# A sex-specific genome-wide association study of blood lipid levels in All of Us

Yunxi Li[1,2], In-Hee Lee[1], Sek Won Kong[1,3*]

Affiliations:
[1]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, 02215, USA
[2]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, 02115, USA
[3]Department of Pediatrics, Harvard Medical School, Boston, MA, 02115, USA

*Correspondence: sekwon.kong@childrens.harvard.edu

## Abstract

Despite widely acknowledged sex differences in lipid metabolism and risks for cardiovascular disease, genetic associations contributing to such differences remain incompletely characterized. Here, we performed a sex-stratified genome-wide association study (GWAS) for four lipid profiles to identify loci exhibiting differential effects between males and females. Using whole-genome sequencing data from All of Us Research Program comprising 124,920 participants of diverse ancestry, we conducted GWAS analyses separately in males, females, and a pooled cohort. Our analyses validated previous findings on genes associated with lipid metabolism. In addition, we have found 5 genes showing significant sex-heterogeneous effects, including *CELSR2* showing stronger association in males for HDL-C ($\beta_{female}$: -0.022, $\beta_{male}$: -0.045); *GPAM* in females for HDL-C ($\beta_{female}$: 0.042, $\beta_{male}$: 0.013); *PLTP* in females for HDL-C ($\beta_{female}$: 0.06, $\beta_{male}$: 0.02); *ZPR1* in females for LDL-C ($\beta_{female}$: 0.050, $\beta_{male}$: 0.014); and *CMIP* in females for TG ($\beta_{female}$: 0.037, $\beta_{male}$: 0.019). These findings highlight sex-specific genetic contributions to lipid metabolism and underscore the importance of including sex in evaluating cardiovascular risk.

## Introduction

Studies have shown that there are sex-based differences in risk of heart disease. Ischaemic heart disease was found to occur 6-8 years earlier in males compared to females, potentially attributed to differences in lipid metabolism [1]. Females with heart failure with reduced ejection fraction are also found to have a lower risk of sudden cardiac death and arterial fibrillation, despite having more symptoms and depression compared to males [1]. In alignment with these findings, two cardiovascular risk-assessment tools suggested by the American Heart Association, Predicting Risk of CVD EVENTs (PREVENT) and Pooled Cohort Equations (PCE), both take sex into consideration and provide sex-specific risk assessments [2,3]. Key risk factors for cardiovascular disease include dietary risks, air pollution, tobacco, and high low density lipoprotein cholesterol (LDL-C) levels [4].

Serum concentrations of blood lipids are biomarkers for cardiovascular disease risk stratification [5]. Lipids are transported in the serum as lipoproteins, such as low density lipoprotein (LDL) and high density lipoprotein (HDL) [6]. LDL carries cholesterol to peripheral tissues via the bloodstream, where its accumulation within arterial walls promotes plaque formation, thereby contributing to increased cardiovascular risk [7]. One study using Mendelian randomization also showed associations between LDL-C and Alzheimer's disease, inflammatory bowel disease, and type 2 diabetes [8]. HDL removes excessive cholesterol in peripheral cells through the reverse cholesterol transport (RCT) process, which contributes to HDL's inverse relationship with cardiovascular risk [9,10]. In clinical blood panel measurements, LDL-C and HDL-C refer to the cholesterol content carried by these two lipoproteins [11]. In addition to LDL-C and HDL-C, concentrations of triglycerides (TG) and total cholesterol (TC) are also commonly measured in the blood lipid panel. TG serves as an efficient energy source and can be stored in adipose tissue [12]. Given its hydrophobic nature, TG is transported in triglyceride-rich lipoproteins like very low density lipoprotein (VLDL), and high levels of plasma TG is associated with increasing risk of atherosclerosis [13]. TC measures all types of plasma cholesterol and can also be used to evaluate cardiovascular risk [14].

Several genes and pathways have been found to affect lipoprotein and lipid metabolism in the liver and central nervous system (CNS). In the hepatic pathway, LDL receptors (*LDLR*) in the liver determines hepatic uptake of LDL, which in turn affects plasma LDL levels [6]. Apolipoprotein E (*APOE*) encodes a lipid transporter protein that transports lipids into cells through LDLR [15]. LDLR degradation is mediated by proprotein convertase subtilisin/kexin type 9 (PCSK9), which binds to LDLR and facilitates its degradation in lysosome [6]. For HDL, the synthesis process starts with binding cholesterol phospholipids from liver and intestine to ApoA-I (pre-beta HDL), facilitated by ATP-binding cassette transporter A1 (ABCA1)[6]. Lecithin-cholesterol acyltransferase (LCAT) facilitates the esterification of free cholesterol bound to HDL particles [6]. From there, HDL can be cleared by the liver through 2 pathways, either mediated by scavenger receptor class B type I (SR-BI) or cholesteryl ester transfer protein (CETP) [6]. Plasma TGs are removed from chylomicrons and VLDL the bloodstream through Lipoprotein lipase (LPL)-mediated lipolysis, where LPL is transported by lipoprotein-binding protein 1 (GPIHBP1) from endothelial cells into capillary lumen [16,17]. In the CNS, limited passage of cholesterol through the blood brain barrier (BBB) makes de novo cholesterol synthesis essential for brain development and brain homeostasis [18]. Cholesterol synthesis takes place mostly during the perinatal and adolescence ages, when myelination occurs on axons [18]. Despite adults having a more stable cholesterol level, cholesterol synthesis in astrocytes is still essential for growth and maintenance of neuronal components like myelin, synapse, dendrites, and axons [18].

