bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

Short running head: Rare Genetic Variants and Alcohol on Lipids

Zhe Wang, Han Chen, Traci M. Bartz, Lawrence F. Bielak, Daniel I. Chasman, Mary F. Feitosa, Nora Franceschini, Xiuqing Guo, Elise Lim, Raymond Noordam, Melissa A. Richard, Heming Wang, Brian Cade, L. Adrienne Cupples, Paul S. de Vries, Franco Giulanini, Jiwon Lee, Rozenn N. Lemaitre, Lisa W. Martin, Alex P. Reiner, Stephen S. Rich, Pamela J. Schreiner, Stephen Sidney, Colleen M. Sitlani, Jennifer A. Smith, Ko Willems van Dijk, Jie Yao, Wei Zhao, Myriam Fornage, Sharon L.R. Kardia, Charles Kooperberg, Ching-Ti Liu, Dennis O Mook-Kanamori, Michael A. Province, Bruce M. Psaty, Susan Redline, Paul M. Ridker, Jerome I. Rotter, Eric Boerwinkle, Alanna C. Morrison, On behalf of the CHARGE Gene-Lifestyle Interactions Working Group.

Affiliations

Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX (Z.W., H.C., P.S.dV., M.F., E.B., A.C.M.); Center for Precision Health, School of Public Health & School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX (H.C.); Cardiovascular Health Research Unit, Department of Biostatistics, University of Washington, Seattle, WA (T.B.); School of Public Health, Department of Epidemiology, University of Michigan, Ann Arbor, MI (L.F.B, J.A.F, W.Z, S.L.R.K.); Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA (D.I.C., F.G., P.M.R.); Harvard Medical School, Boston, MA (D.I.C., H.W., B.C., J.L., S.R., P.M.R.); Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Louis, MO (M.F.F., M.A.P); Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC (N.F.); The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA (X.G., J.Y, J.I.R.); Biostatistics Department, Boston University School of Public Health, Boston, MA (E.L., L.A.C., C.L.); Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden (R.N.); Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX (M.A.R., M.F.); Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA (H.W., B.C., J.L, S.R.); NHLBI Framingham Heart Study, Framingham, MA (L.A.C); Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA (R.N.L, C.M.S., B.M.P.); George Washington University School of Medicine and Health Sciences, Washington, DC (L.W.M.); Fred Hutchinson Cancer Research Center, Seattle, WA (A.P.R., C.K.); Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA (S.S.R.); Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN (P.J.S.); Division of Research, Kaiser Permanente of Northern California, Oakland, CA (S.S.); Institute for Social Research, Survey Research Center, University of Michigan, Ann Arbor, MI (J.A.S., W.Z.); Department of Human Genetics, Leiden University Medical Center, Leiden (K.W.vD.); Division of Endocrinology, Department of Internal Medicine, Leiden University Medical Center, Leiden (K.W.vD.); Department of Clinical Epidemiology, Leiden University Medical Center, Leiden (D.O.M.); Department of Public Health and Primary Care, Leiden University Medical Center, Leiden (D.O.M.); Cardiovascular Health Research Unit, Department

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of Epidemiology, University of Washington, Seattle, WA (B.M.P); Cardiovascular Health Research Unit, Department of Health Services, University of Washington, Seattle, WA (B.M.P.); Kaiser Permanente Washington Health Research Institute, Seattle, WA (B.M.P.); Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX (E.B.)

Corresponding author:

Alanna C. Morrison

RAS E447

1200 Pressler St

Houston, TX, 77030

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

Tel: 713-500-9913

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract (current word count 268, AJCN limit: 300):

Background: Alcohol intake influences plasma lipid levels and such effects may be modulated by genetic variants.

Objective: We aimed to characterize the role of aggregated rare and low-frequency variants in gene by alcohol consumption interactions associated with fasting plasma lipid levels.

Design: 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 European Americans from five discovery studies and 32,275 individuals from six replication studies. Rare and low-frequency protein coding variants (minor allele frequency ≤ 5%) measured by an exome array were aggregated by genes and evaluated by a gene-environment interaction (GxE) test and a joint test of genetic main and GxE 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 was discovered at a Bonferroni corrected significance level (p-value $<5\times10^{-5}$) and replicated (p-value <0.013 for the interaction test) in *SMC5*.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Conclusions: In conclusion, this study applied new gene-based statistical approaches to uncover the role of rare and low-frequency variants in gene-alcohol consumption interactions on lipid levels.

