (Supplemental Table 2). Distributions of QQ plots for meta-analyzing discovery studies are shown in Supplemental Figure 2. In the discovery phase, we observed 31 gene-lipid associations (p-value < 5×10^{-5}) in the joint analysis and 5 gene-lipid associations (p-value < 5×10^{-5}) in the interaction test, with 3 genes (*INDK*, *REM2*, and *SMC5*) overlapping between the two approaches (Supplemental Table 2). These gene-lipid pairs were taken forward for replication, one of which (*IDNK)* was only available in one replication study (the CHS). Therefore, we evaluated 30 gene-lipid associations for replication using the joint test and 4 using the gene-alcohol interaction test (Supplemental Table 2). Thirteen known lipid loci (21 gene-lipid associations) were replicated and one novel interaction at a novel locus was replicated at the borderline for Bonferroni corrected significant level (p_{int} = 0.013) for the *SMC5*-by-current drinker interaction on TG levels (**Table 2**). The average TG levels for SMC5 carriers and non-carriers by current drinker status among discovery studies were showed in **Figure 1**. Among the replicated genes, 4 were shared between TG and HDL-C but none were shared between LDL-C and TG or HDL-C, as shown in a Venn diagram (**Figure 2**).

For the 13 known lipid loci that were replicated through the joint test, additional analyses were conducted following the flowchart shown in **Figure 3**. First, we performed conditional analyses in order to examine whether the gene-based rare variant associations are independent of the common index SNP identified by previous GWAS. In total, 8 loci (*PCSK9, LPA, LPL, LIPG*,

ANGPTL4, APOB, APOC3 and CD300LG) (10 gene-lipid associations) remained significant after conditioning on a common index SNP. However, the genes that were not reported to be associated with lipids themselves but in known lipid loci, such as *BCAM* and *CBLC* on LDL-C, were strongly attenuated after adjusting for rs7412, the index SNP of *APOE* identified by previous GWAS and in part defining the APOE2/3/4 alleles (Supplemental Table 3).

Second, single variant analyses were performed for the 5 gene-lipid associations that were not evaluated in the conditional analyses because they did not have previously reported common SNPs and for the 10 gene-lipid pairs that remained significant following conditional analyses (**Figure 3,** Supplemental Table 3). Single variant tests at these genes confirmed previous known low-frequency lipid variants. For example, rs11591147 in *PCSK9* was associated with LDL-C, and rs77960347 in *LIPG* and rs116843064 in *ANGPTL4* were associated with HDL-C. Additionally, we provide evidence that two of the driving variants underlying the joint test results are novel rare variants associated with LDL-C (Supplemental Table 4). One of them is rs41267813, a variant in the *LPA* gene ($p = 6.55\times10^{-29}$ discovery, $p = 1.83\times10^{-03}$ replication and the other is rs41288783 of *APOB* gene ($p = 5.40\times10^{-08}$ discovery, $p = 7.92\times10^{-07}$ replication). In the joint model, the genetic main effect per A allele of rs41267813 was associated with a 31.6 mg/dL decrease in LDL-C levels ($\beta_{main}[se_{main}] = -31.55[2.78]$), while the estimated interaction effect indicated a positive interaction with regular drinker status (β_{int} [se_{int}] = 27.07[5.66]). In contrast, the genetic main effect of rs41288783 was associated with an increase in LDL-C levels among regular drinkers as well as non-drinkers ($\beta_{main}[se_{main}]$ = 16.18[5.34], β_{int} [se_{int}] = 11.03[7.68]). For the novel interaction between *SMC5* and current drinker on TG levels, we identified the driving variant as rs142488686, a missense mutation (minor allele count (MAC) = 5-7 discovery (ARIC and CARDIA), MAC = 7-17 replication

(WGHS, CHS and MESA)), with a replicating interaction effect (p_{int} = 0.016 discovery, p_{int} = 0.008 replication), while the genetic main effect was modest ($p < 0.1$ discovery and replication, respectively).