Studies have shown that there are sex-based differences in lipid metabolism. More specifically, pre-menopausal women are found to have higher HDL-C and lower LDL-C levels compared to men [19,20]. One theory suggests that sex hormones mostly drive these sex-based differences, as observed in postmenstrual women and women with polycystic ovary syndrome (PCOS) on hormone replacement therapy, as well as transgender individuals on hormonal treatment [21]. In females, estrogen has been found to be associated with RCT process, more specifically with high levels of estrogen promoting hepatic uptake of cholesterol efflux from macrophages and higher ApoA-I synthesis rate, which is essential to HDL synthesis [20,22]. Lower LDL-C in females can potentially be attributed to increased apolipoprotein B-100 fractional catabolic rate [23]. Other studies have also questioned the sex-hormone theory

explaining sex differences in lipid levels, considering that menopause and PCOS not only affects sex hormones but also involve body fat distribution and insulin sensitivity, which can be confounding factors when exploring sex hormone's effect on lipid levels [19].

To date, no large-scale studies have rigorously examined the genetic contributions to male–female differences. A prior study, limited by insufficient sample size and modest effect estimates, yielded inconclusive findings that warranted further investigation [24]. Previous literature exploring genetic associations of lipid profiles are also limited by unequal representation of sex or did not explore sex-specific genetic associations [25,26]. In our study, we report a GWAS study on blood lipids based on the All of Us (AOU) cohort, one of the largest US population dataset with abundant electronic health record and genomic sequencing data. We aim to validate previously reported genetic loci and potentially find novel genetic loci associated with lipid metabolism. In addition, we aim to explore sex-specific genetic influence on lipid metabolism across age groups.

# Materials and methods

## Study population

AOU research program is an initiative led by the National Institute of Health that collects electronic health record and biosample data from 510+ sites across the United States [27]. Since its start of enrollment in 2018, the program has had 865,000+ participants as of July 2025, with a primary focus on participants aged from 18 to 85+ [27]. In our study, controlled tier data release version 8 was accessed in February 2025 [28].

We included participants with LDL-C, HDL-C, TG, Total Cholesterol, short read whole genome sequencing (srWGS) data, and sex assigned at birth. We included 4 concepts of measurements for LDL-C (with logical observation identifiers names and codes (LOINC) values 13457-7, 2089-1, 18262-6, 18261-8, and 49132-4), 3 concepts for HDL-C (LOINC values 2085-9, 49130-8, and 18263-4), 5 concepts for TG (LOINC values 2571-8, 3043-7, 3048-6, 12951-0, and 1644-4), and 1 concept for TC (LOINC 2093-3). Measurements with missing measurement value or unit were removed. Measurements using operators <, ≤, >, ≥ or units other than mg/dL were also excluded from analysis. We did not convert other reported units to mg/dL because the values would be extremely high after conversion, indicating that there are likely some manual entry errors of units. Negative LDL-C values were excluded. For cholesterol measurements, we selected the highest LDL-C value for each person, and retained the participants with HDL-C, TG, and Total Cholesterol measurements available at exactly the same timepoint as LDL-C (Figure 1). A total of 124,920 participants were included in the analysis, comprising 75,127 (60.1%) females and 49,793 (39.9%) males.

Statin usage history within 1 year before the time of measurement was extracted. Given the extensive missingness of end time of statin use records and variable impact of statin withdrawal on lipid profiles, it is not feasible to comprehensively analyze if lipid profile taken under a specific timepoint was under the influence of statin [29]. However, given that the highest LDL-C measurement was selected for each participant followed by matching HDL-C, TG, and TC measured at the same time, the measurements are highly likely to not be affected by statin treatment and most likely due to genetic and dietary factors. Obesity history within 1 year before the time of measurement was extracted. For body mass index (BMI) measurements, the BMI

measurement closest to time of cholesterol measurement was extracted for each participant. Genetic ancestry prediction was extracted from the AOU data release v8 [30].

## Genome-wide association analysis

PLINK2 was used for quality control (QC) of srWGS data prior to pooled and followed by sex-specific GWAS [31]. We followed REGENIE's quality control steps on UK Biobank data, which included filtering out variants with a minor allele frequency (MAF) ≤ 1%, Hardy-Weinberg disequilibrium (HWE) p-value < $1 \times 10^{-15}$, and a genotyping rate < 99%. We performed linkage disequilibrium (LD) pruning using a sliding window of 1,000 markers, a step size of 100 markers, and an $r^2$ threshold of 0.9 [32]. After QC, there are 1,441,940 SNPs to be tested in pooled GWAS, 2,739,534 to be tested in female-specific GWAS, and 2,971,581 to be tested in male-specific GWAS.

We used REGENIE v2.0.2 for pooled and followed by sex-specific whole-genome-wide association study. We selected REGENIE for its capability of handling large scale biobank data and conducting analysis of multiple phenotypes in parallel [26,32]. Rank based inverse normal transformation was performed for each lipid type to normalize lipid panel data [32,33]. In this sense, each unit change in effect size of REGENIE output was predicting inverse normal transformed lipid panel values. Age and first 16 genetic principal components were included as covariates for all GWAS runs. For pooled analysis, sex was additionally included as a covariate. A genome-wide significance threshold of $P < 5 \times 10^{-8}$ was used to determine significant SNPs to account for multiple testing of four lipid types [34]. We also applied Bonferroni correction to account for multiple testing of 4 blood lipid profiles. Beta coefficients were estimated using the reference allele as the effect allele. Regional association plots were created using the locuszoomr package [35].

## Gene-based analysis

REGENIE tested SNPs were converted to GRCh37 to use FUMA to detect genomic risk loci for each lipid profile and sex [36]. Genomic risk loci were defined as independent significant SNPs in summary statistics and surrounding SNPs based on linkage disequilibrium (LD) structure [36]. Region plots for each genomic risk loci were plotted for both pooled and sex-specific GWAS. The nearest gene for each leading SNP was annotated by MAGMA, where the most deleterious annotation is annotated for a region of multiple overlapping genes, and more than 1 gene can be annotated in the case of intergenic regions [36,37]. Leading SNPs across genomic risk loci were cross-referenced with previously reported lipid-associated loci by comparison against the GWAS Catalog (All Associations v1.0.2), using rsID as the identifier [38].