Keywords: gene-environment interactions, lipid levels, alcohol consumption, genome-wide association study, rare variant test

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

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 (3-9), 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 (10, 11). However, the association between alcohol use and LDL-C or TG levels is unclear. Some studies reported positive associations while others reported negative associations (12-21).

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 (22-24). 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 (25-34). These studies have successfully identified common variant loci that were not detected via analysis of main effects alone. However, unlike well-established G×E interaction tests for common variants (35, 36), methods for detecting rare variant G×E interactions are emerging. Recently developed novel approaches for testing rare variant G×E

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 GxE interactions (37).

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

Overview of participating studies

This study includes 66,428 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 (24). The participating studies include the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in Young Adults (CARDIA) study, the Framingham Heart Study (FHS), the Netherlands Epidemiology of Obesity (NEO) study, the Women's Health Initiative (WHI) study, the Cleveland Family Study (CFS), the Cardiovascular Health Study (CHS), the Family Heart Study (FamHS), the Genetic Epidemiology Network of Arteriopathy (GENOA) study, the Multi-Ethnic Study of Atherosclerosis (MESA), and the Women's Genome Health Study (WGHS). Additional detail for these studies is provided in the Supplemental Materials. Each study obtained informed consent from participants and approval from the appropriate institutional review boards. A total of 34,153 participants from five studies participated in the discovery phase (ARIC, FHS, NEO, WHI and CARDIA), and six studies

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

73 involving 32,275 participants were used for replication (WGHS, CFS, CHS, FamHS, GENOA
74 and MESA).

75

76 *Plasma lipids and alcohol consumption*

77 Three fasting (≥ 8 hours) lipid measures were analyzed separately. HDL-C and TG were
78 directly assayed, while LDL-C was either directly assayed (WGHS, FamHS if TG > 400 mg/dL)
79 or estimated using the Friedewald equation (38) (ARIC, FHS, NEO, WHI, CARDIA, CFS, CHS,
80 FamHS, GENOA, MESA) in samples with TG ≤ 400 mg/dL. LDL-C levels were adjusted for
81 use of statins: if LDL-C levels were directly assayed, LDL-C levels were adjusted for lipid-
82 lowering medication use by dividing the original levels by 0.7, otherwise, LDL-C levels were
83 adjusted by first dividing total cholesterol by 0.8, and then using the corrected total cholesterol
84 level in the Friedewald equation. When information on statin-specific use was unavailable, LDL-
85 C levels were adjusted for use of unspecified lipid-lowering medication, but only if lipid
86 measurements were performed after 1994. Due to their skewed distributions, HDL-C and TG
87 were natural log transformed prior to analyses.

88 Alcohol consumption was assessed using two dichotomized self-reported alcohol
89 consumption variables: “current drinker” status, defined as any recurrent drinking behavior, and
90 “regular drinker” status, as the subset of current drinkers who consume at least two drinks per
91 week (33). For this study, definition of “a drink” is approximately 13g of pure ethanol, and this
92 measure was used to standardize the definitions across studies.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genotyping and quality control

Genotyping was performed using the Illumina or Affymetrix Human Exome array v1 or v1.1. To improve accurate calling of rare variants, genotyped data from 10 CHARGE Consortium studies were jointly called (39). Using the curated clustering files from the CHARGE joint calling effort, several cohorts within our study re-called their genotypes. For the remainder of participating studies, genotypes were determined using either BeadStudio or Zcall (40). Detailed information regarding the genotyping platform for each study is presented in Supplemental Table 1. All studies performed the following sample-level quality control steps: call rate <95%, autosomal heterozygosity outliers, gender discordance, GWAS discordance (if GWAS data available), ethnic outlier in a principal components analysis. Variants were removed by filtering for Hardy-Weinberg equilibrium test *p*-value (pHWE) $< 5\times10^{-6}$, call rate <95%, and poorly clustering variants.