Discussion

This is the first large-scale study to evaluate the role of rare and low-frequency variants in lipids by incorporating gene-alcohol consumption interactions. We tested for gene-alcohol interaction effects on lipid levels as well as the joint effects of genetic main effects and gene-alcohol interactions. We replicated 13 gene-lipid associations at known lipid loci, among which 2 leading rare variants in *APOB* and *LPA* genes associated with LDL-c were novel. Only one novel gene-alcohol interaction was identified as significant and replicated at nominal significance level (the interaction between rare and low-frequency variants in *SMC5* and current drinker on TG levels).

Using a single variant test, we confirmed numerous previously identified rare and low-frequency missense lipid variants. For example, rs11591147 (MAF ~1.5%) of *PCSK9* has been associated with LDL-C levels [24, 26], rs77960347 (MAF ~1.2%) of *LIPG* and rs116843064 (MAF ~2.0%) of *ANGPTL4* have been associated with HDL-C levels [27, 28]. A missense mutation in the *APOC3* gene, rs147210663(MAF ~0.07%), has been associated with a more than 40% lower average triglyceride level in individuals carrying one A allele [29, 30]. In the present study, we observed a novel relationship between increased HDL-C levels in individuals carrying rs147210663 (A) allele as rs147210663 was previously reported as a founder mutation in a Pennsylvania Amish population associated with TG [31].

Between the two novel rare driving variants that were identified and replicated, rs41267813 (*LPA*, missense variant, MAF ~0.16%) is located 28kb away from a stop/gain variant rs41267811 *(LPA*, MAF ~0.02%) that was also significantly associated with LDL-C levels in the discovery phase. However, we were unable to replicate the association with rs41267811 as it was only available in one replication study (WGHS) and therefore did not meet our criteria to be included in replication. *LPA* encoded protein constitutes a substantial portion of lipoprotein(a) and associated with inherited conditions including type III hyperlipoproteinemia and familial hyperlipidemia [32]. A stop/gain mutation in this gene would be associated with lower LDL-C levels in carriers, which is true among non-drinkers. However, such association may be modified by alcohol consumption as we observed the carriers of this variant with a higher LDL-C levels compared to non-carriers in a population who had at least two drinks per week in the ARIC study. A previous study of gene-alcohol interaction on lipids focusing on common variants identified rs5014650 (MAF 15%, intergenic), at the *LPA* locus that was associated with LDL-C levels in a joint test [21], suggesting that this locus affects LDL-C levels through both main effects and an interaction with alcohol consumption. Previous studies have reported a relationship between moderate alcohol consumption and lower Lp(a) lipoprotein concentrations [33, 34], yet no published evidence of an association between genetic variation at the *LPA* locus and alcohol consumption. It is possible that that alcohol modifies the *LPA* expression for carriers of rs41267813, changing the Lp(a) lipoprotein concentrations and thereby influencing LDL-C levels. Unfortunately, the LDL-C measurement we used did not distinguish Lp(a) concentrations from the LDL-C levels, further experimental study is warranted to test such hypothesis. However, the observed modification effect should be interpreted with caution as LDL-C levels in ARIC was determined by the Friedewald formula, and this does not distinguish between

cholesterol derived from LDL and lipoprotein(a) and therefore represent the sum of cholesterol from both. It's possible that the observed association represents a relationship with lipoprotein(a) levels.