## Sex-stratified genotype effects on lipid levels

To evaluate genotype effects on lipid levels, we fitted separate age-adjusted linear models for males, females, and the pooled cohort. Within each sex stratum, individuals were grouped according to their genotype (reference homozygote, heterozygote, or alternate homozygote). For each subset, we fitted an analysis of covariance (ANCOVA) model of the form: Value ~ genotype + age. Estimated marginal means (EMMeans) were derived using the emmeans package in R to obtain adjusted mean lipid levels for each genotype after controlling for age. Pairwise genotype

contrasts were evaluated using Tukey's significance adjustment to account for multiple comparisons. Analyses were performed separately for each sex and for the combined cohort.

# Results

## Study cohort and distribution of lipid profiles in female and male

The total cohort size was N = 124,920 including 75,127 (60.1%) female and 49,793 (30.9%) male. The majority of participants were of European ancestry (67.1%), followed by individuals of African ancestry (17.3%). The proportions of participants who had statin use history within 1 year before the date of lipid profile measurement were 13.8% (female) and 22.3% (male).

Females had higher mean LDL-C, HDL-C, and TC levels, while males had higher TG levels (Table 1). When stratified by age, females and males exhibited distinct age-related distributions of blood lipid levels (Figure 2). Females have lower median LDL-C levels in early ages (before age 50). However, females have higher median LDL-C levels after menopause, surpassing males of the same age group. This is consistent with previous findings on females having higher LDL-C from early adulthood to middle-age when compared with males [39]. Male, on the other hand, have higher TG levels than females, especially in early ages when compared to females in the same age groups. This is consistent with previous finding that males have a more atherogenic lipid profile [39]. Total Cholesterol levels for both sexes are similar in early ages, but females have higher total cholesterol levels after menopause (around age 50). This can potentially be attributed to the change in LDL-C levels as age increases. Across all age groups, females have higher median HDL-C compared to males, which is consistent with previous review literature [39].

## Genome-wide association analysis

We identified 1,766 genome-wide significant variants after Boneferroni correction taking into account multiple testing across 4 lipid types, including 541 for HDL-C, 384 for LDL-C, 445 for TG, and 396 for TC (Supplementary Table S1).

LDL-C and TC did not show strong genetic inflation across all 3 GWAS, while HDL and TG showed higher genetic inflation (Supplementary Figure 4). We did not find novel loci compared to previously reported analysis on GWAS catalog as indexed in August 2025.

Gene-based association analysis using MAGMA identified 298 statistically significant genes after applying Bonferroni correction for multiple testing across all tested genes (corrected $p < 0.05$), including 73 for HDL-C, 72 for LDL-C, 84 for TG, and 69 for TC (Supplementary Table S2). Our results validated previous findings of genes associated with lipid profiles as reported on GWAS catalog and we did not find novel loci associated with lipid profiles. We validated *LPL*, *ABCA1*, *CETP* to be associated with HDL-C; *PCSK9*, *CELSR2*, *LDLR*, and *APOE* to be associated with LDL-C; *LPL* to be associated with TG (Figure 3) [40].

## Sex-stratified genome-wide association analysis

Following the pooled GWAS analysis, we performed sex stratified GWAS analysis for all 4 blood lipids (Supplementary Figure 1). In female-specific GWAS, we identified 2,227 genome-wide significant variants after Boneferroni correction taking into account multiple testing across 4 lipid types, including 568 for HDL-C, 506 for LDL-C, 593 for TG, and 560 for

TC (Supplementary Table S3). In male-specific GWAS, we identified 1306 genome-wide significant variants, including 430 for HDL-C, 231 for LDL-C, 455 for TG, and 190 for TC (Supplementary Table S4).

Gene-based association analysis using MAGMA identified 267 statistically significant genes for female-specific GWAS run after applying Bonferroni correction for multiple testing across all tested genes (corrected $p < 0.05$), including 53 for HDL-C, 68 for LDL-C, 69 for TG, and 77 for TC (Supplementary Table S5). For male-specific GWAS run, we found 131 statistically significant genes, including 34 for HDL-C, 30 for LDL-C, 44 for TG, and 23 for TC (Supplementary Table S6). Manual review for removing spurious associations and identifying sex-specific associations further refined the list to 4 loci for HDL-C, 2 loci for LDL-C, and 1 loci for TG (Table 2). Among these sex-specific loci, *CELSR2* was male-specific for HDL-C, and all others were female-specific. Variant level and gene level tables are shown in Table 3 and Table 4, respectively.

Examination of the regional association plot surrounding *CELSR2* region revealed clusters of SNPs in strong linkage disequilibrium exhibiting genome-wide significance after Bonferroni correction in the male and pooled cohorts, but not in females (Figure 4A). We observed a male-specific dosage effect of increasing HDL-C of variant chr1:109274570:A:G (rs7528419) (Figure 5). EMMeans derived from age-adjusted models revealed genotype-dependent differences in HDL-C levels in males but not females. In the male-only cohort, individuals carrying one or two alternate alleles exhibited progressively higher adjusted means compared to reference homozygotes (Tukey-adjusted $p < 0.05$). Such an effect was not observed in the female-only cohort.

In addition, we have found *CMIP* to show female-specific significant association with TG (Figure 4B). *GPAM* and *PLTP* showed female-specific significant association with HDL-C (Supplementary Figure 2A). We observed female-specific dosage effects of increasing TG at chr16:81501185:A:G (rs2925979) for *CMIP* (Supplementary Figure 3D), and female-specific dosage effects of decreasing HDL-C at chr10:112150963:A:G (rs1129555) for *GPAM* (Supplementary Figure 3A) and at chr20:45916308:T:C (rs58847685) for *PLTP* (Supplementary Figure 3B). Interestingly, male participants carrying 1 alternative allele at chr20:45916308:T:C for *PLTP* gene showed statistically significant decrease in HDL-C, but no statistically significant difference was observed between male participants carrying 2 alternative alleles and 2 reference alleles at the loci (Supplementary Figure 3B). We have also found *ZPR1* to show female-specific significant association with LDL-C. We observed a statistically significant difference for female participants carrying 1 alternative allele at chr11:116778201:G:C (rs964184), but no dosage effect was observed at this variant (Supplementary Figure 3C).