Study-specific association analyses

Statistical analyses were performed within each study using the gene-based rareGE R package (37), and were performed for each lipid/alcohol consumption combination for a total of six combinations. Two types of analyses were considered: 1) a GxE test that considers the genetic main effects as fixed/random effects, and 2) a joint analysis of the genetic main and the GxE interaction effects. Rare and low-frequency (MAF $\leq$ 5%) functional variants (i.e. frameshift, nonsynonymous, stop/gain, stop/loss, and splicing) were aggregated within genes. Genes with 0 or only 1 rare and low-frequency variant, or genes with a cumulative minor allele count $\leq$ 10 were not analyzed within each study. Models were adjusted for age, sex, principal components (PCs) and additionally study-specific covariates as presented in Supplemental Table 1.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Meta-analyses

A weighted Z-test using square root of sample sizes as weights was used to meta-analyze study-specific *p*-values for genes present in at least 2 discovery studies (41). Genes of interest from the discovery phase with a *p*-value $< 5\times10^{-5}$ were pursued for replication. For these select genes, we used the same approaches as in discovery studies to perform meta-analysis of the replication studies. Significance was determined using a Bonferroni correction for the number of gene-lipid pairs taken forward to replication (*p*-value $< 0.05/30 = 0.0017$ for analysis of the joint test of genetic main and interaction effects, *p*-value $<0.05/4 = 0.013$ for analysis of the interaction effects).

Additional analyses: conditional and single variant tests

For each replicated gene-lipid pair, additional analyses were conducted following the flowchart shown in **Figure 1**. For genes +/-500kb bp from previously reported lipid loci (42), conditional analyses were performed to identify aggregated rare and low-frequency variants associated with lipids independent of the previously reported common index single nucleotide polymorphism (SNP) (42). Results from study-specific conditional analyses were meta-analyzed using a weighted Z-test, separately in discovery and replication. For novel genes and known genes that remained significant after conditional analyses, we performed single variant tests for each variant (MAF $\leq$ 5% and minor allele count $\geq$ 5) that was included in the aggregate test in order to identify the driving variants within these genes. We obtained robust estimates of covariance matrices and robust standard errors from each study and implemented METAL to jointly meta-analyze the genetic main and interaction effects (36, 43), and to meta-analyze the interaction coefficients alone using inverse-variance weighted meta-analysis for each single variant within selected genes.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

Descriptive statistics for the discovery and 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 for the discovery studies as compared to the replication studies.

Overall, meta-analysis 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 1. 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 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 gene-alcohol interaction (Supplemental Table 2). Thirteen known lipid loci (21 gene-lipid associations) were replicated and one novel interaction were replicated for the *SMC5*-by-current drinker interaction on TG levels (**Table 2**). 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, we performed conditional analyses in order to examine whether the gene-based rare variant effects 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, genes at known lipid

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

loci yet these genes themselves were not previously reported to be associated with lipids, such as *BCAM* and *CBLC* on LDL-C, were strongly attenuated after adjusting for rs7412, the index SNPs of *APOE* identified by previous GWAS (Supplemental Table 3).

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 1,** 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). For the novel interaction between *SMC5* and current drinker on TG levels, we identified the driving variant as rs142488686, a missense mutation (MAC = 5-7 discovery (ARIC and CARDIA), MAC = 7-17 replication (WGHS, CHS and MESA)), with positive interaction effect ($p = 0.016$ discovery, $p = 0.008$ 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 effect on lipid levels as well as the joint effects of genetic main and gene-alcohol

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

185 interactions. We replicated 13 gene-lipid associations at known lipid loci, among which 2
186 leading rare variants in *APOB* and *LPA* genes associated with LDL-c were novel. Only one novel
187 gene-alcohol interaction was identified as significant and successfully replicated (the interaction
188 between rare and low-frequency variants in *SMC5* and current drinker on TG levels).

189 Using a single variant test, we confirmed previously identified rare and low-frequency
190 lipid variants. For example, rs11591147 of *PCSK9* has been associated with LDL-c levels (44),
191 rs77960347 of *LIPG* and rs116843064 of *ANGPTL4* have been associated with HDL-c levels
192 (45, 46). A loss of function mutation in the *APOC3* gene, rs147210663, has been associated with
193 a more than 40% lower average triglyceride level in individuals carrying one A allele (47, 48). In
194 the present study, we observed a novel relationship between increased HDL-c levels in
195 individuals carrying rs147210663 (A) allele as rs147210663 was previously reported as a
196 founder mutation in a Pennsylvania Amish population (49).