In addition to the variant described above, the other driving rare variant had not been previously associated with a lipid trait, rs41288783 (p.Pro994Leu), is a deleterious variant in *APOB* gene (*missense variant, MAF ~0.10%*). A previous study reported its existence in a patient who was clinically diagnosed as familial hypercholesterolaemia (FH) without a detectable mutation [35]. FH is characterized by very high levels of LDL-C, and we observed an association with higher LDL-C levels though jointly testing the effects of rs41288783 and its interaction with alcohol consumption. Nevertheless, the exact biological function of rs41288783 remains unknown. We note that a Mendelian randomization study has suggested a causal role of alcohol consumption in reducing plasma apo B and LDL-C levels in a general population [36]. Considering this, alcohol consumption may have contributed to the observed significant joint effect of *APOB* and alcohol consumption on LDL-C levels. It is also worth noting that these two novel rare driving variants showed 4 to 16 times larger main effect sizes on LDL-C levels as compared to the effect sizes of previously identified common variants (rs1367117 and rs1564348) [3]. Such observations supported the hypothesis of rare alleles of large effect [37] and pinpointed the importance of analyzing rare variant G×E interactions.

For the significant gene-alcohol interaction effect we observed on TG levels, the driving variant was identified as rs142488686, a missense mutation in *SMC5* (Structural Maintenance Of Chromosomes 5). *SMC5* encodes a core component involved in repair of DNA double-strand breaks and required for telomere maintenance [38-40]. Variants in *SMC5* have been previously reported to be associated with body mass index in a Japanese population [41], but not with lipid

levels nor alcohol consumption, and it is unknown whether the interaction we observed between *SMC5* locus and current drinking behavior on TG levels has a biological aspect. As the gene level interaction test results just missed Bonferroni corrected significance level for replication, further studies are warranted to validate such findings.

A limitation of this study is the imbalance in percentage of alcohol consumers between discovery (on average 48.7% regular drinker, 78.5% current drinker) and replication studies (on average 29.8% regular drinker, 57.2% current drinker) which may have impacted our ability to identify and replicate additional loci beyond what is reported here. All participating studies used similar questionnaires (either interviewer-administered or self-reported, Supplemental Table 1) to collect alcohol consumption information. The variation in percentage of current and regular drinkers may represent the heterogeneity of drinking behaviors across populations, and therefore may contribute to the generalizability of our results. Additionally, as self-reported alcohol consumption was used and may very likely be underreported, this study may suffer from loss of statistical power due to potential misclassification [42]. Similarly, dichotomizing alcohol consumption into regular drinkers and current drinkers may also reduce power as compared to treating it as a continuous variable [43]. It is possible that a more comprehensive characterization of alcohol consumption could reveal associations that were missed in the present study. In addition, although the sample size of 66,428 may provide sufficient power for a traditional GWAS, on the identification of rare variants and gene-environment interactions may require even larger sample sizes or bigger effect sizes [17, 44].

In conclusion, this study applied emerging statistical approaches to investigate the role of rare and low-frequency variants in gene-alcohol consumption interaction effects on lipid levels, and identified 2 novel rare variants at know lipid loci for LDL-C levels, with larger effect sizes

than those of the previously known common variants, and suggested 1 novel locus for gene-alcohol interaction on TG levels. Our results show promise for other larger scale studies analyzing rare variant G×E interactions to refine association signals at previously identified loci to reveal novel biology.

Sources of Funding: The various Gene-Lifestyle Interaction projects, including this one, are largely supported by a grant from the U.S. National Heart, Lung, and Blood Institute (NHLBI), the National Institutes of Health, R01HL118305. Paul S. de Vries is supported by American Heart Association Grant 18CDA34110116. Nora Franceschini is supported by the NIH grants R01-DK117445, R01-MD012765 and R21-HL140385. Full set of study-specific funding sources appear in the Supplemental Materials.

Disclosures: The authors declare no competing conflicts of interests except for the following. Dennis O Mook-Kanamori is a part-time research consultant with Metabolon, Inc; and Bruce M Psaty serves on the Steering Committee of the Yale Open Access Project funded by Johnson & Johnson.

References:

1. Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: The scandinavian simvastatin survival study (4s). *Lancet*. 1994;344:1383-1389

2. American College of Physicians. Clinical guideline, part 1: Guidelines for using serum cholesterol, high-density lipoprotein cholesterol, and triglyceride levels as screening tests for preventing coronary heart disease in adults. *Ann Intern Med*. 1996;124:515-517..

3. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010;466:707-713.

4. Asselbergs FW, Guo Y, van Iperen EP, Sivapalaratnam S, Tragante V, Lanktree MB, Lange LA, Almoguera B, Appelman YE, Barnard J, et al. Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am J Hum Genet*. 2012;91:823-838.

5. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274-1283.

6. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitziel NO, Brody JA, Khetarpal SA, Crosby JR, Fornage M, et al. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet*. 2014;94:223-232.

7. Surakka I, Horikoshi M, Magi R, Sarin AP, Mahajan A, Lagou V, Marullo L, Ferreira T, Miraglio B, Timonen S, et al. The impact of low-frequency and rare variants on lipid levels. *Nat Genet*. 2015;47:589-597.

8. Tang CS, Zhang H, Cheung CY, Xu M, Ho JC, Zhou W, Cherny SS, Zhang Y, Holmen O, Au KW, et al. Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in chinese. *Nat Commun*. 2015;6:10206.

9. Muth ND, Laughlin GA, von Muhlen D, Smith SC, Barrett-Connor E. High-density lipoprotein subclasses are a potential intermediary between alcohol intake and reduced risk of cardiovascular disease: The rancho bernardo study. *Br J Nutr*. 2010;104:1034-1042.

10. Gardner CD, Tribble DL, Young DR, Ahn D, Fortmann SP. Associations of hdl, hdl(2), and hdl(3) cholesterol and apolipoproteins a-i and b with lifestyle factors in healthy women and men: The stanford five city project. *Prev Med*. 2000;31:346-356.

11. Wakabayashi I. Relationship between alcohol intake and lipid accumulation product in middle-aged men. *Alcohol Alcohol*. 2013;48:535-542.

12. Whitfield JB, Heath AC, Madden PA, Pergadia ML, Montgomery GW, Martin NG. Metabolic and biochemical effects of low-to-moderate alcohol consumption. *Alcohol Clin Exp Res*. 2013;37:575-586.

13. Rakic V, Puddey IB, Dimmitt SB, Burke V, Beilin LJ. A controlled trial of the effects of pattern of alcohol intake on serum lipid levels in regular drinkers. *Atherosclerosis*. 1998;137:243-252.

14. Brien SE, Ronksley PE, Turner BJ, Mukamal KJ, Ghali WA. Effect of alcohol consumption on biological markers associated with risk of coronary heart disease: Systematic review and meta-analysis of interventional studies. *BMJ*. 2011;342:d636.

15. Rimm EB, Williams P, Fosher K, Criqui M, Stampfer MJ. Moderate alcohol intake and lower risk of coronary heart disease: Meta-analysis of effects on lipids and haemostatic factors. *BMJ*. 1999;319:1523-1528.

16. Thomas D. Gene–environment-wide association studies: Emerging approaches. *Nat Rev Genet*. 2010;11:259-272.

17. Hunter DJ. Gene–environment interactions in human diseases. *Nat Rev Genet.* 2005;6:287-298.

18. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, Gauderman WJ, Rice K, Munroe PB, Psaty BM, et al. Multiancestry study of gene-lifestyle interactions for cardiovascular traits in 610 475 individuals from 124 cohorts: Design and rationale. *Circ Cardiovasc Genet*. 2017;10.

19. Feitosa MF, Kraja AT, Chasman DI, Sung YJ, Winkler TW, Ntalla I, Guo X, Franceschini N, Cheng C-Y, Sim X, et al. Novel genetic associations for blood pressure identified via gene-alcohol interaction in up to 570k individuals across multiple ancestries. *PLOS ONE*. 2018;13:e0198166.

20. Kilpeläinen TO, Bentley AR, Noordam R, Sung YJ, Schwander K, Winkler TW, Jakupović H, Chasman DI, Manning A, Ntalla I, et al. Multi-ancestry study of blood lipid levels identifies four loci interacting with physical activity. *Nat Commun.* 2019;10:376.