## Discussion

We observed age-group- and sex- specific differences in lipid profile in AOU data. To investigate genetic control underlying such difference, we ran sex-specific GWAS and found sex-specific association analysis revealed novel sex-specific genetic risk loci for HDL-C, LDL-C, and TG that has not previously been reported to be sex-specifically related to lipid metabolism.

Previous research in age- and sex- specific lipid profiles has been conducted in a dutch cohort study [41]. While informative, the study was limited by the relatively homogeneous composition of the study population [41]. In our study, we observed consistent patterns of lipid

variation between females and males across age groups. Leveraging the ancestry-diverse AOU easured differences in mean lipid levels using Cohen's *d* to provide standardized effect estimates.

Although our study did not reveal the mechanisms behind sex specificity of these genes in their association to lipid profiles, future research can look into the potential mechanisms. Previous research has shown that *CELSR2–PSRC1–SORT1* gene cluster is associated with lipid levels and coronary heart disease [42]. *CELSR2* knockdown experiment in L02 human fetal hepatocyte cell line showed decreased lipid accumulation due to dysregulation of endoplasmic reticulum (ER) homeostasis, more specifically due to impairment in unfolded protein response [43]. Another mice experiment has indicated that sexual dimorphism before puberty in ER stress may be due to lack of testosterone in males [44]. We hypothesize that the male-specific association of *CELSR2* with HDL-C levels may be related to sex dimorphism in ER stress related to sex hormones. Future studies can investigate how male-specific regulation in ER stress is related to lipid metabolism.

The prevailing hypothesis for *CMIP*'s female-specific effects is an interaction with female hormones and fat distribution. An intronic *CMIP* polymorphism (rs2925979) has been associated with adiponectin levels, which is an insulin-sensitizing adipokine that raises HDL-C and lowers TG [45,46]. A previous meta-analysis of 114 studies involving 320,485 individuals showed *CMIP* to have association for waist-to-hip ratio (adjusted for BMI), indicating that *CMIP* may be involved in abdomen fat accumulation. On the other hand, another study shows that the *CMIP* risk allele (rs2925979*T) paradoxically correlates with lower BMI and waist circumference yet higher type 2 diabetes risk in females [47]. It is possible that in women the variant induces insulin resistance and dyslipidemia through pathways beyond generalized adiposity. In summary, we suspect that *CMIP* variants affect TG values in a sex-specific pattern through mediating adiponectin and shift in fat distribution. Further studies are warranted to provide more mechanistics behind such sex-specific effects.

Our study has several limitations including calculated LDL-C values. The first one is the dependent relationship between LDL-C values and other lipid measurements. While TC and HDL-C are directly measured from the serum, LDL-C is commonly calculated by the Friedewald equation, which is LDL-C = TC − (HDL-C) − (TG/5), which brings bias in itself and when LDL-C values are explored in GWAS [48]. Previous research has shown that calculated LDL-C values have higher variation and bias at lower LDL-C levels and advised against using LDL-C smaller than 70mg/dL for treatment decisions [49]. In GWAS for LDL-C values, genetic variants associated with other lipid types may also indirectly appear associated with LDL-C. In fact, in our analysis, *ZPR1* showed female-specific association with LDL-C, which was further validated by observation of statistically significant difference in LDL-C values for female participants with C/C (alt/alt) genotype compared to C/G (alt/ref), but not for males. *ZPR1* also demonstrated similar, despite weaker, female-specific association with TC. Future research can look deeper into sex-specific association of *ZPR1* with LDL-C after taking into consideration the calculative nature of many LDL-C measurements.

Another limitation of our study is the extensive missingness of statin use record. In our research cohort, 21,422 (17.1%) participants had statin usage history with start time within 1 year before the lipid measurement time in our study, among which 6,919 (32.3%) did not have statin use end time available. Extensive missingness in statin use end times, along with variable time it takes for statin to take effect in different individuals, make it not feasible to definitively determine if the LDL-C values used in study was affected by statin. Previous research in lipid GWAS using electronic health record or biobank data have handled repeated lipid measurements

and lipid-lowering therapy in several ways. One study defined pre-treatment LDL-C to be the median of all LDL-C values before the first mention of any statin treatment [50]. Another biobank-level study on LDL-C adjusted values for statin use by dividing measured LDL-C by 0.7 to approximate pre-treatment levels [51]. However, such adjustment did not take into account interindividual variability of lipid lowering effect of statin [52]. Given the extent of missingness in statin end dates in our dataset and the resulting uncertainty about treatment status at any given blood draw, we opted instead to use the highest recorded LDL-C value across each participant's lifetime electronic health record, reasoning that this "peak" LDL-C is most likely to precede initiation of lipid-lowering therapy and that the accompanying lipid panel is therefore less likely to be influenced by statins. Nevertheless, future research with access to more comprehensive statin use history can more thoroughly investigate genetic risk factors underlying hyperlipidemia while taking into consideration the effect of lipid lowering medications.

## Conclusion

In our research, we investigated lipid panel values across 124,920 participants in the AOU research program. We characterized sex-stratified HDL-C, LDL-C, TG, TC distributions across lifespan. Our findings revealed sex-specific gene associations with lipid profiles such as *CELSR2* (for male) and *CMIP* (for female). The study paves the way for further evaluation in mechanisms underlying sex-specific genetic associations with lipid profiles and provides foundation for sex-specific risk evaluation of cardiovascular risk.

## Funding Statement

This work was supported by the NIH R01NS129188 and U54HG012513 (SWK).

## Acknowledgement

We gratefully acknowledge AOU participants for their contributions, without whom this research would not have been possible.We also thank the National Institutes of Health's AOU Research Program for making available the participant data examined in this study.

References


1. Regitz-Zagrosek V, Gebhard C. Gender medicine: effects of sex and gender on cardiovascular disease manifestation and outcomes. Nat Rev Cardiol. 2023 Apr;20(4):236–47.

2. Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, Greenland P, Lackland DT, Levy D, O'Donnell CJ, Robinson JG, Schwartz JS, Shero ST, Smith SC Jr, Sorlie P, Stone NJ, Wilson PWF, Jordan HS, Nevo L, Wnek J, Anderson JL, Halperin JL, Albert NM, Bozkurt B, Brindis RG, Curtis LH, DeMets D, Hochman JS, Kovacs RJ, Ohman EM, Pressler SJ, Sellke FW, Shen WK, Smith SC Jr, Tomaselli GF, American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines: A report of the American college of cardiology/American heart association task force on practice guidelines. Circulation. 2014 June 24;129(25 Suppl 2):S49–73.

3. Khan SS, Matsushita K, Sang Y, Ballew SH, Grams ME, Surapaneni A, Blaha MJ, Carson AP, Chang AR, Ciemins E, Go AS, Gutierrez OM, Hwang SJ, Jassal SK, Kovesdy CP, Lloyd-Jones DM, Shlipak MG, Palaniappan LP, Sperling L, Virani SS, Tuttle K, Neeland IJ, Chow SL, Rangaswami J, Pencina MJ, Ndumele CE, Coresh J, Chronic Kidney Disease Prognosis Consortium and the American Heart Association Cardiovascular-Kidney-Metabolic Science Advisory Group. Development and validation of the American heart association's PREVENT equations. Circulation. 2024 Feb 6;149(6):430–49.

4. Vaduganathan M, Mensah GA, Turco JV, Fuster V, Roth GA. The global burden of cardiovascular diseases and risk: A compass for future health. J Am Coll Cardiol. 2022 Dec 20;80(25):2361–71.

5. Garady L, Soota A, Shouche Y, Chandrachari KP, K S V, Shankar P, Sharma SV, C K, Munnyal S, Gopi A, Devyani A. A narrative review of the role of blood biomarkers in the risk prediction of cardiovascular diseases. Cureus. 2024 Dec;16(12):e74899.

6. Feingold KR. Introduction to lipids and lipoproteins. In: Endotext. South Dartmouth (MA): MDText.com, Inc.; 2000.

7. Bhatt A. Cholesterol: Understanding HDL vs. LDL [Internet]. Harvard Health. 2020 [cited 2025 Aug 17]. Available from: https://www.health.harvard.edu/blog/understanding-cholesterol-hdl-vs-ldl-2018041213608

8. Schmidt AF, Joshi R, Gordillo-Marañón M, Drenos F, Charoen P, Giambartolomei C, Bis JC, Gaunt TR, Hughes AD, Lawlor DA, Wong A, Price JF, Chaturvedi N, Wannamethee G, Franceschini N, Kivimaki M, Hingorani AD, Finan C. Biomedical consequences of elevated cholesterol-containing lipoproteins and apolipoproteins on cardiovascular and non-cardiovascular outcomes. Commun Med (Lond). 2023 Jan 20;3(1):9.

9. Playford MP, Neufeld EB, Zubirán R, Remaley AT. Reverse cholesterol transport: current assay methods, alterations with disease and response to therapeutic intervention. Front Cardiovasc Med. 2025 July 10;12(1608384):1608384.

10. Kang H, Song J, Cheng Y. HDL regulates the risk of cardiometabolic and inflammatory-related diseases: Focusing on cholesterol efflux capacity. Int Immunopharmacol. 2024 Sept 10;138(112622):112622.

11. Wadhera RK, Steen DL, Khan I, Giugliano RP, Foody JM. A review of low-density lipoprotein

cholesterol, treatment strategies, and its impact on cardiovascular disease morbidity and mortality. J Clin Lipidol. 2016 May;10(3):472–89.

12. Metabolism of Triglyceride-Rich Lipoproteins. Available from: https://www.ncbi.nlm.nih.gov/books/NBK584295/

13. Borén J, Taskinen MR, Björnson E, Packard CJ. Metabolism of triglyceride-rich lipoproteins in health and dyslipidaemia. Nat Rev Cardiol. 2022 Sept;19(9):577–92.

14. Total cholesterol [Internet]. Mount Sinai Health System. [cited 2025 Aug 17]. Available from: https://www.mountsinai.org/health-library/tests/total-cholesterol

15. Yang LG, March ZM, Stephenson RA, Narayan PS. Apolipoprotein E in lipid metabolism and neurodegenerative disease. Trends Endocrinol Metab. 2023 Aug;34(8):430–45.

16. Basu D, Goldberg IJ. Regulation of lipoprotein lipase-mediated lipolysis of triglycerides. Curr Opin Lipidol. 2020 June;31(3):154–60.

17. Young SG, Fong LG, Beigneux AP, Allan CM, He C, Jiang H, Nakajima K, Meiyappan M, Birrane G, Ploug M. GPIHBP1 and lipoprotein lipase, partners in plasma triglyceride metabolism. Cell Metab. 2019 July 2;30(1):51–65.

18. Vanherle S, Loix M, Miron VE, Hendriks JJA, Bogie JFJ. Lipid metabolism, remodelling and intercellular transfer in the CNS. Nat Rev Neurosci. 2025 Apr;26(4):214–31.

19. Wang X, Magkos F, Mittendorfer B. Sex differences in lipid and lipoprotein metabolism: it's not just about sex hormones. J Clin Endocrinol Metab. 2011 Apr;96(4):885–93.

20. Schaefer EJ, Zech LA, Jenkins LL, Bronzert TJ, Rubalcaba EA, Lindgren FT, Aamodt RL, Brewer HB Jr. Human apolipoprotein A-I and A-II metabolism. J Lipid Res. 1982 Aug;23(6):850–62.

21. Robinson GA, Peng J, Peckham H, Radziszewska A, Butler G, Pineda-Torra I, Jury EC, Ciurtin C. Sex hormones drive changes in lipoprotein metabolism. iScience. 2021 Nov 19;24(11):103257.