197 Between the two novel rare driving variants we identified and replicated, rs41267813
198 (*LPA)* is located close to a stop/gain variant rs41267811 *(LPA)* that was also significantly
199 associated with LDL-c levels in the discovery phase. However, we were unable to replicate the
200 association with rs41267811 as it was only available in one replication study (WGHS) and
201 therefore did not meet our criteria to be included in replication. *LPA* encoded protein constitutes
202 a substantial portion of lipoprotein(a) and associated with inherited conditions including type III
203 hyperlipoproteinemia and familial hyperlipidemia (50). A stop/gain mutation in this gene would
204 be associated with lower LDL-C levels in carriers, which is true among non-drinkers. However,
205 such effect may be modified by alcohol consumption as we observed the carriers of this variant
206 with a higher LDL-C levels compared to non-carriers in a population who had at least two drinks
207 per week in the ARIC study. Previous studies have reported a relationship between moderate

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

alcohol consumption and lower Lp(a) lipoprotein concentrations (51, 52), but there was no existing evidence linking genetic variants of *LPA* to alcohol consumption. Although the underlying biology of the observed modification effect of alcohol consumption on rs41267813 and LDL-c associations remained unclear, we hypothesize that alcohol modifies the *LPA* expression for carriers of rs41267813, therefore modifies LDL-c 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), a deleterious variant in *APOB* gene. A previous study reported its existence in a patient who was clinically diagnosed as familial hypercholesterolaemia (FH) without a detectable mutation (53). 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 (54). Considering this, alcohol consumption may have contributed to the observed significant joint effect of *APOB* and alcohol consumption on LDL-c levels.

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 (55-57). Variants in *SMC5* have not been previously reported to be associated with lipid levels nor alcohol consumption, and it is unknown whether the interaction between *SMC5* locus and current drinking behavior on TG levels has a biological aspect.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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. 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 (58). Similarly, dichotomizing alcohol consumption into regular drinkers and current drinkers may also reduce power as compared to treating it as a continuous variable (59). 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 seem sufficient for a traditional GWAS, to identify additional novel loci while focusing on rare variants and gene-environment interactions may require larger sample size or bigger effect size (23, 60).

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 and 1 novel gene-alcohol interaction for TG levels. Our results show promise for other larger scale studies analyzing rare variant GxE interactions to refine association signals at previously identified loci to reveal novel biology.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgments

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. Full set of study-specific funding sources and acknowledgments appear in the Supplemental Materials.

Conflicts of Interests

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.

Authors' Contributions

The authors' contributions were as follows – ZW, DIC, RN, HW, LAC, KWvD., JAS, SSR, MF, SLRK, CL, DOM, MAP, PMR, JIR, EB, and ACM: designed research (project conception, development of overall research plan, and study oversight); LFB, DIC, MFF, NF, XG, BC, KWvD, JL, LWM, APR, SSR, PJS, SS, JAS, WZ, SLRK, CK, DOM, MAP, BMP, SR, PMR, and JIR: conducted research (hands-on conduct of the experiments and data collection) and provided essential reagents or provided essential materials (applies to authors who contributed by providing animals, constructs, databases, etc, necessary for the research); WZ, HC, TMB, LFB, DIC, MFF, NF, XG, EL, RN, MAR, HW, LAC, FG, KWvD, JL, JY, MF, SLRK, CL, DOM, BMP, and JIR: analyzed

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

data or performed statistical analysis; ZW and ACM: wrote the paper; and all authors: read, reviewed, and approved of the final manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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). The Lancet 1994;344(8934):1383-9.
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. Annals of Internal Medicine 1996;124(5):515-7. doi: 10.7326/0003-4819-124-5-199603010-00012.
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(7307):707-13. doi: 10.1038/nature09270.
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(5):823-38. doi: 10.1016/j.ajhg.2012.08.032.
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(11):1274-83. doi: 10.1038/ng.2797.
6. Albrechtsen A, Grarup N, Li Y, Sparso T, Tian G, Cao H, Jiang T, Kim SY, Korneliussen T, Li Q, et al. Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. Diabetologia 2013;56(2):298-310. doi: 10.1007/s00125-012-2756-1.
7. 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. American journal of human genetics 2014;94(2):223-32.
8. 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(6):589-97. doi: 10.1038/ng.3300.
9. 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. doi: 10.1038/ncomms10206.
10. 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(7):1034-42. doi: 10.1017/S0007114510001595.
11. 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(4):346-56. doi: 10.1006/pmed.2000.0716.
12. Wakabayashi I. Relationship between alcohol intake and lipid accumulation product in middle-aged men. Alcohol Alcohol 2013;48(5):535-42. doi: 10.1093/alcalc/agt032.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