21. de Vries PS, Brown MR, Bentley AR, Sung YJ, Winkler TW, Ntalla I, Schwander K, Kraja AT, Guo X, Franceschini N, et al. Multi-ancestry genome-wide association study of lipid levels incorporating gene-alcohol interactions. *Am J Epidemiol.* 2019;188:1033-1054.

22. Sung YJ, Winkler TW, de Las Fuentes L, Bentley AR, Brown MR, Kraja AT, Schwander K, Ntalla I, Guo X, Franceschini N, et al. A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure. *Am J Hum Genet*. 2018;102:375-400.

23. Kraft P, Yen Y-C, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered.* 2007;63:111-119.

24. Manning AK, LaValley M, Liu CT, Rice K, An P, Liu Y, Miljkovic I, Rasmussen-Torvik L, Harris TB, Province MA, et al. Meta-analysis of gene-environment interaction: Joint estimation of snp and snp x environment regression coefficients. *Genet Epidemiol*. 2011;35:11-18.

25. Chen H, Meigs JB, Dupuis J. Incorporating gene-environment interaction in testing for association with rare genetic variants. *Hum Hered.* 2014;78:81-90.

26. Cohen JC, Boerwinkle E, Mosley TH, Jr., Hobbs HH. Sequence variations in pcsk9, low ldl, and protection against coronary heart disease. *N Engl J Med*. 2006;354:1264-1272.

27. Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC. Population-based resequencing of angptl4 uncovers variations that reduce triglycerides and increase hdl. *Nat Genet*. 2007;39:513-516.

28. Kanoni S, Masca NG, Stirrups KE, Varga TV, Warren HR, Scott RA, Southam L, Zhang W, Yaghootkar H, Muller-Nurasyid M, et al. Analysis with the exome array identifies multiple new independent variants in lipid loci. *Hum Mol Genet*. 2016;25:4094-4106.

29. Jorgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A. Loss-of-function mutations in apoc3 and risk of ischemic vascular disease. *N Engl J Med*. 2014;371:32-41.

30. Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitziel NO, Lange LA, Lu Y, Tang ZZ, Zhang H, Hindy G, et al. Loss-of-function mutations in apoc3, triglycerides, and coronary disease. *N Engl J Med*. 2014;371:22-31.

31. Crawford DC, Dumitrescu L, Goodloe R, Brown-Gentry K, Boston J, McClellan B, Jr., Sutcliffe C, Wiseman R, Baker P, Pericak-Vance MA, et al. Rare variant apoc3 r19x is associated with cardio-protective profiles in a diverse population-based survey as part of the epidemiologic architecture for genes linked to environment study. *Circ Cardiovasc Genet.* 2014;7:848-853.

32. Langsted A, Kamstrup PR, Benn M, Tybjærg-Hansen A, Nordestgaard BG. High lipoprotein(a) as a possible cause of clinical familial hypercholesterolaemia: A prospective cohort study. *Lancet Diabetes Endocrinol*. 2016;4:577-587.

33. Sharpe PC, Young IS, Evans AE. Effect of moderate alcohol consumption on lp(a) lipoprotein concentrations: Reduction is supported by other studies. *BMJ*. 1998;316:1675-1675.

34. Paassilta M, Kervinen K, Rantala AO, Savolainen MJ, Lilja M, Reunanen A, Kesaniemi YA. Social alcohol consumption and low lp(a) lipoprotein concentrations in middle aged finnish men: Population based study. *BMJ*. 1998;316:594-595.

35. Alves AC, Etxebarria A, Soutar AK, Martin C, Bourbon M. Novel functional apob mutations outside ldl-binding region causing familial hypercholesterolaemia. *Hum Mol Genet.* 2014;23:1817-1828.