22. Palmisano BT, Zhu L, Eckel RH, Stafford JM. Sex differences in lipid and lipoprotein metabolism. Mol Metab. 2018 Sept;15:45–55.

23. Karjalainen A, Heikkinen J, Savolainen MJ, Bäckström AC, Kesäniemi YA. Mechanisms regulating LDL metabolism in subjects on peroral and transdermal estrogen replacement therapy. Arterioscler Thromb Vasc Biol. 2000 Apr;20(4):1101–6.

24. Tabassum R, Ruotsalainen S, Ottensmann L, Gerl MJ, Klose C, Tukiainen T, Pirinen M, Simons K, Widén E, Ripatti S. Lipidome- and genome-wide study to understand sex differences in circulatory lipids. J Am Heart Assoc. 2022 Oct 4;11(19):e027103.

25. Klarin D, Damrauer SM, Cho K, Sun YV, Teslovich TM, Honerlaw J, Gagnon DR, DuVall SL, Li J, Peloso GM, Chaffin M, Small AM, Huang J, Tang H, Lynch JA, Ho YL, Liu DJ, Emdin CA, Li AH, Huffman JE, Lee JS, Natarajan P, Chowdhury R, Saleheen D, Vujkovic M, Baras A, Pyarajan S, Di Angelantonio E, Neale BM, Naheed A, Khera AV, Danesh J, Chang KM, Abecasis G, Willer C, Dewey FE, Carey DJ, Global Lipids Genetics Consortium, Myocardial Infarction Genetics (MIGen) Consortium, Geisinger-Regeneron DiscovEHR Collaboration, VA Million Veteran Program, Concato J, Gaziano JM, O'Donnell CJ, Tsao PS, Kathiresan S, Rader DJ, Wilson PWF, Assimes TL. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. Nat Genet.

2018 Nov;50(11):1514–23.

26. Koyama S, Yu Z, Choi SH, Jurgens SJ, Selvaraj MS, Klarin D, Huffman JE, Clarke SL, Trinh MN, Ravi A, Dron JS, Spinks C, Surakka I, Bhatnagar A, Lannery K, Hornsby W, Damrauer SM, Chang KM, Lynch JA, Assimes TL, Tsao PS, Rader DJ, Cho K, Peloso GM, Ellinor PT, Sun YV, Wilson PW, Program MV, Natarajan P. Exome wide association study for blood lipids in 1,158,017 individuals from diverse populations [Internet]. medRxiv. 2024. Available from: http://medrxiv.org/lookup/doi/10.1101/2024.09.17.24313718

27. Data Snapshots [Internet]. [cited 2025 Aug 4]. Available from: https://www.researchallofus.org/data-tools/data-snapshots/

28. Curated Data Repository (CDR) version 8 Release Notes [Internet]. User Support. [cited 2025 Aug 18]. Available from: https://support.researchallofus.org/hc/en-us/articles/30294451486356-Curated-Data-Repository-CDR-version-8-Release-Notes

29. Alvarez-Jimenez L, Morales-Palomo F, Moreno-Cabañas A, Mora-Gonzalez D, Turrillas MDCM, Mora-Rodriguez R. Time-course atherogenic blood lipid response to statin discontinuation in dyslipidemic adults. Nutr Metab Cardiovasc Dis. 2024 Oct;34(10):2334–43.

30. Controlled CDR Directory [Internet]. User Support. [cited 2025 Aug 18]. Available from: https://support.researchallofus.org/hc/en-us/articles/29475233432212-Controlled-CDR-Directory

31. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015 Feb 25;4(1):7.

32. Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, Benner C, O'Dushlaine C, Barber M, Boutkov B, Habegger L, Ferreira M, Baras A, Reid J, Abecasis G, Maxwell E, Marchini J. Computationally efficient whole-genome regression for quantitative and binary traits. Nat Genet. 2021 July;53(7):1097–103.

33. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, Beckmann JS, Bragg-Gresham JL, Chang HY, Demirkan A, Den Hertog HM, Do R, Donnelly LA, Ehret GB, Esko T, Feitosa MF, Ferreira T, Fischer K, Fontanillas P, Fraser RM, Freitag DF, Gurdasani D, Heikkilä K, Hyppönen E, Isaacs A, Jackson AU, Johansson Å, Johnson T, Kaakinen M, Kettunen J, Kleber ME, Li X, Luan J 'an, Lyytikäinen LP, Magnusson PKE, Mangino M, Mihailov E, Montasser ME, Müller-Nurasyid M, Nolte IM, O'Connell JR, Palmer CD, Perola M, Petersen AK, Sanna S, Saxena R, Service SK, Shah S, Shungin D, Sidore C, Song C, Strawbridge RJ, Surakka I, Tanaka T, Teslovich TM, Thorleifsson G, Van den Herik EG, Voight BF, Volcik KA, Waite LL, Wong A, Wu Y, Zhang W, Absher D, Asiki G, Barroso I, Been LF, Bolton JL, Bonnycastle LL, Brambilla P, Burnett MS, Cesana G, Dimitriou M, Doney ASF, Döring A, Elliott P, Epstein SE, Ingi Eyjolfsson G, Gigante B, Goodarzi MO, Grallert H, Gravito ML, Groves CJ, Hallmans G, Hartikainen AL, Hayward C, Hernandez D, Hicks AA, Holm H, Hung YJ, Illig T, Jones MR, Kaleebu P, Kastelein JJP, Khaw KT, Kim E, Klopp N, Komulainen P, Kumari M, Langenberg C, Lehtimäki T, Lin SY, Lindström J, Loos RJF, Mach F, McArdle WL, Meisinger C, Mitchell BD, Müller G, Nagaraja R, Narisu N, Nieminen TVM, Nsubuga RN, Olafsson I, Ong KK, Palotie A, Papamarkou T, Pomilla C, Pouta A, Rader DJ, Reilly MP, Ridker PM, Rivadeneira F, Rudan I, Ruokonen A, Samani N, Scharnagl H, Seeley J, Silander K, Stančáková A, Stirrups K, Swift AJ, Tiret L, Uitterlinden AG, van Pelt LJ, Vedantam S, Wainwright N, Wijmenga C, Wild SH, Willemsen G, Wilsgaard T, Wilson JF, Young EH, Zhao JH, Adair LS, Arveiler D, Assimes TL, Bandinelli S, Bennett F, Bochud M, Boehm BO, Boomsma DI, Borecki IB, Bornstein SR, Bovet P, Burnier M,

