Assessing long-term pleiotropic effects of potential novel triglyceride-lowering medications using variants identified by Mendelian randomization

Elliot Outland, BS[1#]; Yi Xin[1,2#], MS; Alyson L. Dickson, MA[3]; Sergio Mundo[1], PhD; Ran Tao[4], PhD; Xue Zhong[5], PhD; Gul Karakoc[3], MS; Sevim Serley[3], MS; Lan Jiang[3], MS; Nancy J. Cox[3,5], PhD; Wei-Qi Wei[1,2], PhD; C. Michael Stein[3,6], MD; QiPing Feng[1,3,5]*, PhD

[1] Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN
[2] Department of Computer Science, Vanderbilt University, Nashville, TN
[3] Division of Genetic Medicine and Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN
[4] Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee
[5] Vanderbilt Genetics Institute, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN
[6] Department of Pharmacology, Vanderbilt University, Nashville, TN

Role of the Funder/Sponsor: The funders had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

*Corresponding Author: QiPing Feng, PhD, Vanderbilt University Medical Center, 2222 Pierce Ave., Nashville, TN 37232, USA (Tel:615-936-3551, qiping.feng@vumc.org)

[#] The authors contributed equally.

Word count (not including Abstract, Methods, References and figure legends): 4284

Abstract (200 words)

Drugs targeting triglyceride (TG)-associated genes have the potential to improve cardiovascular outcomes. However, we know little regarding the potential additional benefits or deleterious effects of such targeting. As such, in ancestry-specific cohorts of European ancestry (EA) and African ancestry (AA) at Vanderbilt (BioVU), we used PheWAS to test associations between measured TG levels and validated variants previously associated with TG levels in Mendelian randomization studies. We replicated results in All of Us (AoU). In the BioVU EA cohort (n=63,094), 9 of 10 validated SNPs had significant or suggestive associations with lipid and cardiovascular phenotypes, largely consistent with AoU participants of EA (n=97,545). In the BioVU AA cohort (n=12,515) and AoU AA cohort (n=31,710), results were more limited; only 1 of 6 validated SNPs was significantly associated with a lipid or cardiovascular phenotype in either BioVU or AoU, and no significant or suggestive associations were consistent across both cohorts. We detected few secondary effects in either EA or AA BioVU patients, and none were replicated. These results suggest that there may be limited additional benefits, but few deleterious effects, associated with targeting known TG-associated genes. However, these targets may not be as effective for mitigating cardiovascular risk among individuals of AA.

Keywords
Triglycerides; mendelian randomization; phewas; drug side effects

INTRODUCTION

Elevated triglyceride (TG) levels are associated with an increased risk of coronary heart disease (CHD), the leading cause of death globally.[1,2] In the United States, 31% of adults have triglyceride (TG) levels ≥150 mg/dL (i.e., mildly elevated and above),[2] with studies demonstrating that the risks for CHD increase 14% for men and 37% for women for each mmol (88 mg/dL) rise in TG levels.[3] Indeed, targeting TG can have benefits for reducing CHD risk, even among patients already receiving standard lipid-lowering treatments (e.g., statins, fibrates, omega 3 fatty acids, and niacin).[4–6]

As such, identifying effective therapeutic targets associated with TG levels is an important goal for the reduction of CHD risk; however, successful targeting requires not only a decrease in TG levels, but also an associated secondary effect of improved cardiovascular outcomes. The degree to which and the mechanisms whereby TGs are associated with CHD has been the subject of debate, particularly given the potential for confounding and the interrelated effects of other lipid traits (e.g., low density lipoprotein cholesterol [LDL-C] and apolipoprotein B [ApoB]).[2] One relatively recent methodological approach has offered a means to identify whether TGs are in the causal pathway of CHD more accurately—namely, Mendelian randomization (MR). MR seeks to mirror clinical trials (although it differs in some aspects, including susceptibility to Type 1 selection bias), but rather than random assignment, randomization is achieved through the arbitrary assignment of gene variants passed from parents to offspring.[7] Beneficially, this approach offers the ability to assess associations between exposures and outcomes while limiting unmeasured confounding, since genetic information is fixed at conception.[8–13] Indeed, applying such an approach, researchers have used

functional variations in genes with robust effects on TG levels (e.g., *LPL*, *APOA5,* and *APOC3*) to establish that TGs are on the causal pathway of CHD.[12,14–19] These insights have spurred new drug development, targeting specific TG-associated genes for more effective CHD prevention.[20] In particular, current efforts have focused on increasing lipoprotein lipase (LPL) activity; both the *LPL* gene and genes in the LPL pathways (e.g., *ANGPTL3*, *ANGPTL4*, and *APOC3*) have been considered as potential drug targets.[14,21–27]

In addition to establishing these associations, it is also possible to assess other secondary effects of targeting TG-associated genes identified by MR, including adverse effects. During the approval process, clinical trials are designed to test not just the efficacy of new treatments, but also to identify potential side effects (or lack thereof). However, not all phenotypes (effects) can be tested in clinical trials given their relatively limited power and the short periods of follow-up that can obscure long-term effects; these effects typically are only later detected by case reports and retrospective cohort studies of patients that have experienced those effects (e.g., the discovery of cardiovascular risk associated with rofecoxib[28,29]). For example, statins are among the most extensively studied drugs in history and millions of patients take them; however, their increased risk of diabetes was only identified after statins had been on the market for 21 years.[9] Beneficially, combining MR results with PheWAS approaches can facilitate hypothesis generation for potential effects by screening the phenotypic consequences of genetically proxied drug target modulation,[30] offering the advantage of pre-emptively detecting potential long-term effects of such therapies. More specifically, this approach can help with the initial identification of broad on-target

effects, even if the primary driver of the causal pathway is another highly-correlated lipid trait. As such, this framework (i.e., MR results combined with PheWAS) presents researchers with the opportunity to generate hypotheses regarding other potential beneficial or deleterious effects of TG-lowering treatments under development.[31] Given that TG-associated genes (particularly *LPL*) are implicated in numerous pathological conditions (e.g., atherosclerosis, diabetes, pancreatitis, and Alzheimer's disease and related dementias [ADRD]),[32–35] targeting such genes with TG-lowering treatments may have additional effects on these and other conditions. Such approaches are a crucial component of necessary monitoring for potential pleiotropic effects—both beneficial or deleterious) by modelling the real-world deployment of TG-lowering drugs and assessing how known targets may impact patient outcomes within and across populations based in the United States.

Despite this promise, studies of non-CHD effects associated with either TG levels or candidate drugs for TG-lowering have remained limited.[36] In particular, the research to date has focused almost exclusively on patients of European ancestry (EA), predominantly from Finnish and British cohorts;[37–39] this relatively narrow emphasis has potentially hindered our ability to detect effective TG-related targets and predict associated secondary effects, given that lipid profiles can vary significantly by ancestry. Notably, average triglyceride levels are lower in populations of African ancestry (AA) than those of EA,[40] including the lower proportion of individuals with very high TG levels.[40–46] Although genetic studies on inter-individual TG variability have begun to incorporate ancestries beyond those of patients of EA,[46–48] relatively little is known about the relationship between TG-associated genes and secondary effects in

individuals of AA. Accordingly, it is important to determine whether associations validated among patients of EA also apply to patients of AA as well as to determine whether there are ancestry-specific secondary effects (either beneficial or deleterious).

As such, to more broadly assess the potential benefits and risks for TG-lowering medications currently in development, we deployed PheWAS to assess the potential pleiotropic effects of MR-identified genetic predictors of TG levels in ancestry-specific cohorts of patients of EA and AA at Vanderbilt University Medical Center [VUMC], with replication among All of Us Research Program [AoU] participants (**Figure 1**). First, we validated associations between 12 genetic variants reported in previous MR studies from 5 genes identified as TG-lowering targets (i.e., *APOA5*, *LPL*, *APOC3*, *ANGPTL3*, and *ANGPTL4*)[37,49] and measured TG levels. Then, among the validated variants, we conducted PheWAS to assess whether these variants were associated with selected prespecified 1) lipid phenotypes, 2) cardiovascular phenotypes, and 3) other phenotypes previously associated with TGs (including diabetes, pancreatitis, and ADRD), as well as 4) unspecified other phenotypes (i.e., without previously reported associations with TGs). Given the complexity of the causal pathway and the derivation of candidate SNPs (from predominantly EA cohorts), we conducted multiple secondary analyses, including additional adjustments for measured LDL-C, genetic variant set-based tests, predicted gene expression testing, interaction assessments among significant results, and interrogation of instrumental variables (IVs) for TGs derived from Global Lipids Genetics Consortium (GLGC)[50–55] cohorts of individuals with AA.

RESULTS

Cohort characteristics

The EA cohort from BioVU included 63,094 individuals; 35,323 (56.0%) were female, the mean age at the most recent visit was 54.4 ± 21.8, and the mean length of electronic health record (EHR) was 12.0 ± 7.9 years (**Table 1**). The AA cohort from BioVU included 12,515 individuals; 7,737 (61.8%) were female, the mean age was 45.2 ± 20.5, and the mean length of EHR was 11.8 ± 8.7 years (**Table 1**). The adjusted median TG levels for the EA and AA cohorts were 160.9 ± 105.9 mg/dL and 125.5 ± 88.9 mg/dL, respectively. Characteristics for the replication EA (n=97,545) and AA (n=31,710) cohorts in AoU are available in **Supplementary Table 1**.

European ancestry

Eleven variants met the 5 selection criteria for inclusion in our study (i.e., reported as functional, previous use as an MR proxy, previous GWAS significance, availability in BioVU, and BioVU ancestry-specific minor allele frequencies (MAF) consistent with ancestral population-wide MAF available from public databases; **Supplementary Table 2**). We then tested the association of the TG-increasing allele in these SNPs with adjusted median measured TGs to validate their use as instrumental variables. Ten variants were significantly associated with measured TGs in the EA cohort (**Supplementary Table 3**), consistent with the direction of previous reporting; there was no minor allele carriage for the remaining SNP (rs147210663) to test. We restricted subsequent analyses in the BioVU EA cohort and replication AoU cohort to those significantly associated 10 SNPs.