36. Vu KN, Ballantyne CM, Hoogeveen RC, Nambi V, Volcik KA, Boerwinkle E, Morrison AC. Causal role of alcohol consumption in an improved lipid profile: The atherosclerosis risk in communities (aric) study. *PloS One*. 2016;11:e0148765-e0148765.

37. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. Shifting paradigm of association studies: Value of rare single-nucleotide polymorphisms. *Am J Hum Genet*.. 2008;82:100-112.

38. Potts PR, Yu H. The smc5/6 complex maintains telomere length in alt cancer cells through sumoylation of telomere-binding proteins. *Nat Struct Mol Biol* .2007;14:581-590.

39. Behlke-Steinert S, Touat-Todeschini L, Skoufias DA, Margolis RL. Smc5 and mms21 are required for chromosome cohesion and mitotic progression. *Cell Cycle.* 2009;8:2211-2218.

40. Jeppsson K, Kanno T, Shirahige K, Sjogren C. The maintenance of chromosome structure: Positioning and functioning of smc complexes. *Nat Rev Mol Cell Biol.* 2014;15:601-614.

41. Okada Y, Kubo M, Ohmiya H, Takahashi A, Kumasaka N, Hosono N, Maeda S, Wen W, Dorajoo R, Go MJ, et al. Common variants at cdkal1 and klf9 are associated with body mass index in east asian populations. *Nat Genet*. 2012;44:302-306.

42. Livingston M, Callinan S. Underreporting in alcohol surveys: Whose drinking is underestimated? *J Stud Alcohol Drugs*. 2015;76:158-164.

43. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*. 2006;332:1080.

44. VanderWeele TJ. Sample size and power calculations for additive interactions. *Epidemiol Methods.* 2012;1:159-188.

Table 1. Descriptive characteristics for discovery and replication studies

	Study*	Design	N	CurDrinker† (%)	RegDrinker† (%)
Discovery	ARIC	Unrelated	10989	64.9	36.8
	FHS	Family	7258	83.6	65.5
	NEO	Unrelated	5718	86.8	69.0
	WHI	Unrelated	8021	76.5	32.6
	CARDIA	Unrelated	2167	68.7	59.6
	Total/Average		34,153	75.5	48.7
Replication	WGHS	Unrelated	22478	56.7	29.3
	CFS	Family	253	50.2	25.1
	CHS	Unrelated	3690	53.8	25.0
	FamHS	Family	1735	50.7	28.3
	GENOA	Family	1543	53.1	29.2
	MESA	Unrelated	2578	71.9	43
	Total/Average		32,277	57.2	29.8
Overall			66,430	66.6	39.5

*ARIC: the Atherosclerosis Risk in Communities study; CARDIA: the Coronary Artery Risk Development in Young Adults study; FHS: the Framingham Heart Study; NEO: the Netherlands Epidemiology of Obesity study; WHI: the Women's Health Initiative study; CFS: the Cleveland Family Study, CHS: the Cardiovascular Health Study; FamHS: the Family Heart Study; GENOA: the Genetic Epidemiology Network of Arteriopathy study; MESA: the Multi-Ethnic Study of Atherosclerosis; and WGHS: the Women's Genome Health Study

† CurDrinker and RegDrinker represents the current and regular drinkers, respectively

Table 2. Genes discovered and replicated by the joint test or interaction only test