Campbell H, Chakravarti A, Chambers JC, Chen YDI, Collins FS, Cooper RS, Danesh J, Dedoussis G, de Faire U, Feranil AB, Ferrières J, Ferrucci L, Freimer NB, Gieger C, Groop LC, Gudnason V, Gyllensten U, Hamsten A, Harris TB, Hingorani A, Hirschhorn JN, Hofman A, Hovingh GK, Hsiung CA, Humphries SE, Hunt SC, Hveem K, Iribarren C, Järvelin MR, Jula A, Kähönen M, Kaprio J, Kesäniemi A, Kivimaki M, Kooner JS, Koudstaal PJ, Krauss RM, Kuh D, Kuusisto J, Kyvik KO, Laakso M, Lakka TA, Lind L, Lindgren CM, Martin NG, März W, McCarthy MI, McKenzie CA, Meneton P, Metspalu A, Moilanen L, Morris AD, Munroe PB, Njølstad I, Pedersen NL, Power C, Pramstaller PP, Price JF, Psaty BM, Quertermous T, Rauramaa R, Saleheen D, Salomaa V, Sanghera DK, Saramies J, Schwarz PEH, Sheu WHH, Shuldiner AR, Siegbahn A, Spector TD, Stefansson K, Strachan DP, Tayo BO, Tremoli E, Tuomilehto J, Uusitupa M, van Duijn CM, Vollenweider P, Wallentin L, Wareham NJ, Whitfield JB, Wolffenbuttel BHR, Ordovas JM, Boerwinkle E, Palmer CNA, Thorsteinsdottir U, Chasman DI, Rotter JI, Franks PW, Ripatti S, Cupples LA, Sandhu MS, Rich SS, Boehnke M, Deloukas P, Kathiresan S, Mohlke KL, Ingelsson E, Abecasis GR, Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013 Nov;45(11):1274–83.

34. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008 May;32(4):381–5.

35. Lewis MJ, Wang S. locuszoomr: an R package for visualizing publication-ready regional gene locus plots. Bioinform Adv. 2025 Jan 16;5(1):vbaf006.

36. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017 Nov 28;8(1):1826.

37. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sept;38(16):e164.

38. Burdett T, Hastings E, Welter D, SPOT, EMBL-EBI, NHGRI. GWAS Catalog [Internet]. [cited 2025 Aug 18]. Available from: https://www.ebi.ac.uk/gwas/docs/file-downloads

39. Holven KB, Roeters van Lennep J. Sex differences in lipids: A life course approach. Atherosclerosis. 2023 Nov;384(117270):117270.

40. Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, Heath SC, Timpson NJ, Najjar SS, Stringham HM, Strait J, Duren WL, Maschio A, Busonero F, Mulas A, Albai G, Swift AJ, Morken MA, Narisu N, Bennett D, Parish S, Shen H, Galan P, Meneton P, Hercberg S, Zelenika D, Chen WM, Li Y, Scott LJ, Scheet PA, Sundvall J, Watanabe RM, Nagaraja R, Ebrahim S, Lawlor DA, Ben-Shlomo Y, Davey-Smith G, Shuldiner AR, Collins R, Bergman RN, Uda M, Tuomilehto J, Cao A, Collins FS, Lakatta E, Lathrop GM, Boehnke M, Schlessinger D, Mohlke KL, Abecasis GR. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. Nat Genet. 2008 Feb;40(2):161–9.

41. Balder JW, de Vries JK, Nolte IM, Lansberg PJ, Kuivenhoven JA, Kamphuisen PW. Lipid and lipoprotein reference values from 133,450 Dutch Lifelines participants: Age- and gender-specific baseline lipid values and percentiles. J Clin Lipidol. 2017 July;11(4):1055–64.e6.

42. Arvind P, Nair J, Jambunathan S, Kakkar VV, Shanker J. CELSR2-PSRC1-SORT1 gene expression and association with coronary artery disease and plasma lipid levels in an Asian Indian cohort. J Cardiol. 2014 Nov;64(5):339–46.

43. Tan J, Che Y, Liu Y, Hu J, Wang W, Hu L, Zhou Q, Wang H, Li J. CELSR2 deficiency suppresses lipid accumulation in hepatocyte by impairing the UPR and elevating ROS level. FASEB J. 2021 Oct;35(10):e21908.

44. Rossetti CL, de Oliveira Costa HM, Barthem CS, da Silva MH, de Carvalho DP, da-Silva WS. Sexual dimorphism of liver endoplasmic reticulum stress susceptibility in prepubertal rats and the effect of sex steroid supplementation. Exp Physiol. 2019 May;104(5):677–90.

45. Yanai H, Yoshida H. Beneficial effects of adiponectin on glucose and lipid metabolism and atherosclerotic progression: Mechanisms and perspectives. Int J Mol Sci. 2019 Mar 8;20(5):1190.

46. Wu Y, Gao H, Li H, Tabara Y, Nakatochi M, Chiu YF, Park EJ, Wen W, Adair LS, Borja JB, Cai Q, Chang YC, Chen P, Croteau-Chonka DC, Fogarty MP, Gan W, He CT, Hsiung CA, Hwu CM, Ichihara S, Igase M, Jo J, Kato N, Kawamoto R, Kuzawa CW, Lee JJM, Liu J, Lu L, McDade TW, Osawa H, Sheu WHH, Teo Y, Vadlamudi S, Van Dam RM, Wang Y, Xiang YB, Yamamoto K, Ye X, Young TL, Zheng W, Zhu J, Shu XO, Shin C, Jee SH, Chuang LM, Miki T, Yokota M, Lin X, Mohlke KL, Tai ES. A meta-analysis of genome-wide association studies for adiponectin levels in East Asians identifies a novel locus near WDR11-FGFR2. Hum Mol Genet. 2014 Feb 15;23(4):1108–19.