Association between validated variants and clinical phenotypes. For the primary analysis, we first examined potential associations between the 10 validated variants and 21 prespecified clinical phenotypes (**Supplementary Table 4**). We found 17 Bonferroni-corrected significant associations in BioVU patients of EA (**Figure 2; Supplementary Table 5**); there were 21 additional suggestive associations with lipid and cardiovascular phenotypes. Nine of the 10 validated variants were significantly associated with at least one of these phenotypes in the expected direction. Among the potential prespecified secondary effects, we identified suggestive associations between the TG-increasing alleles and increased risk of gout (rs1801177, OR=0.75, p=0.025), diseases of the pancreas (rs1801177, OR=1.37, p=0.024), chronic pancreatitis (rs1801177, OR=1.53, p=0.039); there was lowered risk of urinary tract infection (rs268, OR=0.88, p=0.036) and acute pancreatitis (rs268, OR=0.59, p=0.012; **Figure 2**). For the unspecified phenotypes in the global PheWAS (n=1587), there were additional associations with increased risk of cataplexy and narcolepsy (rs118204057, OR=26.84, $p=1.74x10^{-5}$), atherosclerosis of the renal artery (rs1801177, OR=2.10, $p=1.39x10^{-5}$), pleurisy/pleural effusion (rs1801177, OR=1.30, $p=1.30x10^{-5}$), complication of colostomy or enterostomy (rs1801177, OR=2.20, $p=1.81x10^{-5}$), congenital deformities of feet (rs1801177, OR=2.87, $p=3.10x10^{-5}$), and disorders of lipoid metabolism (rs651821, OR=1.51, $p=1.85x10^{-5}$); there was decreased risk of glaucoma (rs138326449, OR=0.35, $p=1.07x10^{-5}$; **Supplementary Tables 6 and 7**).

Replication. In the AoU EA cohort, the validated variants (n=10) had MAF that were consistent with ancestral population wide MAF (**Supplementary Table 8**). For the 21 prespecified phenotypes, we replicated 26 significant and suggestive associations

($p<0.05$) with these variants (**Figure 2**; **Supplementary Table 9**). Consistent with the BioVU cohort of patients of EA, the AoU cohort had significant associations between prespecified lipid phenotypes and genetic variants (i.e., rs268, rs328, rs651821, rs3135506, rs11207977, rs115849089, rs1801177, rs118204057, and rs116843064). Additionally, we replicated associations with prespecified cardiovascular phenotypes, including angina pectoris (rs328 and rs115849089) and coronary atherosclerosis (rs328 and rs115849089). No other prespecified or unspecified phenotypes identified in the BioVU EA cohort were replicated in the AoU EA cohort (**Supplementary Tables 6 and 9**).

Secondary analyses. While some of the significant associations from the primary analysis were attenuated after the additional adjustment for LDL-C, results remained largely consistent (**Supplementary Table 10**). For the genetic variant set-based tests (i.e., locus-based analyses), we grouped SNPs from *LPL* and *APOA5*, with no notable additional results (**Supplementary Table 11**). For the interactions, we conducted interaction analyses for lipid and cardiovascular phenotypes among validated variants in four TG-associated genes—*APOA5* (rs651821), *LPL* (rs328), *ANGPTL3* (rs11207977), and *ANGPTL4* (rs116843064)—in the BioVU EA cohort using logistic regression models adjusted for the same covariates as the primary analysis. No statistically significant interactions were observed between any of the variant pairs. Likelihood ratio tests further confirmed that including interaction terms did not significantly improve model fit (**Supplementary Table 12**). For predicted gene expression, we found that 4 TG-associated genes had sufficient tissue samples to conduct cross-tissue analysis (*APOA5* only had a single qualifying tissue). Of these, both *LPL* and *ANGPTL3* yielded

significant associations with prespecified lipid phenotypes; *LPL* also had suggestive associations with prespecified cardiovascular phenotypes (**Table 2; Supplementary Table 13**). Among the prespecified other phenotypes, *ANGPTL3* had suggestive associations with gout and migraine, while *APOC3* had a suggestive association with dementias. Single-tissue analyses did not yield any additional significant results (**Supplementary Table 14**).

African ancestry

The same eleven variants met the 5 selection criteria for inclusion in our study (i.e., reported as functional, previous use as an MR proxy, previous GWAS significance, availability in BioVU, and BioVU ancestry-specific MAF consistent with ancestral population-wide MAF available from public databases; **Supplementary Table 2**). However, while all associations were consistent with the direction of previous reporting (**Supplementary Table 3**), only 6 of these variants were significantly associated with measured TGs; 5 variants were not associated, and there was no minor allele carriage for the remaining SNP (**Supplementary Table 3**).

<u>Association between validated variants and clinical phenotypes</u>. In the BioVU AA cohort, there were fewer significant associations among known TG-associated variants. We identified 1 suggestive association between the 6 validated TG-increasing variants and the prespecified lipid or cardiovascular clinical phenotypes: mixed hyperlipidemia (rs11207977, OR=1.14, $p=4.08 \times 10^{-3}$; **Figure 3**; **Supplementary Table 15**). Additionally, among prespecified other phenotypes, we found suggestive associations with lower risk for dizziness (rs11207977, OR=0.89, p=0.023) and increased risk for Alzheimer's

disease (rs651821, OR=1.57, p=0.025). For the 1065 unspecified phenotypes in the global PheWAS, there were significant associations with increased risk for heart valve replacement (rs651821, OR=1.79, $p=1.36x10^{-5}$), hemorrhagic disorder due to intrinsic circulating anticoagulants (rs651821, OR=2.58, $p=3.89x10^{-6}$), and decreased libido (rs1801177, OR=3.68, $p=5.44x10^{-6}$), as well as decreased risk for stricture of artery (rs147210663, OR=0.08, $p=2.00x10^{-5}$) and insect bite (rs147210663, OR=0.09, $p=1.56x10^{-5}$; **Figure 3; Supplementary Tables 16 and 17**). No associations replicated in the AoU AA cohort (**Supplementary Tables 16 and 18**).

Secondary analyses. There were no substantive differences in the primary results after the additional adjustment for LDL-C (**Supplementary Table 19**) or locus-based analyses (**Supplementary Table 20**). There were no results consistent across genes for interaction testing. Leveraging the independent GLGC[50–55] TG GWAS AA cohorts, we identified 12 instrumental variables for TGs (**Supplementary Table 21**), 2 of which were included among the variants of the primary analysis (i.e., rs116843064 in *ANGPTL4* and rs3135506 in *APOA5*). In the BioVU AA cohort, only 7 of these variants were associated with measured TGs (**Supplementary Table 21**); when testing these 7 variants with the 21 prespecified phenotypes, we found largely consistent patterns of limited pleiotropic associations (**Supplementary Table 22**), with only a single significant result among lipid phenotypes (mixed hyperlipidemia and rs35529421, OR=1.16, $p=7.99x10^{-4}$) and a single significant result among cardiovascular phenotypes (angina pectoris and rs10096633; OR=1.38, $p=1.68x10^{-4}$).

DISCUSSION

In this study, we applied PheWAS approaches to MR-identified variants, paralleling previous work to interrogate drug efficacy and adverse effects,[8,56–59] to assess the potential for both deleterious and beneficial effects of targeting TG-associated genes in populations of both EA *and* AA using two extensive EHR databases: BioVU and AoU. We found that for individuals of EA, the lipid and cardiovascular benefits were largely confirmed, while we detected few secondary effects (either previously identified or unknown) that might pose increased risk by targeting these TG-associated genes. Among individuals of AA, however, we found only limited evidence for lipid and cardiovascular benefits; as with individuals of EA, there were few secondary effects associated with these genes. The secondary analyses were largely consistent with the primary findings.

Our results support current efforts to target TG-associated genes[60,61] for the prevention and treatment of heart disease and were consistent with previous findings,[37–39] particularly among individuals of EA. Indeed, all 9 validated variants had suggestive or significant association with lipid phenotypes in the expected direction in the BioVU cohort, and all of these had at least one hyperlipidemia phenotype validated in the AoU cohort. Moreover, 6 of the validated variants had suggestive associations with heart disease-related phenotypes in at least one cohort in the expected direction; 5 variants had suggestive significance in both cohorts (although the phenotypes did not replicate in all instances). In contrast, for individuals of AA, the associations between genetic variants and hyperlipidemia-related phenotypes were more limited, consistent with previous findings.[46,62] Only 6 out of 11 previously

reported variants showed significant associations with adjusted measured TG levels; of these, only one of the variants had a significant association with hyperglyceridemia (phecode 272.12) in either the BioVU or AoU cohort, and two variants showed suggestive significance; none of these results were replicated in both cohorts. None of the 6 validated SNPs were significantly associated with other cardiovascular phenotypes, although one SNP showed suggestive significance in the AoU cohort. Notably, while analysis for the IVs identified in independent GLGC AA cohorts yielded some additional significant and suggestive results in the BioVU AA cohort, these findings were limited. Given that individuals of AA tend to have lower average TG levels and lower frequencies of severe or moderate hypertriglyceridemia compared to individuals of EA,[46] it remains unclear whether the same variants have comparable effect sizes on the cardiovascular outcomes in populations of AA and EA.