Trait	Gene	CHR	Alcohol*	Test	N.discovery†	cMAF Range‡	*p*.discovery	N.replication†	*p*.replication
	LPL	8	Both	Joint	5	0.036 - 0.040	8.76E-22	5	4.25E-21
	APOC3	11	Both	Joint	3	0.001 - 0.001	2.82E-06	2	4.62E-06
HDL-C	*CD300LG*	17	Both	Joint	5	0.031 - 0.055	2.64E-12	6	5.94E-10
	LIPG	18	Both	Joint	5	0.014 - 0.019	7.65E-17	5	4.09E-11
	ANGPTL4	19	Both	Joint	5	0.024 - 0.031	2.34E-20	5	5.53E-09
	HNF4A	20	Both	Joint	5	0.031 - 0.034	3.37E-10	5	3.20E-07
	CELSR2	1	Both	Joint	5	0.079 - 0.093	1.63E-10	6	3.21E-08
	MYBPHL	1	Both	Joint	5	0.044 - 0.051	7.26E-09	6	6.49E-06
	PCSK9	1	Both	Joint	5	0.050 - 0.055	3.16E-62	6	9.06E-11
LDL-C	*APOB*	2	Both	Joint	5	0.174 - 0.226	5.33E-18	6	1.20E-15
	LPA	6	RegDrink	Joint	5	0.096 - 0.147	2.28E-05	6	3.7E-04
	APOH	17	Both	Joint	5	0.074 - 0.081	1.11E-05	6	1.18E-05
	BCAM	19	Both	Joint	5	0.120 - 0.166	1.49E-18	6	1.77E-37
	CBLC	19	Both	Joint	5	0.084 - 0.104	7.48E-22	6	1.64E-35
	LPL	8	Both	Joint	5	0.036 - 0.040	8.55E-19	5	7.30E-16
	APOA4	11	Both	Joint	5	0.019 - 0.024	8.83E-09	6	3.77E-09
	APOA5	11	Both	Joint	5	0.025 - 0.033	8.93E-07	5	2.3E-04
TG	*APOC3*	11	Both	Joint	3	0.001 - 0.001	2.09E-10	3	7.92E-08
	MAP1A	15	Both	Joint	5	0.129 - 0.166	1.70E-06	6	4.30E-05
	CD300LG	17	Both	Joint	5	0.031 - 0.055	1.39E-09	6	5.26E-08
	ANGPTL4	19	Both	Joint	5	0.024 - 0.031	1.33E-24	5	3.56E-15
	SMC5	9	CurDrink	Interaction	4	0.001 - 0.002	6.65E-06	4	0.013§