13. Klop B, do Rego AT, Cabezas MC. Alcohol and plasma triglycerides. Curr Opin Lipidol 2013;24(4):321-6. doi: 10.1097/MOL.0b013e3283606845.
14. 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(4):575-86. doi: 10.1111/acer.12015.
15. Rifler JP, Lorcerie F, Durand P, Delmas D, Ragot K, Limagne E, Mazue F, Riedinger JM, d'Athis P, Hudelot B, et al. A moderate red wine intake improves blood lipid parameters and erythrocytes membrane fluidity in post myocardial infarct patients. Mol Nutr Food Res 2012;56(2):345-51. doi: 10.1002/mnfr.760.
16. Brinton EA. Effects of ethanol intake on lipoproteins. Curr Atheroscler Rep 2012;14(2):108-14. doi: 10.1007/s11883-012-0230-7.
17. Kechagias S, Zanjani S, Gjellan S, Leinhard OD, Kihlberg J, Smedby O, Johansson L, Kullberg J, Ahlstrom H, Lindstrom T, et al. Effects of moderate red wine consumption on liver fat and blood lipids: a prospective randomized study. Ann Med 2011;43(7):545-54. doi: 10.3109/07853890.2011.588246.
18. Wakabayashi I. Associations of alcohol drinking and cigarette smoking with serum lipid levels in healthy middle-aged men. Alcohol Alcohol 2008;43(3):274-80. doi: 10.1093/alcalc/agn005.
19. 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(2):243-52.
20. 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. doi: 10.1136/bmj.d636.
21. 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(7224):1523-8.
22. Thomas D. Gene–environment-wide association studies: emerging approaches. Nature Reviews Genetics 2010;11(4):259-72.
23. Hunter DJ. Gene–environment interactions in human diseases. Nature Reviews Genetics 2005;6(4):287-98.
24. 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(3). doi: 10.1161/CIRCGENETICS.116.001649.
25. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, Rybin D, Liu C-T, Bielak LF, Prokopenko I, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nature genetics 2012;44(6):659-69. doi: 10.1038/ng.2274.
26. Winkler TW, Justice AE, Graff M, Barata L, Feitosa MF, Chu S, Czajkowski J, Esko T, Fall T, Kilpeläinen TO, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. PLOS Genetics 2015;11(10):e1005378. doi: 10.1371/journal.pgen.1005378.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

27. Simino J, Shi G, Bis Joshua C, Chasman Daniel I, Ehret Georg B, Gu X, Guo X, Hwang S-J, Sijbrands E, Smith Albert V, et al. Gene-Age Interactions in Blood Pressure Regulation: A Large-Scale Investigation with the CHARGE, Global BPgen, and ICBP Consortia. The American Journal of Human Genetics 2014;95(1):24-38. doi: https://doi.org/10.1016/j.ajhg.2014.05.010.
28. Taylor JY, Schwander K, Kardia SLR, Arnett D, Liang J, Hunt SC, Rao DC, Sun YV. A Genome-wide study of blood pressure in African Americans accounting for gene-smoking interaction. Scientific Reports 2016;6:18812. doi: 10.1038/srep18812.
29. Justice AE, Winkler TW, Feitosa MF, Graff M, Fisher VA, Young K, Barata L, Deng X, Czajkowski J, Hadley D, et al. Genome-wide meta-analysis of 241,258 adults accounting for smoking behaviour identifies novel loci for obesity traits. Nature Communications 2017;8:14977. doi: 10.1038/ncomms14977.
30. Taylor KC, Carty CL, Dumitrescu L, Bůžková P, Cole SA, Hindorff L, Schumacher FR, Wilkens LR, Shohet RV, Quibrera PM, et al. Investigation of gene-by-sex interactions for lipid traits in diverse populations from the population architecture using genomics and epidemiology study. BMC Genetics 2013;14(1):33. doi: 10.1186/1471-2156-14-33.
31. 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(6):e0198166. doi: 10.1371/journal.pone.0198166.
32. 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. Nature Communications 2019;10(1):376. doi: 10.1038/s41467-018-08008-w.
33. 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. American journal of epidemiology 2019. doi: 10.1093/aje/kwz005.
34. 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(3):375-400. doi: 10.1016/j.ajhg.2018.01.015.
35. Kraft P, Yen Y-C, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. Human heredity 2007;63(2):111-9.
36. Manning AK, LaValley M, Liu C-T, 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 × environment regression coefficients. Genetic Epidemiology 2011;35(1):11-8. doi: 10.1002/gepi.20546.
37. 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(2):223-32. doi: 10.1016/j.ajhg.2014.01.009.
38. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin Chem 1972;18(6):499-502.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

39. Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, Hansen M, Borecki IB, Cupples LA, Fornage M, et al. Best Practices and Joint Calling of the HumanExome BeadChip: The CHARGE Consortium. PLoS One 2013;8(7):e68095. doi: 10.1371/journal.pone.0068095.
40. Goldstein JI, Crenshaw A, Carey J, Grant GB, Maguire J, Fromer M, O'Dushlaine C, Moran JL, Chambert K, Stevens C, et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. Bioinformatics 2012;28(19):2543-5. doi: 10.1093/bioinformatics/bts479.
41. Zaykin DV. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. Journal of evolutionary biology 2011;24(8):1836-41. doi: 10.1111/j.1420-9101.2011.02297.x.
42. Liu DJ, Peloso GM, Yu H, Butterworth AS, Wang X, Mahajan A, Saleheen D, Emdin C, Alam D, Alves AC, et al. Exome-wide association study of plasma lipids in >300,000 individuals. Nature Genetics 2017;49:1758. doi: 10.1038/ng.3977

https://www.nature.com/articles/ng.3977#supplementary-information.
43. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190-1. doi: 10.1093/bioinformatics/btq340.
44. 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(12):1264-72. doi: 10.1056/NEJMoa054013.
45. 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(4):513-6. doi: 10.1038/ng1984.
46. 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(18):4094-106. doi: 10.1093/hmg/ddw227.
47. 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(1):32-41. doi: 10.1056/NEJMoa1308027.
48. 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(1):22-31. doi: 10.1056/NEJMoa1307095.
49. 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(6):848-53. doi: 10.1161/circgenetics.113.000369.
50. 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. The Lancet Diabetes & Endocrinology 2016;4(7):577-87. doi: 10.1016/S2213-8587(16)30042-0.
51. Sharpe PC, Young IS, Evans AE. Effect of moderate alcohol consumption on Lp(a) lipoprotein concentrations: Reduction is supported by other studies. BMJ : British Medical Journal 1998;316(7145):1675-.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

52. 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(7131):594-5.
53. Alves AC, Etxebarria A, Soutar AK, Martin C, Bourbon M. Novel functional APOB mutations outside LDL-binding region causing familial hypercholesterolaemia. Human Molecular Genetics 2014;23(7):1817-28. doi: 10.1093/hmg/ddt573.
54. 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(2):e0148765-e. doi: 10.1371/journal.pone.0148765.
55. Potts PR, Yu H. The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. Nature structural & molecular biology 2007;14(7):581-90. doi: 10.1038/nsmb1259.
56. Behlke-Steinert S, Touat-Todeschini L, Skoufias DA, Margolis RL. SMC5 and MMS21 are required for chromosome cohesion and mitotic progression. Cell cycle (Georgetown, Tex) 2009;8(14):2211-8. doi: 10.4161/cc.8.14.8979.
57. Jeppsson K, Kanno T, Shirahige K, Sjogren C. The maintenance of chromosome structure: positioning and functioning of SMC complexes. Nature reviews Molecular cell biology 2014;15(9):601-14. doi: 10.1038/nrm3857.
58. Livingston M, Callinan S. Underreporting in alcohol surveys: whose drinking is underestimated? J Stud Alcohol Drugs 2015;76(1):158-64.
59. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332(7549):1080. doi: 10.1136/bmj.332.7549.1080.
60. VanderWeele TJ. Sample Size and Power Calculations for Additive Interactions. Epidemiologic methods 2012;1(1):159-88. doi: 10.1515/2161-962X.1010.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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	3688	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		32,275	57.2	29.8
Overall			66,428	66.6	39.5