Along with the consideration of efficacy for cardiovascular benefits, it was also important to consider the other potential long-term benefits and risks of these drugs; although the limited studies investigating the impact of TG-lowering therapies on non-cardiovascular phenotypes associated with TG candidate genes found no additional effects, these studies focused on European-based populations of EA.[38,39]

Pancreatitis, diabetes, and dementia/Alzheimer's disease were among the most serious conditions previously associated with TGs and TG-related genes. Previous findings have indicated increased risk of pancreatitis with increased risk of high triglycerides;[49,63–68], our analysis of pancreatitis phenotypes among individuals of EA yielded limited results, with two SNPs (rs1801177 and rs3135506) showing suggestive significance for increased risk of related phenotypes in either the BioVU or

AoU EA cohorts (not replicated), and one SNP showing suggestive significance of decreased risk (rs268). Similarly, results regarding potential benefits for diabetes prevention were not conclusive, despite previous research indicating this possibility.[35,37,69–73] Likewise, despite prior findings indicating potential linkages between TG-associated genes and ADRD, we found limited evidence of robust associations (i.e., secondary analysis of predicted gene expression found a suggestive association between dementias and *APOC3* in patients of EA) which supports ideas that TG-lowering treatment could reduce the risk of dementia or Alzheimer's diseases.[74–78] The lack of consistent results parallel research assessing the relationship between LDL-C lowering treatments (i.e., statins) and dementias; indeed, the association between lipid management and the benefit for Alzheimer's disease remains controversial.[79–84] We found no consistent evidence to support these potential secondary beneficial effects among individuals of AA.

Our study featured several notable strengths. First, application of PheWAS to MR-identified variants enabled an unbiased investigation of over 1,000 clinical phenotypes. Second, leveraging longitudinal EHR data from BioVU and AoU allowed us to evaluate long-term effects in individuals of EA and AA. This is particularly significant as populations of AA have historically been underrepresented in genetic and pharmacogenetic research. Moreover, the use of two large cohorts offered more robust results. Finally, by focusing on previously reported variants, our study contributed to a more nuanced understanding of the long-term effects—both beneficial and deleterious—of TG-lowering medications.

However, there were limitations. First, despite conducting analyses in two large biobanks, the cohorts of AA remained comparatively small, which limited the statistical power compared to individuals of EA, especially for low-frequency variants. Indeed, in post-hoc power calculations using PS[85,86] (with a Type 1 error=0.05 and uncorrected chi-squared statistics), given an OR=1.5 for a phenotype in the exposed subjects (as observed in the association between rs1801177 and the prespecified phenotype chronic pancreatitis [phecode 577.2] in individuals of EA), we would need 473 cases and 4730 controls (assuming a 1:10 case:control ratio) to detect an association with a genetic variant of MAF 0.05 (10% carriers), and 4067 cases and 40,670 controls to detect association with a genetic variant of MAF 0.005 (1% carriers). Likewise, given an OR=1.28 for a phenotype in the exposed subjects (as observed in the association between rs268 and the prespecified phenotype other chronic ischemic heart disease [phecode 411.8] in individuals of EA), we would need 1384 cases and 13,840 controls (assuming a 1:10 case:control ratio) to detect association with a genetic variant of MAF 0.05 (10% carriers), and 12,190 cases and 121,900 controls to detect association with a genetic variant of MAF 0.005 (1% carriers). These calculations confirm that several associations identified in EA cohorts would be underpowered for detection in the AA cohorts, particularly for variants with low allele frequency or modest effect size. These observations also apply to other ancestral populations with smaller proportions in BioVU and AoU (e.g., individuals of Asian ancestry). Future validation studies using larger, multi-ancestry cohorts are needed. As AoU continues to recruit participants, this analysis may be more informed with an updated dataset.

Second, this study focused exclusively on previously reported variants within candidate TG-lowering targets which have been identified in cohorts of EA. Given that there are differences in genetic architecture and gene-environment interactions between populations of EA and AA (e.g., effect sizes, minor allele frequences, and LD structure),[87,88] these factors, combined with cohort size, could have hampered replication of results in the AA cohorts (relative to the findings among the EA cohorts). To address this limitation, we tested instrumental variables identified in an independent AA cohort, and the results were comparable to the primary analysis (with extremely limited associations). Notably, most of the identified IVs would not pass the variant selection criteria for the primary analysis; only 7 variants were associated with measured TGs, and of these, only 1 was previously identified as a functional variant (i.e., rs3135506, a variant which overlapped with our primary analysis). Moreover, we did not perform ancestry-specific fine-mapping or trans-ancestry colocalization, which would be required to comprehensively identify causal variants in populations of AA; while we agree this analysis could be informative for identifying additional variants for targeting, this analysis fell outside the scope of our study. Further, there may be other unidentified functional variants whose impact on the association results remain unknown, particularly in populations of AA. Given the historical underrepresentation of populations of AA and other ancestral groups in genetic studies, it is possible that some important ancestry-specific TG-related variants remain unidentified. With the expansion of multi-ancestry biobanks and whole-genome sequencing data as well as the application of additional drug-targeted MR approaches, future studies could more thoroughly assess diverse ancestry groups, including investigation of low-frequency

genetic variants, identification of additional targets, and identification of drug-repurposing candidates.

Third, our study cohorts were composed of patients who have survived to receive clinical care and to be included in EHR-linked biobanks. This introduces the possibility of Type 1 selection bias—particularly if the genetic variants under study influence both disease risk and survival. Variants with beneficial cardiometabolic effects may be overrepresented among older or healthier patients, potentially attenuating associations or obscuring deleterious effects. However, other than Mendelian disease and well-recognized, strongly pathogenic variants, the individual effects of most genetic variations, particularly common genetic variations, on survival is likely to be small and seem unlikely to lead to important depletion of susceptibles in a population spanning a broad age group.[7] Moreover, BioVU offers an opportunity to follow at least some patients from birth, with records spanning more than 30 years. While this follow-up may not completely mitigate issues of survivorship bias, it can help attenuate such concerns.

Fourth, additional factors (e.g., comorbidities and social determinants of health) could be unmeasured confounders; moreover, these factors could help explain why lipid-lowering effects do not correspond automatically to reduced cardiovascular risk. Notably, current efforts are underway (e.g., AoU) to systematically collect data regarding lifestyle, healthcare access and utilization, emotional health, behavioral health, and social factors of health. As more comprehensive data become available, we will have greater opportunities to integrate lifestyle, sociodemographic, and environmental covariates in future analyses.

Fifth, such factors combined with differences between the data sources (e.g., record-keeping practices and patterns of healthcare utilization) could limit our ability to replicate findings. Specifically, while both BioVU and AoU include large, EHR-linked populations, some differences (e.g., age distribution, geographic region, healthcare delivery systems, and duration of EHR follow-up) could influence phenotype prevalence, comorbidity patterns, or data completeness, which could in turn affect statistical power and reproducibility.

Sixth, we were unable to complete analysis of *APOA5* in the predicted expression secondary analysis, given the limited availability of tissues. Moreover, the limited availability of ancestry-diverse expression quantitative trait loci (eQTL) resources impeded our ability to conduct predicted expression analyses for the BioVU and AoU AA cohorts. The ongoing development of such resources[89–92] can facilitate future ancestry-matched transcriptome-wide association studies.

Seventh, detecting protective effects can be more challenging than demonstrating adverse effects related to known risk alleles (e.g., lowered diabetes risk). It is possible that there will be protective benefits from drugs targeting TG-lowering genes, but we do not yet have sufficient power to detect these effects.

Finally, while the combined MR-PheWAS approach was designed to generate hypotheses regarding vertical pleiotropic effects of functional variants, we note that alterations in the selected TG-related genes may lead to additional clinical consequences via causal pathways that are not solely affected by TGs—namely, horizontal pleiotropic effects—because the candidate genes under examination may

have biological functions beyond TG regulation. In particular, as noted in the Introduction, TGs are closely related to other atherogenic lipids, including LDL-C and ApoB, which are well-established causal factors for CHD. Horizontal pleiotropy can complicate clinical interpretation (e.g., biasing effect estimates) and parsing the relationships remains important; indeed, recent multivariable MR studies suggest that associations between TG and CHD were attenuated after adjusting for ApoB-containing lipoproteins, rather than TG per se. Nevertheless, our findings remain relevant for evaluating the clinical consequences of modulating these genes by focusing on the detection of potential effects of the TG-related pathways, especially as several therapeutics targeting these genes have been approved (e.g., evinacumab, an ANGPTL3 inhibitor) or are currently in development. Most importantly, this analysis was not intended to be the conclusive determination of such effects, but rather an important hypothesis-generating component of the larger project of assessing such potential effects; future work should include testing potential alternative pathways as well as monitoring effects in populations using TG-targeting therapies.

In conclusion, by leveraging two large EHR-based biobanks (BioVU and AoU), we applied genetic approaches to evaluate the associations between clinical phenotypes and previously reported TG variants in populations of EA and AA. We confirmed the associations with TG-related phenotypes for most of the candidate genes in individuals of EA; these effects could have implications for preventing cardiovascular disease. Our results also suggest the potential for a few additional beneficial effects, with few deleterious effects, for individuals of EA using TG-lowering treatments; however, these results do not provide comparable evidence to assess the efficacy and safety of these

treatments for individuals of AA. As datasets continue to grow in breadth (e.g., whole genome sequencing and survey data regarding social drivers of health) and length of follow-up, testing in populations other than those of EA is a priority.

METHODS

The primary analysis was conducted in ancestry-specific cohorts of patients of EA and AA from BioVU, a biobank at VUMC. BioVU offers genome-wide genotyping data for ~90,000 patients on the Illumina Infinium Expanded Multi-Ethnic Genotyping Array plus custom content platform (MEGA), which are linked to de-identified copies of the patients' EHRs. These EHRs include demographics, clinical notes, lab results, medications, and diagnostic and procedure codes (e.g., International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM] codes and Tenth Revision [ICD-10-CM] codes).[93] Replication analysis was conducted among comparable cohorts of participants of EA and AA in AoU (**Supplementary §1**). This study was considered non-human subjects and was approved by the VUMC Institutional Review Board (IRB# 220725).