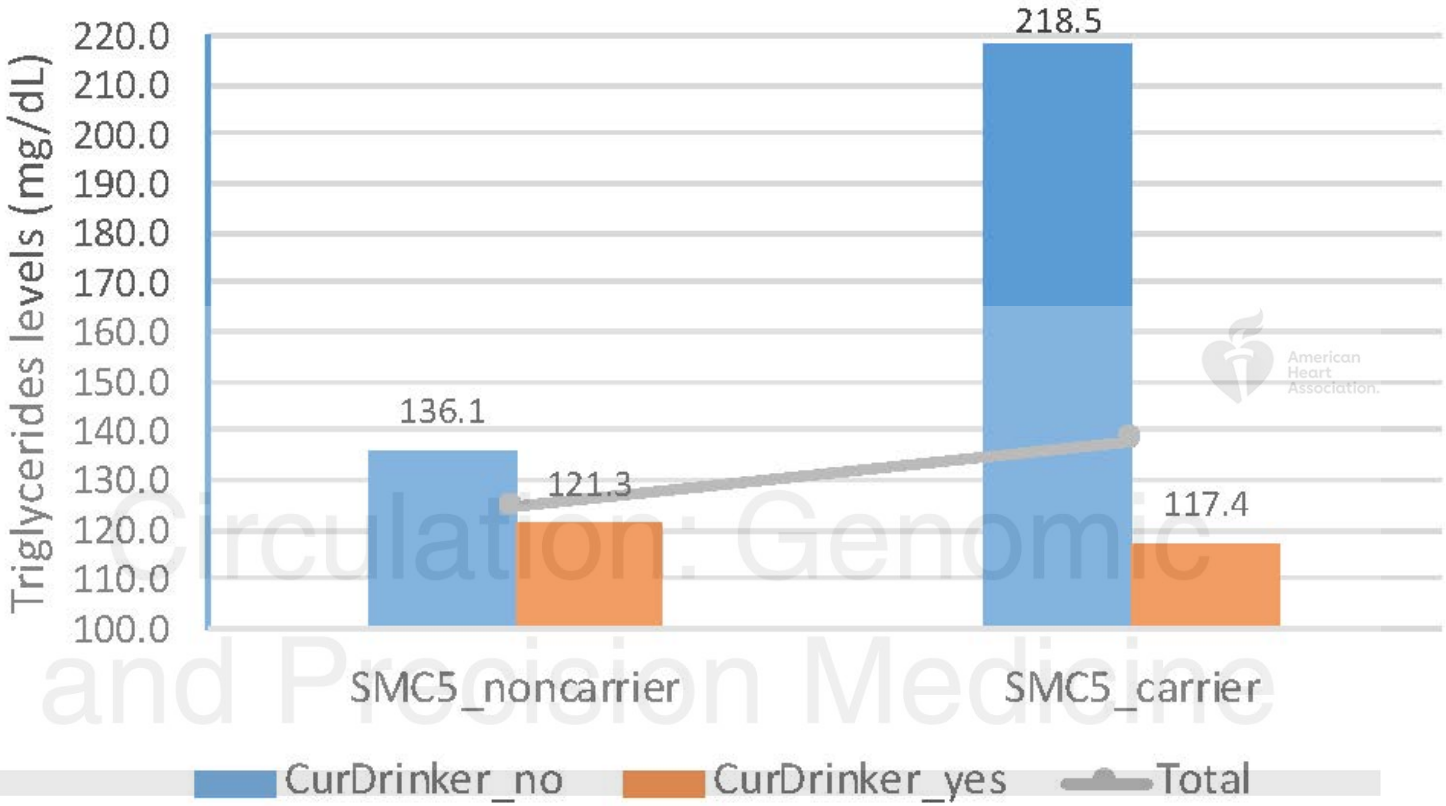
* Both indicates the gene-lipid pair was identified through using both current and regular drinker as the alcohol consumption variable.
† N.discoery and N.replication represent the number of studies included in the respective meta-analyses.
‡ cMAF Range represents the cumulative minor allele frequency for variants aggregated in the genes across studies involved in discovery phase for that gene.
§ *SMC5*-current drinking interaction on TG levels just missed the Bonferroni corrected threshold of significance ($p = 0.0125$) for replication but reached nominal significance.
Significant threshold for replication was set as $p < 0.0017$ for joint test and $p < 0.0125$ for interaction test using Bonferroni correction.

Figure Legends:

Figure 1. *SMC5*, current alcohol consumption and average TG levels across four discovery studies: the Atherosclerosis Risk in Communities study; the Framingham Heart Study; the Netherlands Epidemiology of Obesity study; the Women's Health Initiative study

Figure 2. Genes as revealed by G×E interaction test or jointly testing the gene and G×E interaction effects in association with plasma lipid levels. **Bolded** genes were genes remained significant after conditioning on common index SNPs. Genes in red were not previously reported to be associated with one or more lipid traits but they are in known lipid loci

Figure 3. Flowchart of follow-up analyses, including conditional analysis and single variant test to identify driving rare variants**.** For conditional analysis, significant results were defined as *p*-value < 5×10^{-5} in meta-analysis of discovery studies, and *p*-value < 0.05/10 (Bonferroni correction for 10 gene-lipid pairs with *p*-value < 5×10^{-5} in discovery phase) in meta-analysis of replication studies. For single variant test to identify driving rare variants, we included variants with minor allele count at least 5 and present in at least 2 studies. Bonferroni correction for number of SNPs tested in discovery phase and number of SNPs taken forward to replication were applied separately for joint test and interaction test for each lipid trait.



Sample size	CurDrinker_no	CurDrinker_yes
SMC5_noncarrier	7,448	23,636
SMC5_carrier	30	117