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

Trait	Gene	CHR	Alcohol[1]	Test	N.discovery[2]	cMAF Range[3]	*p*.discovery	N.replication[2]	*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	INT	4	0.001 - 0.002	6.65E-06	4	0.013

[1] Both indicates the gene-lipid pair was identified through using both current and regular drinker as the alcohol consumption variable.
[2] N.discoery and N.replication represent the number of studies included in the respective meta-analyses.
[3] cMAF Range represents the cumulative minor allele frequency for variants aggregated in the genes across studies involved in discovery phase for that gene.
Significant threshold for replication was set as $p < 0.0017$ for joint test and $p < 0.013$ for interaction test using Bonferroni correction.

bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

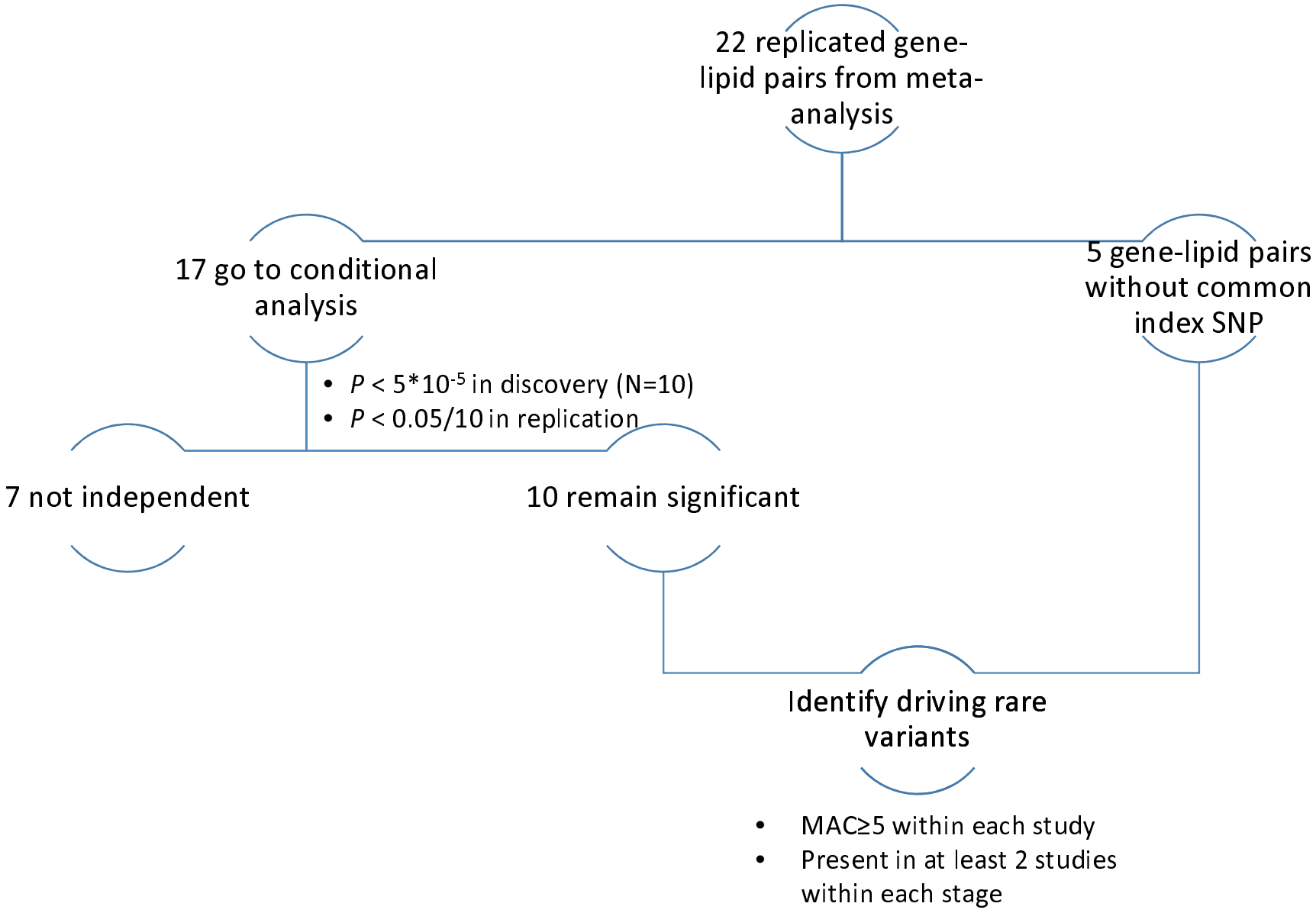
bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1. Flowchart of follow-up analyses, including conditional analysis and single variant test to identify driving rare variants