Cohort Eligibility

Primary cohort eligibility criteria included BioVU patients with available MEGA genotyping data that passed standard quality controls (see below; **Supplementary §2**), and at least 2 ICD codes recorded in their EHR through study end date (January 11, 2023). We restricted the study to patients with genetically-determined predominantly EA or AA).[94] Although socially identified EHR-reported race is largely consistent with genetic ancestry in BioVU,[95–97] to minimize the conflation of genetic and social factors, we further restricted the cohorts to patients with consistent EHR-reported race (i.e., White race for patients of EA and Black race for patients of AA); individuals with

more than one reported race were excluded. We further included adjustment for the first 10 principal components (PCs) to help account for intra-ancestral variability.

Clinical characteristics

Patient characteristics were extracted from the EHR, including age (at most recent EHR visit), EHR-reported gender, length of EHR (i.e., the total time between an individual's first and most recent EHR record through the end of the study), lipid levels (i.e., TG, HDL-C, and LDL-C), and whether patients were ever prescribed lipid-lowering medications (statins, niacin, or fibrates; **Supplementary Table 23**); median lipid values were calculated using all measurements after quality control (excluding those that were biologically implausible, e.g. negative values). We adjusted median TG levels for use of lipid-lowering medications for each individual.[46,98]

Genotyping and Imputation

Genotyping was performed on the MEGA platform with quality control as performed previously (**Supplementary §2**).[99–101] Genetically-determined ancestry and the first 10 PCs for ancestry were determined following previous reports.[94]

Genetic variants in TG gene targets

Primary analysis. We assessed variants selected from 5 genes that have been identified previously by MR as TG-lowering targets: *APOA5*, *LPL*, *APOC3*, *ANGPTL3*, and *ANGPTL4* (**Supplementary Table 2**).[37,49] To be included in the analysis, variants

were required to demonstrate the following: (1) relevant function (i.e., protein-coding missense variants or variants known to influence gene expression); (2) prior use as proxies in MR studies to test associations between target genes and clinical phenotypes other than TG levels; (3) availability in BioVU; (4) ancestry-specific previously-reported significant associations ($p < 5\times10^{-}$□) in large-scale GWAS; and (5) consistent ancestry-specific MAF with 1000 Genomes reference data (**Figure 1**).[102] For SNPs meeting these criteria, we then assessed whether the TG-increasing allele of the genotype was significantly associated with adjusted measured TG levels in the direction previously reported in separate analyses of the EA and AA BioVU cohorts; further analyses testing associations with phenotypes (see below) were restricted to these TG-increasing variants significantly associated with measured TG levels ($p<0.05$) in the expected direction.

Secondary analyses. We conducted the following secondary analyses:

1) Additional measured LDL-C adjustment. While elevated TG levels are consistently associated with increased CHD risk, disentangling their independent causal role remains complex due to strong correlations with other lipid traits, particularly ApoB and LDL-C. As such, we conducted analyses in the BioVU EA and AA cohorts with an additional adjustment for median LDL-C levels. As ApoB levels are not routinely collected in clinical practice, we could not additionally adjust for ApoB.

2) Genetic variant set-based tests. We used the Generalized Berk-Jones (GBJ) test to analyze the validated variants in the same locus together for increased power in the BioVU EA and AA cohorts.

3) Predicted gene expression. We conducted a secondary analysis testing the association between the prespecified phenotypes (see below) and the predicted gene expression of the 5 TG-lowering genes; we restricted this analysis to individuals of EA, since most eQTL-based gene expression prediction models—such as those we used from GTEx v8—are largely derived from European ancestry populations[103] and may be less accurate for individuals of other ancestral groups.[104–106] Selecting tissues for availability and biological relevance, we assessed cross-tissue effects relative to the 21 prespecified phenotypes, using (1) kidney and liver tissues for *ANGPTL3*; (2) tibial atrial, kidney, and heart atrial tissues for *ANGPTL4*; (3) pancreatic tissue for *APOA5*; (4) heart left ventricular and kidney tissues for *APOC3*; and (5) kidney and whole blood tissues for *LPL*. We prioritized cross-tissue analyses because most drug effects are systemic and may not be limited to a single biologically "relevant" tissue; further, this approach accounts for potential correlation between tissues and enhances power to detect consistent effects while reducing false positives from single-tissue noise. However, we also conducted single tissue analyses to improve interpretability.

4) Interaction in significant results. We conducted analyses assessing interactions between variants for phenotypes with significant associations in more than one gene within a BioVU cohort.

5) Instrumental variables identified by querying an independent AA cohort. Because the functional variants associated with TGs have been identified among cohorts of European ancestry, we conducted an additional analysis to examine instrumental variables in the 5 TG-related genes derived from a world-wide, multi-institution cohort of individuals with AA. We leveraged the GLGC[50–55] TG GWAS within AA cohorts and

selected AA-specific IVs for TGs. Specifically, we selected variants within +/- 100kb of the candidate gene regions which were associated with measured TG levels at genome-wide significance ($p<5x10^{-8}$) in the GLGC AA cohorts. We also conducted clumping to select independent variants ($R^2<0.01$). Among the identified IVs, we assessed whether the variants were associated with measured TGs in the BioVU AA cohort; for the validated variants, we assessed associations with the 21 preselected phenotypes (see below) in the BioVU AA cohort.

Phenotypes

Phecodes. We established the presence or absence of clinical phenotypes using phecodes, a system based on clinical diagnosis codes (i.e., ICD-9-CM and ICD-10-CM).[56,57] Cases were individuals with 2 or more occurrences of the phecode; controls had zero occurrences; and individuals with 1 mention of the code or related codes were excluded from the analysis to limit misclassification. We required ≥50 cases to include a phecode for analysis.

Phenotype categories. We separated the qualifying phecodes (i.e., ≥50 cases) into 4 categories for analysis in both cohorts, based on previous evidence of association with measured TGs or genetically-determined TG levels (**Supplementary Table 4**): (1) 4 prespecified lipid phenotypes [i.e., hyperlipidemia, hypercholesterolemia, hyperglyceridemia, and mixed hyperlipidemia], to confirm the efficacy of the target; (2) 4 prespecified cardiovascular phenotypes [i.e., essential hypertension, angina pectoris, coronary atherosclerosis, and other chronic ischemic heart disease, unspecified], to confirm the desired secondary cardiovascular benefits; (3) 13 prespecified other

phenotypes previously associated with TGs [i.e., type 2 diabetes, gout, dementias, Alzheimer's disease, vascular dementia, delirium due to conditions classified elsewhere, headache, migraine, dizziness, diseases of pancreas, acute pancreatitis, chronic pancreatitis, and urinary tract infection], to determine if reduced lipid or cardiovascular risk could coincide with elevated risk for known potential secondary outcomes; and (4) all remaining qualifying phecodes in a global PheWAS, to assess whether there are unknown outcomes associated with targeting TG-lowering genes.

Statistical Analysis

We used linear regressions adjusted for age and sex to assess whether qualifying genotypes were significantly associated with adjusted measured TG levels in separate analyses of the EA and AA BioVU cohorts.

For the primary analysis among the EA and AA cohorts in BioVU and AoU, we used logistic regression to test the associations between phecodes and validated TG-increasing variants, with adjustment for age at most recent clinical visit, sex, length of EHR, and the first 10 PCs for ancestry. In BioVU, for associations with the prespecified clinical phenotypes (i.e., categories 1-3; n=21), we considered $p<0.05$ as suggestive, $p<0.0024$ (0.05/21) as significant (Bonferroni-corrected to account for multiple testing), and $p<3.15\times10^{-5}$ and p<4.69×10-5 (Bonferroni-corrected significance for the global PheWAS, as below) as highly significant in the EA and AA cohorts, respectively; we also report significance ($p<0.05$) corrected for false discovery rate (FDR) for greater interpretability of the results. For the remaining phenotypes in BioVU we considered $p<3.15\times10^{-5}$ (0.05/1587 unspecified phecodes tested) and $p<4.69\times10^{-5}$ (0.05/1065

unspecified phecodes tested) as significant in patients of EA and AA, respectively. For the AoU replication cohorts, $p<0.05$ was considered significant.

In the secondary analysis, 1) for the additional LDL-C adjustment, we used logistic regression adjusted for median LDL-C, age at most recent clinical visit, sex, length of EHR, and the first 10 PCs for ancestry. We considered $p<0.0024$ (0.05/21) as significant and $p<0.05$ as suggestive for these tests. 2) For genetic variant set-based tests, we used the GBJ test[107] to analyze the variants in the same locus for increased power in the BioVU EA and AA cohorts. We considered $p<0.0024$ (0.05/21) as significant and $p<0.05$ as suggestive for these tests. 3) For interactions, in selected phenotypes, we conducted a logistic regression for selected variants from genes with significant results, adjusted for age at most recent clinical visit, sex, length of EHR, and the first 10 PCs for ancestry (Model 2). We additionally adjusted for variant-pair interactions (Model 1). We then conducted likelihood ratio tests between these models; we considered $p<0.05$ as significant for these tests. 4) For the predicted gene expression, comparing PrediXcan, UTMOST, and JTI methods,[34,108,109] we identified the best performing gene expression model for each gene by selecting the model with the highest performance R^2 (see SNPs for each tissue in **Supplementary Table 24**). Then, we assessed the cross-tissue effects between (1) the predicted gene expressions for each gene with more than one available tissue and (2) the 21 pre-specified clinical phenotypes (**Supplementary Table 4**), using the GBJ test[107] to account for potential expression correlation among the different tissues.[110,111] We considered $p<0.0024$ (0.05/21) as significant and $p<0.05$ as suggestive for these tests. We also conducted tissue-specific analyses to provide context for interpreting the cross-tissue results; for single-tissue results, we

used logistic regressions to test the associations between the 21 prespecified phenotypes and the predicted expression for each tissue, adjusted for age at most recent clinical visit, sex, length of EHR, and the first 10 PCs for ancestry. And 5) for the instrumental variables identified in an independent GLGC AA cohort, we used we used logistic regression adjusted for age at most recent clinical visit, sex, length of EHR, and the first 10 PCs for ancestry.

Individuals with missingness for any covariate were omitted from primary and secondary analyses. We present categorical variables as numbers and percentages and continuous variables as means and standard deviation (SD). All analyses were conducted using R version 4.1.0 and PLINK version 2.0.

Table 1. Demographic summary of BioVU cohorts.

Figure 1. Study design. TG = triglyceride. MR = Mendelian randomization. GWAS = genome-wide association study. MAF = minor allele frequency. EA = European ancestry. AA = African ancestry. PheWAS = phenome-wide association study. BioVU = Vanderbilt University Medical Center biobank. AoU = All of Us Research Program.

Figure 2. Significant associations between prespecified phenotypes and genetic variants in the BioVU European ancestry cohort, with All of Us replication indicated. ○ = replicated in All of Us. Dark red = $p<3.15\times10^{-5}$ (0.05/1587) [logistic regression]; highly significantly increased risk associated with TG-increasing allele (full PheWAS Bonferroni-corrected). Medium red = $p<0.0024$ (0.05/21) [logistic regression]; significantly increased risk associated with TG-increasing allele (Bonferroni-corrected). Light red = $p<0.05$ [logistic regression]; suggestive significance of increased risk associated with TG-increasing allele. Light blue = $p<0.05$ [logistic regression]; suggestive significance of decreased risk associated with TG-increasing allele.

Table 2. Significant associations between prespecified phenotypes and genetically predicted expression of TG-associated genes among BioVU patients of EA.

Figure 3. Significant associations between prespecified phenotypes and genetic variants in the BioVU African ancestry cohort, with All of Us replication indicated. ○ = replicated in All of Us. Dark red = $p<4.69\times10^{-5}$ (0.05/1065) [logistic regression]; highly significantly increased risk associated with TG-increasing allele (full PheWAS Bonferroni-corrected). Medium red = $p<0.0024$ (0.05/21) [logistic regression]; significantly increased risk associated with TG-increasing allele (Bonferroni-corrected). Light red = $p<0.05$ [logistic regression]; suggestive significance of increased risk associated with TG-increasing allele. Light blue = $p<0.05$ [logistic regression; suggestive significance of decreased risk associated with TG-increasing allele.

List of abbreviations

AA = African ancestry
ADRD = Alzheimer's disease and related dementias
AoU = All of Us Research Program
ApoB = apolipoprotein B
CHD = coronary heart disease
EA = European ancestry
EHR = electronic health record
eQTL = expression quantitative trait loci
FDR = false discovery rate
GBJ = generalized Berk-Jones
GLGC = Global Lipids Genetics Consortium
HDL-C = high-density lipoprotein cholesterol
ICD-9-CM = International Classification of Diseases, Ninth Revision, Clinical Modification
ICD-10-CM = International Classification of Diseases, Tenth Revision, Clinical Modification
IRB = Institutional Review Board
LDL-C = low density lipoprotein cholesterol
MAF = minor allele frequencies
MEGA = Illumina Infinium Expanded Multi-Ethnic Genotyping Array
MR = Mendelian randomization
PC = principal component
PheWAS = phenome wide association study
SD = standard deviation
TG = triglyceride
VUMC = Vanderbilt University Medical Center

DECLARATIONS

Ethics approval and consent to participate

The study was approved by the VUMC Institutional Review Board (IRB# 220725).

Availability of data and materials

BioVU data and code

Statistical code: Available at https://github.com/FengLabVUMC/Triglyceride-PheWAS.
Data set: Full summary statistics for validated variants and qualifying phenotypes for the BioVU cohorts are available in file Supplementary Tables 7-24. For individual level data, VUMC EHR data are de-identified using Safe-harbor methods. BioVU genomic data are linked to de-identified records and are further protected by BioVU data use agreements ensuring that researchers will not attempt re-identification. BioVU participant consenting procedures limit access to individual level data. Limited primary cohort data are available by request to Dr. Feng (e-mail, Qiping.feng@vumc.org), pending BioVU approval and a data use agreement.

All of Us data and code

In line with the privacy standards set by the All of Us Research Program, data and code used for these cohorts are available to approved researchers who register for access to the Researcher Workbench platform at https://workbench.researchallofus.org/login. This analysis was run on the All of Us Research Program Controlled Tier Dataset version 7, production release C2022Q4R9.

Declaration of Interests

Drs. Mundo and Zhong reported receiving support from National Institutes of Health (NIH) grants during the conduct of the study. Ms. Karakoc reported receiving support from the Vanderbilt Physician Scientist Development Award during the conduct of the study. Dr. Cox reported receiving an honorarium from the Endocrine Society during the conduct of the study. Dr. Stein reported receiving support from NIH grants, book royalties, and an honorarium for writing a textbook chapter during the conduct of the study. The remaining authors declared no competing interests for this work.

Funding acknowledgements: R01HL163854 (Q.F.), R01GM120523 (Q.F.), R01HL133786 (W.Q.W.), R01HL171809 (Q.F, W.Q.W) and Vanderbilt Faculty Research Scholar Fund (Q.F.).

The primary dataset(s) were obtained from Vanderbilt University Medical Center's BioVU, which is supported by institutional funding, 1S10RR025141-01, and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Additional funding provided by the NIH through grants P50GM115305 and U19HL065962. The authors wish to acknowledge the expert technical support of the VANTAGE and VANGARD core facilities, supported in part by the Vanderbilt-Ingram Cancer Center (P30 CA068485) and Vanderbilt Vision Center (P30 EY08126).

The validation dataset(s) were obtained from the All of Us Research Program. This program would not be possible without the partnership of its participants. Additionally, the All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276.

Role of Funders
The funders had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Contributors
Conceptualization (ALD, WQW, CMS, QF); data curation (EO, YX, ALD, SM, RT, GK, SS, LJ, NJC, WQW, QF); formal analysis (EO, YX); investigation (EO, YX, ALD, SM, RT, XZ, CMS, QF); funding acquisition (WQW, QF); methodology (EO, YX); validation (YX, ALD, SM, WQW, CMS, QF); project administration (ALD, WQW, CMS, QF) , visualization (EO, YX); supervision (ALD, WQW, CMS, QF); resources (SS, LJ, NJC, WQW, QF); writing – original draft (EO, YX, ALD, CMS, QF); and writing - review and editing (EO, YX, ALD, SM, RT, XZ, GK, SS, LJ, NJC, CMS, QF)

The first and corresponding authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors read and approved the final version of the manuscript.

References

1. The top 10 causes of death. https://test-cms.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death (2025).
2. Miller, M. *et al.* Triglycerides and Cardiovascular Disease A Scientific Statement From the American Heart Association. *Circulation* **123**, 2292–2333 (2011).
3. Hokanson, J. E. & Austin, M. A. Plasma triglyceride level is a risk factor for cardiovascular disease independent of high-density lipoprotein cholesterol level: a meta-analysis of population-based prospective studies. *J Cardiovasc Risk* **3**, 213–219 (1996).
4. Nichols, G. A., Philip, S., Reynolds, K., Granowitz, C. B. & Fazio, S. Increased Cardiovascular Risk in Hypertriglyceridemic Patients With Statin-Controlled LDL Cholesterol. *The Journal of Clinical Endocrinology & Metabolism* **103**, 3019–3027 (2018).
5. Mszar, R., Bart, S., Sakers, A., Soffer, D. & Karalis, D. G. Current and Emerging Therapies for Atherosclerotic Cardiovascular Disease Risk Reduction in Hypertriglyceridemia. *Journal of Clinical Medicine* **12**, 1382 (2023).
6. Hussain, A., Al Rifai, M., Hermel, M., Slipczuk, L. & Virani, S. S. What is really new in triglyceride guidelines? *Current Opinion in Endocrinology, Diabetes and Obesity* **30**, 73 (2023).
7. Gkatzionis, A. & Burgess, S. Contextualizing selection bias in Mendelian randomization: how bad is it likely to be? *Int J Epidemiol* **48**, 691–701 (2019).
8. Liu, G. *et al.* A Mendelian Randomization Approach Using 3-HMG-Coenzyme-A Reductase Gene Variation to Evaluate the Association of Statin-Induced Low-Density Lipoprotein Cholesterol Lowering With Noncardiovascular Disease Phenotypes. *JAMA Netw Open* **4**, e2112820 (2021).
9. Smit, R. A. J. *et al.* Statin-induced LDL cholesterol response and type 2 diabetes: a bidirectional two-sample Mendelian randomization study. *Pharmacogenomics J* **20**, 462–470 (2020).
10. Gill, D. *et al.* Use of Genetic Variants Related to Antihypertensive Drugs to Inform on Efficacy and Side Effects. *Circulation* **140**, 270–279 (2019).

11. Satterfield, B. A. *et al.* Associations of Genetically Predicted Lp(a) (Lipoprotein [a]) Levels With Cardiovascular Traits in Individuals of European and African Ancestry. *Circ Genom Precis Med* **14**, e003354 (2021).

12. Ference, B. A. *et al.* Association of Triglyceride-Lowering LPL Variants and LDL-C–Lowering LDLR Variants With Risk of Coronary Heart Disease. *JAMA* **321**, 364–373 (2019).