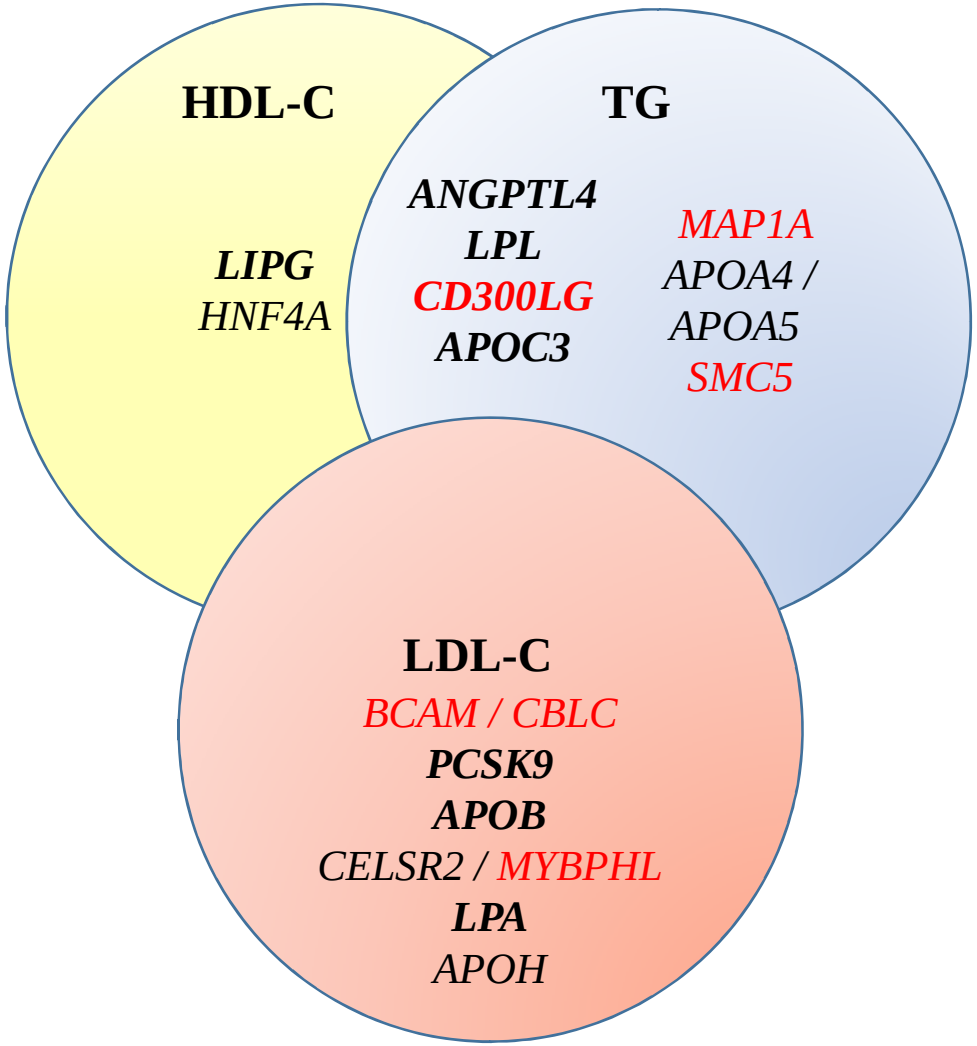
For conditional analysis, significant results were defined as *p*-value < 5×10^{-5} in meta-analysis of discovery studies, and *p*-value < 0.05/10 (Bonferroni correction for 10 gene-lipid pairs with *p*-value < 5×10^{-5} in discovery phase) in meta-analysis of replication studies.

For single variant test to identify driving rare variants, we applied Bonferroni correction for number of SNPs tested in discovery phase and number of SNPs taken forward to replication separately for joint test and interaction test for each lipid trait.

Figure 2. Genes as revealed by GxE interaction test or jointly testing the gene and GxE 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



bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



bioRxiv preprint doi: https://doi.org/10.1101/561225; this version posted February 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.