13. Ference, B. A., Majeed, F., Penumetcha, R., Flack, J. M. & Brook, R. D. Effect of Naturally Random Allocation to Lower Low-Density Lipoprotein Cholesterol on the Risk of Coronary Heart Disease Mediated by Polymorphisms in NPC1L1, HMGCR, or Both: A 2 × 2 Factorial Mendelian Randomization Study. *Journal of the American College of Cardiology* **65**, 1552–1561 (2015).

14. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute *et al.* Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med* **371**, 22–31 (2014).

15. Do, R. *et al.* Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102–106 (2015).

16. Douvris, A. *et al.* Functional Analysis of the TRIB*1* Associated Locus Linked to Plasma Triglycerides and Coronary Artery Disease. *Journal of the American Heart Association* **3**, e000884 (2014).

17. Do, R. *et al.* Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* **45**, 1345–1352 (2013).

18. Triglyceride Coronary Disease Genetics Consortium and Emerging Risk Factors Collaboration *et al.* Triglyceride-mediated pathways and coronary disease: collaborative analysis of 101 studies. *Lancet* **375**, 1634–1639 (2010).

19. Holmes, M. V. *et al.* Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* **36**, 539–550 (2015).

20. Lotta, L. A. *et al.* Association of Genetically Enhanced Lipoprotein Lipase-Mediated Lipolysis and Low-Density Lipoprotein Cholesterol-Lowering Alleles With Risk of Coronary Disease and Type 2 Diabetes. *JAMA Cardiol* **3**, 957–966 (2018).

21. Musunuru, K. *et al.* Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N. Engl. J. Med.* **363**, 2220–2227 (2010).

22. Dewey, F. E. *et al.* Genetic and Pharmacologic Inactivation of ANGPTL3 and Cardiovascular Disease. *N Engl J Med* **377**, 211–221 (2017).

23. Jørgensen, A. B., Frikke-Schmidt, R., Nordestgaard, B. G. & Tybjærg-Hansen, A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med* **371**, 32–41 (2014).

24. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators *et al.* Coding Variation in ANGPTL4, LPL, and SVEP1 and the Risk of Coronary Disease. *N Engl J Med* **374**, 1134–1144 (2016).

25. Dewey, F. E. *et al.* Inactivating Variants in ANGPTL4 and Risk of Coronary Artery Disease. *N. Engl. J. Med.* **374**, 1123–1133 (2016).

26. Nordestgaard, B. G. *et al.* Heterozygous Lipoprotein Lipase Deficiency. *Circulation* **96**, 1737–1744 (1997).

27. Wittrup, H. H., Tybjærg-Hansen, A. & Nordestgaard, B. G. Lipoprotein Lipase Mutations, Plasma Lipids and Lipoproteins, and Risk of Ischemic Heart Disease. *Circulation* **99**, 2901–2907 (1999).

28. Ray, W. A. *et al.* COX-2 selective non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease. *Lancet* **360**, 1071–1073 (2002).

29. Ray, W. A., Stein, C. M., Hall, K., Daugherty, J. R. & Griffin, M. R. Non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease: an observational cohort study. *Lancet* **359**, 118–123 (2002).

30. Safarova, M. S. *et al.* A phenome-wide association study to discover pleiotropic effects of PCSK9, APOB, and LDLR. *NPJ Genom Med* **4**, 3 (2019).

31. Li, Z. *et al.* Genetic association of lipids and lipid-lowering drug target genes with non-alcoholic fatty liver disease. *EBioMedicine* **90**, 104543 (2023).

32. Mead, J. R. & Ramji, D. P. The pivotal role of lipoprotein lipase in atherosclerosis. *Cardiovascular Research* **55**, 261–269 (2002).

33. Mead, J. R., Irvine, S. A. & Ramji, D. P. Lipoprotein lipase: structure, function, regulation, and role in disease. *J Mol Med* **80**, 753–769 (2002).

34. Hu, Y. *et al.* A statistical framework for cross-tissue transcriptome-wide association analysis. *Nat Genet* **51**, 568–576 (2019).

35. Digenio, A. *et al.* Antisense-Mediated Lowering of Plasma Apolipoprotein C-III by Volanesorsen Improves Dyslipidemia and Insulin Sensitivity in Type 2 Diabetes. *Diabetes Care* **39**, 1408–1415 (2016).

36. Park, J. K. *et al.* Phenome-wide Mendelian randomization study of plasma triglyceride levels and 2600 disease traits. *eLife* **12**, e80560 (2023).

37. Wang, Q. *et al.* Metabolic profiling of angiopoietin-like protein 3 and 4 inhibition: a drug-target Mendelian randomization analysis. *Eur Heart J* **42**, 1160–1169 (2021).

38. Gagnon, E., Bourgault, J., Gobeil, É., Thériault, S. & Arsenault, B. J. Impact of loss-of-function in angiopoietin-like 4 on the human phenome. *Atherosclerosis* **393**, (2024).

39. Gagnon, E. *et al.* Evaluating the Cardiometabolic Efficacy and Safety of Lipoprotein Lipase Pathway Targets in Combination With Approved Lipid-Lowering Targets: A Drug Target Mendelian Randomization Study. *Circ Genom Precis Med* **18**, e004933 (2025).

40. Miller, M. *et al.* Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. *Circulation* **123**, 2292–2333 (2011).

41. Frank, A. T. H. *et al.* Racial/Ethnic Differences in Dyslipidemia Patterns. *Circulation* **129**, 570–579 (2014).

42. Meeks, K. A. C. *et al.* Ancestral and environmental patterns in the association between triglycerides and other cardiometabolic risk factors. *EBioMedicine* **91**, 104548 (2023).

43. Sumner, A. E. Ethnic Differences in Triglyceride Levels and High-Density Lipoprotein Lead to Underdiagnosis of the Metabolic Syndrome in Black Children and Adults. *J Pediatr* **155**, S7.e7-S7.11 (2009).

44. Lin, S. X., Carnethon, M., Szklo, M. & Bertoni, A. Racial/Ethnic Differences in the Association of Triglycerides with Other Metabolic Syndrome Components: The Multi-Ethnic Study of Atherosclerosis. *Metabolic Syndrome and Related Disorders* **9**, 35 (2011).

45. Li, C., Ford, E. S., Meng, Y.-X., Mokdad, A. H. & Reaven, G. M. Does the association of the triglyceride to high-density lipoprotein cholesterol ratio with fasting serum insulin differ by race/ethnicity? *Cardiovascular Diabetology* **7**, 4 (2008).

46. Jiang, L. *et al.* Characterizing genetic profiles for high triglyceride levels in U.S. patients of African ancestry. *J Lipid Res* **65**, 100569 (2024).

47. Gill, P. K. *et al.* Ancestry-specific profiles of genetic determinants of severe hypertriglyceridemia. *J Clin Lipidol* **15**, 88–96 (2021).

48. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).

49. Hansen, S. E. J., Madsen, C. M., Varbo, A., Tybjærg-Hansen, A. & Nordestgaard, B. G. Genetic Variants Associated With Increased Plasma Levels of Triglycerides, via Effects on the Lipoprotein Lipase Pathway, Increase Risk of Acute Pancreatitis. *Clin Gastroenterol Hepatol* **19**, 1652-1660.e6 (2021).

50. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679 (2021).

51. Graham, S. E. *et al.* Author Correction: The power of genetic diversity in genome-wide association studies of lipids. *Nature* **618**, E19–E20 (2023).

52. Ramdas, S. *et al.* A multi-layer functional genomic analysis to understand noncoding genetic variation in lipids. *Am J Hum Genet* **109**, 1366–1387 (2022).

53. Dron, J. S., Natarajan, P. & Peloso, G. M. The breadth and impact of the Global Lipids Genetics Consortium. *Curr Opin Lipidol* **36**, 61–70 (2025).

54. Kanoni, S. *et al.* Implicating genes, pleiotropy, and sexual dimorphism at blood lipid loci through multi-ancestry meta-analysis. *Genome Biol* **23**, 268 (2022).

55. GLGC. https://www.lipidgenetics.org/#top (2025).

56. Wei, W.-Q. *et al.* Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS ONE* **12**, e0175508 (2017).

57. Wu, P. *et al.* Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Medical Informatics* **7**, e14325 (2019).

58. Safarova, M. S. *et al.* A phenome-wide association study to discover pleiotropic effects of PCSK9, APOB, and LDLR. *npj Genomic Med* **4**, 3 (2019).

59. Denny, J. C. *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–10 (2010).

60. Feingold, K. R. Triglyceride Lowering Drugs. in *Endotext* (eds Feingold, K. R. et al.) (MDText.com, Inc., South Dartmouth (MA), 2000).

61. Center for Drug Evaluation and Research. FDA approves add-on therapy for patients with genetic form of severely high cholesterol. *FDA* https://www.fda.gov/drugs/news-events-human-drugs/fda-approves-add-therapy-patients-genetic-form-severely-high-cholesterol-0 (2024).

62. Patel, R. A. *et al.* Genetic interactions drive heterogeneity in causal variant effect sizes for gene expression and complex traits. *Am J Hum Genet* **109**, 1286–1297 (2022).

63. Wang, B. *et al.* Lipid levels and risk of acute pancreatitis using bidirectional Mendelian randomization. *Sci Rep* **14**, 6267 (2024).

64. Mi, J. *et al.* Mendelian randomization in blood metabolites identifies triglycerides and fatty acids saturation level as associated traits linked to pancreatitis risk. *Front Nutr* **9**, 1021942 (2022).

65. Yuan, S., Giovannucci, E. L. & Larsson, S. C. Gallstone disease, diabetes, calcium, triglycerides, smoking and alcohol consumption and pancreatitis risk: Mendelian randomization study. *NPJ Genom Med* **6**, 27 (2021).

66. Adiamah, A., Psaltis, E., Crook, M. & Lobo, D. N. A systematic review of the epidemiology, pathophysiology and current management of hyperlipidaemic pancreatitis. *Clin Nutr* **37**, 1810–1822 (2018).

67. Brunzell, J. D. & Deeb, S. S. Familial Lipoprotein Lipase Deficiency, Apo C-II Deficiency, and Hepatic Lipase Deficiency. in *The Online Metabolic and Molecular Bases of Inherited Disease* (eds Valle, D. L., Antonarakis, S., Ballabio, A., Beaudet, A. L. & Mitchell, G. A.) (McGraw-Hill Education, New York, NY, 2019).

68. Burnett, J. R., Hooper, A. J. & Hegele, R. A. Familial Lipoprotein Lipase Deficiency. in *GeneReviews®* (eds Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).

69. Robciuc, M. R. *et al.* Angptl3 Deficiency Is Associated With Increased Insulin Sensitivity, Lipoprotein Lipase Activity, and Decreased Serum Free Fatty Acids. *Arteriosclerosis, Thrombosis, and Vascular Biology* **33**, 1706–1713 (2013).

70. Graham, M. J. *et al.* Cardiovascular and Metabolic Effects of ANGPTL3 Antisense Oligonucleotides. *New England Journal of Medicine* **377**, 222–232 (2017).

71. Goldberg, R. B. *et al.* Effects of Extended-Release Niacin Added to Simvastatin/Ezetimibe on Glucose and Insulin Values in AIM-HIGH. *Am J Med* **129**, 753.e13–22 (2016).

72. Robciuc, M. R. *et al.* Angptl3 Deficiency Is Associated With Increased Insulin Sensitivity, Lipoprotein Lipase Activity, and Decreased Serum Free Fatty Acids. *Arteriosclerosis, Thrombosis, and Vascular Biology* **33**, 1706–1713 (2013).

73. Digenio, A. *et al.* Antisense-Mediated Lowering of Plasma Apolipoprotein C-III by Volanesorsen Improves Dyslipidemia and Insulin Sensitivity in Type 2 Diabetes. *Diabetes Care* **39**, 1408–1415 (2016).

74. Dimache, A. M., Șalaru, D. L., Sascău, R. & Stătescu, C. The Role of High Triglycerides Level in Predicting Cognitive Impairment: A Review of Current Evidence. *Nutrients* **13**, 2118 (2021).

75. Koch, M. *et al.* High density lipoprotein and its apolipoprotein-defined subspecies and risk of dementia. *J Lipid Res* **61**, 445–454 (2020).

76. Bruce, K. D., Tang, M., Reigan, P. & Eckel, R. H. Genetic Variants of Lipoprotein Lipase and Regulatory Factors Associated with Alzheimer's Disease Risk. *International Journal of Molecular Sciences* **21**, 8338 (2020).

77. Gong, H. *et al.* Lipoprotein lipase (LPL) is associated with neurite pathology and its levels are markedly reduced in the dentate gyrus of Alzheimer's disease brains. *J Histochem Cytochem* **61**, 857–868 (2013).

78. Xian, X. *et al.* Presynaptic defects underlying impaired learning and memory function in lipoprotein lipase-deficient mice. *J Neurosci* **29**, 4681–4685 (2009).

79. Jick, H., Zornberg, G. L., Jick, S. S., Seshadri, S. & Drachman, D. A. Statins and the risk of dementia. *Lancet* **356**, 1627–1631 (2000).

80. Wolozin, B., Kellman, W., Ruosseau, P., Celesia, G. G. & Siegel, G. Decreased prevalence of Alzheimer disease associated with 3-hydroxy-3-methyglutaryl coenzyme A reductase inhibitors. *Arch Neurol* **57**, 1439–1443 (2000).

81. Feldman, H. H. *et al.* Randomized controlled trial of atorvastatin in mild to moderate Alzheimer disease. *Neurology* **74**, 956–964 (2010).

82. Sano, M. *et al.* A randomized, double-blind, placebo-controlled trial of simvastatin to treat Alzheimer disease. *Neurology* **77**, 556–563 (2011).

83. Wei, B.-Z. *et al.* The Relationship of Omega-3 Fatty Acids with Dementia and Cognitive Decline: Evidence from Prospective Cohort Studies of Supplementation, Dietary Intake, and Blood Markers. *Am J Clin Nutr* **117**, 1096–1109 (2023).

84. Quinn, J. F. *et al.* Docosahexaenoic acid supplementation and cognitive decline in Alzheimer disease: a randomized trial. *JAMA* **304**, 1903–1911 (2010).

85. Dupont, W. D. & Plummer, W. D. Power and sample size calculations for studies involving linear regression. *Control Clin Trials* **19**, 589–601 (1998).

86. Power and Sample Size. https://cqsclinical.app.vumc.org/ps/ (2025).

87. Deo, R. C. *et al.* Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: the Jackson Heart Study. *PLoS Genet* **5**, e1000342 (2009).

88. Kuchenbaecker, K. *et al.* The transferability of lipid loci across African, Asian and European cohorts. *Nat Commun* **10**, 4330 (2019).

89. Kelly, D. E. *et al.* The genetic and evolutionary basis of gene expression variation in East Africans. *Genome Biol* **24**, 35 (2023).

90. Sengupta, D., Choudhury, A. & Ramsay, M. H3Africa: a model for implementing biobank-based genomic research in resource-constrained settings. *Human Molecular Genetics* ddaf113 (2025) doi:10.1093/hmg/ddaf113.

91. Mulder, N. *et al.* H3Africa: current perspectives. *Pharmgenomics Pers Med* **11**, 59–66 (2018).

92. Shang, L. *et al.* Genetic Architecture of Gene Expression in European and African Americans: An eQTL Mapping Study in GENOA. *The American Journal of Human Genetics* **106**, 496–512 (2020).

93. Wei, W.-Q. & Denny, J. C. Extracting research-quality phenotypes from electronic health records to support precision medicine. *Genome Med* **7**, 41 (2015).

94. Zhang, D., Dey, R. & Lee, S. Fast and robust ancestry prediction using principal component analysis. *Bioinformatics* **36**, 3439–3446 (2020).

95. Dumitrescu, L. *et al.* Assessing the accuracy of observer-reported ancestry in a biorepository linked to electronic medical records. *Genet Med* **12**, 648–650 (2010).

96. Hall, J. B., Dumitrescu, L., Dilks, H. H., Crawford, D. C. & Bush, W. S. Accuracy of Administratively-Assigned Ancestry for Diverse Populations in an Electronic Medical Record-Linked Biobank. *PLoS One* **9**, e99161 (2014).

97. Dickson, A. L. *et al.* Race, Genotype, and Azathioprine Discontinuation: A Cohort Study. *Ann Intern Med* **175**, 1092–1099 (2022).

98. Hu, Y. *et al.* Minority-centric meta-analyses of blood lipid levels identify novel loci in the Population Architecture using Genomics and Epidemiology (PAGE) study. *PLoS Genet* **16**, e1008684 (2020).

99. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279–1283 (2016).

100. Jiang, L. *et al.* Association between APOL1 risk variants and the occurrence of sepsis in patients hospitalized with infections. *eLife* **12**, (2023).

101. Liu, G. *et al.* The relationship between high density lipoprotein cholesterol and sepsis: A clinical and genetic approach. *Clin Transl Sci* **16**, 489–501 (2023).

102. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

103. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

104. Yang, G., Mishra, M. & Perera, M. A. Multi-Omics Studies in Historically Excluded Populations: The Road to Equity. *Clin Pharmacol Ther* **113**, 541–556 (2023).

105. Mikhaylova, A. V. & Thornton, T. A. Accuracy of Gene Expression Prediction From Genotype Data With PrediXcan Varies Across and Within Continental Populations. *Front Genet* **10**, 261 (2019).

106. Saitou, M., Dahl, A., Wang, Q. & Liu, X. Allele frequency impacts the cross-ancestry portability of gene expression prediction in lymphoblastoid cell lines. *Am J Hum Genet* **111**, 2814–2825 (2024).

107. Sun, R. & Lin, X. Genetic Variant Set-Based Tests Using the Generalized Berk-Jones Statistic with Application to a Genome-Wide Association Study of Breast Cancer. *J Am Stat Assoc* **115**, 1079–1091 (2020).

108. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091–1098 (2015).

109. Zhou, D. *et al.* A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nat Genet* **52**, 1239–1246 (2020).

110. Uellendahl-Werth, F. *et al.* Cross-tissue transcriptome-wide association studies identify susceptibility genes shared between schizophrenia and inflammatory bowel disease. *Commun Biol* **5**, 1–15 (2022).

111. Gaynor, S. M., Sun, R., Lin, X. & Quackenbush, J. Identification of differentially expressed gene sets using the Generalized Berk-Jones statistic. *Bioinformatics* **35**, 4568–4576 (2019).

	EA cohort	AA cohort
Total N	63,094	12,515
Female, n (%)	35,323 (56.0%)	7,737 (61.8%)
Age (mean ± SD, years)	54.4 ± 21.8	45.2 ± 20.5
Female	53.6 ± 21.2	44.7 ± 20.1
Male	55.5 ± 22.5	46.1 ± 21.3
Length of EHR (mean ± SD, years)	12.0 ± 7.9	11.8 ± 8.7
Female	12.7 ± 8.0	12.3 ± 8.8
Male	11.2 ± 7.8	11.1 ± 8.6
Lipid measures (mean ± SD, mg/dL)		
Triglycerides*	160.9 ± 105.9 (n=29,327)	125.5 ± 88.9 (n=5,587)
Female	150.0±99.3 (n=16,459)	117.6 ± 102.9 (n=3,312)
Male	174.8±112.2 (n=12,868)	137.1 ± 76.8 (n=2,275)
LDL-C	102.3 ± 31.5 (n=27,712)	103.8 ± 33.0 (n=5,405)
Female	106.5 ± 31.1 (n=15,602)	105.4 ± 32.5 (n=3,221)
Male	96.9 ± 31.2 (n=12,110)	101.4 ± 33.4 (n=2,184)
HDL-C	51.6 ± 16.7 (n=28,513)	52.2 ± 15.8 (n=5,521)
Female	57.5 ± 17.0 (n=16,059)	55.4 ± 15.9 (n=3,289)
Male	44.1 ± 12.9 (n=12,454)	47.4 ± 14.4 (n=2,232)
Lipid-lowering medications, n (%)		
Statins	21,646 (34.3%)	3,070 (24.5%)
Female	10,289 (29.1%)	1699 (22.0%)
Male	11,357 (40.9%)	1371 (28.7%)
Niacin	2,185 (3.5%)	134 (1.1%)
Female	856 (2.4%)	66 (0.9%)
Male	1,329 (4.8%)	68 (1.4%)
Fibrates	3,058 (4.8%)	166 (1.3%)
Female	1,298 (3.7%)	75 (1.0%)
Male	1,760 (6.3%)	91 (1.9%)

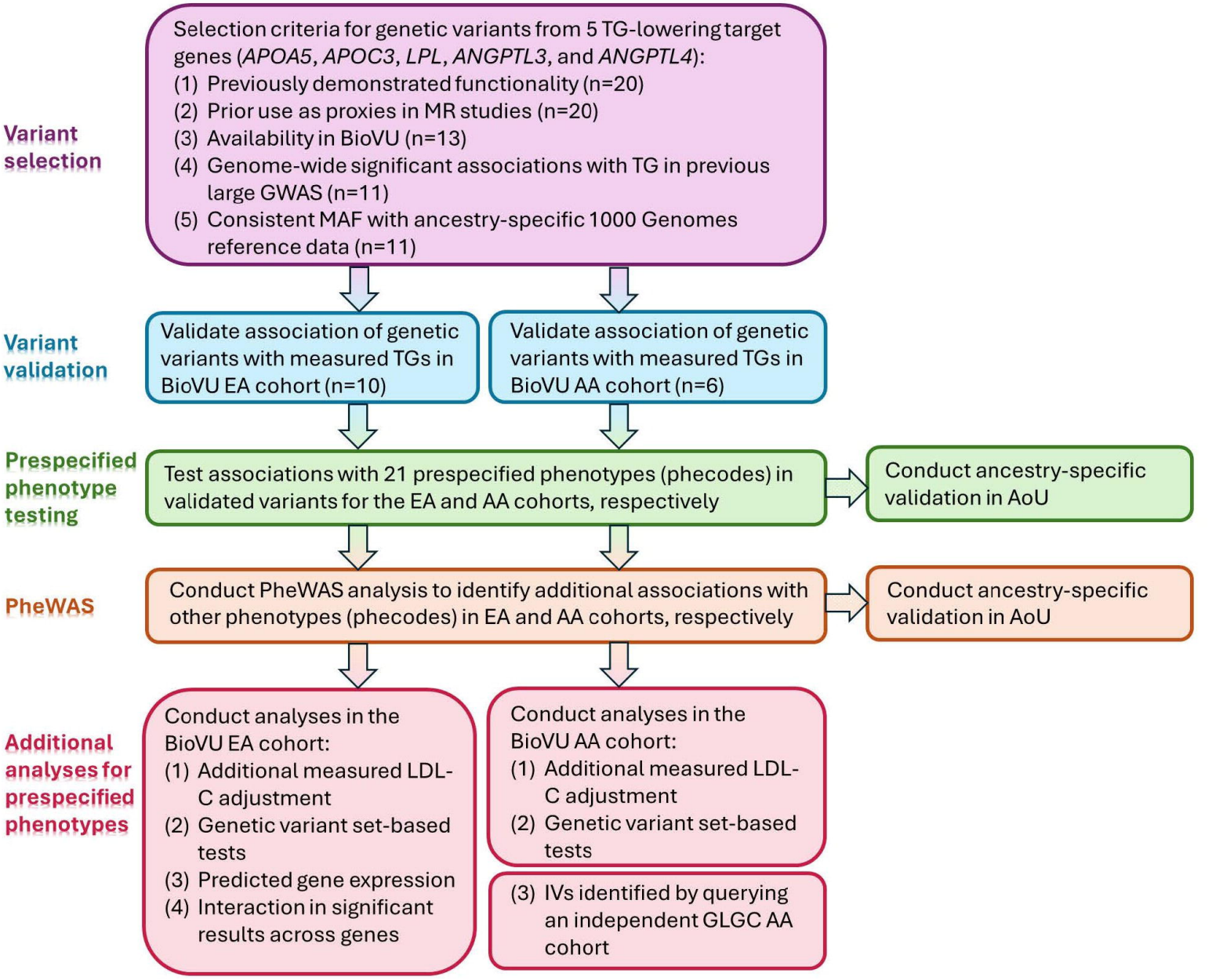
Table 1. Demographic summary of BioVU cohorts

*Adjusted for statin, niacin, and fibrate use (see Supplementary Table 23).
EA = European ancestry; AA = African ancestry; EHR = electronic health record

Table 2. Significant associations between prespecified phenotypes and genetically predicted expression of TG-associated genes among BioVU patients of EA

			LPL		*APOC3*		*ANGPTL3*	
Phenotype group	**Phecode**	**Phenotype**	**GBJ_Statistic**	**p-value**	**GBJ_Statistic**	**p-value**	**GBJ_Statistic**	**p-value**
Lipid phenotypes	272.1	Hyperlipidemia	**5.9657**	**0.001****	1.7557	0.087	**5.5513**	**0.002****
	272.11	Hypercholesterolemia	**2.4444**	**0.045***	0.7635	0.248	**4.7191**	**0.004***
	272.12	Hyperglyceridemia	**5.9906**	**0.001****	0.5192	0.328	**2.1245**	**0.049***
	272.13	Mixed hyperlipidemia	**4.5951**	**0.005***	1.5638	0.106	**5.8732**	**0.001****
Cardiovascular phenotypes	411.3	Angina pectoris	**4.8790**	**0.004***	0	1	0	1
	411.4	Coronary atherosclerosis	**3.4456**	**0.016***	0.1981	0.486	0.0578	0.490
	411.8	Other chronic ischemic heart disease, unspecified	**4.0202**	**0.009***	0	1	0	1
Previously reported other phenotypes	274.1	Gout	0.1353	0.557	0	1	**4.1693**	**0.006***
	290.1	Dementias	0	1	**3.4496**	**0.016***	0	1
	340	Migraine	0	1	0	1	**2.4559**	**0.035***

**p<0.0024 (0.05/21); significant association with genetically predicted expression of gene (Bonferroni-corrected)
*p<0.05; suggestive significance of association with genetically predicted expression of gene
bold=at least suggestive significance; red=Bonferroni-corrected significant
EA = European ancestry; GBJ = Generalized Berk Jones


Variant selection
Selection criteria for genetic variants from 5 TG-lowering target genes (APOA5, APOC3, LPL, ANGPTL3, and ANGPTL4):
(1) Previously demonstrated functionality (n=20)
(2) Prior use as proxies in MR studies (n=20)
(3) Availability in BioVU (n=13)
(4) Genome-wide significant associations with TG in previous large GWAS (n=11)
(5) Consistent MAF with ancestry-specific 1000 Genomes reference data (n=11)
Variant validation
Validate association of genetic variants with measured TGs in BioVU EA cohort (n=10)
Validate association of genetic variants with measured TGs in BioVU AA cohort (n=6)
Prespecified phenotype testing
Test associations with 21 prespecified phenotypes (phecodes) in validated variants for the EA and AA cohorts, respectively
Conduct ancestry-specific validation in AoU
PheWAS
Conduct PheWAS analysis to identify additional associations with other phenotypes (phecodes) in EA and AA cohorts, respectively
Conduct ancestry-specific validation in AoU
Additional analyses for prespecified phenotypes
Conduct analyses in the BioVU EA cohort:
(1) Additional measured LDL-C adjustment
(2) Genetic variant set-based tests
(3) Predicted gene expression
(4) Interaction in significant results across genes
Conduct analyses in the BioVU AA cohort:
(1) Additional measured LDL-C adjustment
(2) Genetic variant set-based tests
(3) IVs identified by querying an independent GLGC AA cohort

					APOA5		LPL					*ANGPTL3*	*ANGPTL4*
	phecode	**phenotype**	**case**	**control**	**rs651821**	**rs3135506**	**rs1801177**	**rs118204057**	**rs268**	**rs328**	**rs115849089**	**rs11207977**	**rs116843064**
Lipid	272.1	Hyperlipidaemia	16,809	32,956	●	●			●	●	●	●	●
	272.11	Hypercholesterolemia	8,726	32,956	●						●	●	
	272.12	Hyperglyceridaemia	894	32,956	●	●	●	●		●	●	●	●
	272.13	Mixed hyperlipidaemia	11,426	32,956	●					●	●	●	●
Cardiovascular	401.1	Essential hypertension	29,279	25,351									
	411.3	Angina pectoris	3,504	39,647						●	●		
	411.4	Coronary atherosclerosis	12,456	39,647						●	●		
	411.8	Other chronic ischemic heart disease, unspecified	4,361	39,647									
Previously reported other phenotypes	274.1	Gout	2,495	53,540									
	577	Diseases of pancreas	1,262	54,811									
	577.1	Acute pancreatitis	977	54,811									
	577.2	Chronic pancreatitis	501	54,811									
	591	Urinary tract infection	11,240	39,665									

	phecode	phenotype	case	control	APOA5 rs651821	ANGPTL3 rs11207977
Lipid	272.13	Mixed hyperlipidaemia	1847	7808		
Previously reported other phenotypes	290.11	Alzheimer's disease	87	9,087		
	386.9	Dizziness	1226	9,552		