47. Shi G, Yin RX, Guo T, Sun JQ, Chen X, Pan L, Huang J, Li WJ, Gao H, Zhang QH. Association of the CMIP rs 2925979 polymorphism and several environmental factors with serum lipid levels in the Chinese Jing and Han populations. 2016; Available from: https://e-century.us/files/ijcep/9/12/ijcep0034515.pdf

48. 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 June;18(6):499–502.

49. Meeusen JW, Snozek CL, Baumann NA, Jaffe AS, Saenger AK. Reliability of calculated low-density lipoprotein cholesterol. Am J Cardiol. 2015 Aug 15;116(4):538–40.

50. Wei WQ, Li X, Feng Q, Kubo M, Kullo IJ, Peissig PL, Karlson EW, Jarvik GP, Lee MTM, Shang N, Larson EA, Edwards T, Shaffer CM, Mosley JD, Maeda S, Horikoshi M, Ritchie M, Williams MS, Larson EB, Crosslin DR, Bland HT, Pacheco JA, Rasmussen-Torvik LJ, Cronkite D, Hripcsak G, Cox NJ, Wilke RA, Stein CM, Rotter JI, Momozawa Y, Roden DM, Krauss RM, Denny JC. LPA variants are associated with residual cardiovascular risk in patients receiving statins. Circulation. 2018 Oct 23;138(17):1839–49.

51. Natarajan P, Peloso GM, Zekavat SM, Montasser M, Ganna A, Chaffin M, Khera AV, Zhou W, Bloom JM, Engreitz JM, Ernst J, O'Connell JR, Ruotsalainen SE, Alver M, Manichaikul A, Johnson WC, Perry JA, Poterba T, Seed C, Surakka IL, Esko T, Ripatti S, Salomaa V, Correa A, Vasan RS, Kellis M, Neale BM, Lander ES, Abecasis G, Mitchell B, Rich SS, Wilson JG, Cupples LA, Rotter JI, Willer CJ, Kathiresan S, NHLBI TOPMed Lipids Working Group. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. Nat Commun. 2018 Aug 23;9(1):3391.

52. Makhmudova U, Schatz U, Perakakis N, Kassner U, Schumann F, Axthelm C, Stürzebecher P, Sinning DL, Doevelaar A, Rohn B, Westhoff T, Vogt A, Scholl M, Kästner U, Geiling JA, Stach K, Mensch J, Lorenz E, Paitazoglou C, Eitel I, Baessler A, Steinhagen-Thiessen E, Koenig W, Schulze PC, Landmesser U, Laufs U, Weingärtner O, German Inclisiran Network (GIN). High interindividual variability in LDL-cholesterol reductions after inclisiran administration in a real-world multicenter setting in Germany. Clin Res Cardiol. 2023 Nov;112(11):1639–49.

# Tables

**Table 1.** Baseline characteristics of study cohort

Table 1: Baseline characteristics of study cohort

| | **Female** (N = 75,127) | **Male** (N = 49,793) |
|---|---|---|
| **Age (years)** | | |
| Mean (SD) | 51.60 (15.24) | 55.22 (14.69) |
| **BMI (kg/m$^2$)** | | |
| Mean (SD) | 30.55 (9.05) | 29.63 (7.16) |
| **Ancestry** | | |
| African/African American | 14,144 (18.8%) | 7,461 (15.0%) |
| American Admixed/Latino | 9,714 (12.9%) | 5,028 (10.1%) |
| East Asian | 1,741 (2.3%) | 984 (2.0%) |
| European | 48,509 (64.6%) | 35,340 (71.0%) |
| Middle Eastern | 237 (0.3%) | 245 (0.5%) |
| South Asian | 782 (1.0%) | 735 (1.5%) |
| **Ethnicity** | | |
| Hispanic or Latino | 10,570 (14.1%) | 5,206 (10.5%) |
| Not Hispanic or Latino | 62,703 (83.5%) | 43,130 (86.6%) |
| PMI: Prefer Not To Answer | 312 (0.4%) | 234 (0.5%) |
| PMI: Skip | 868 (1.2%) | 707 (1.4%) |
| Race Ethnicity None Of These | 674 (0.9%) | 516 (1.0%) |
| **Statin use within 1 year** | | |
| Yes | 10,332 (13.8%) | 11,090 (22.3%) |
| No | 64,795 (86.2%) | 38,703 (77.7%) |
| **Obesity within 1 year** | | |
| Yes | 4,218 (5.6%) | 3,363 (6.8%) |
| No | 70,909 (94.4%) | 46,430 (93.2%) |
| **LDL-C (mg/dL)** | | |
| Mean ± SD (range) | $126.47 \pm 40.44$ (3 – 473) | $122.80 \pm 40.37$ (5 – 1035) |
| **HDL-C (mg/dL)** | | |
| Mean ± SD (range) | $59.32 \pm 17.51$ (3 – 214) | $48.31 \pm 14.17$ (2 – 330) |
| **TG (mg/dL)** | | |
| Mean ± SD (range) | $120.21 \pm 77.36$ (2 – 5709) | $135.38 \pm 86.08$ (9 – 7920) |
| **TC (mg/dL)** | | |
| Mean ± SD (range) | $209.04 \pm 47.11$ (17 – 626) | $196.64 \pm 45.83$ (4 – 1801) |

*Extreme values of 10,000,000 mg/dL were excluded from calculating lipid measurement summary statistics as they are likely due to unit mistakes in data entry.

**Table 2** Sex-specific genes associated with blood lipid levels

| | HDL-C | LDL-C | TG | TC |
|---|---|---|---|---|
| **Male** | *CELSR2* | - | - | - |
| **Female** | *GPAM, PLTP* | *ZPR1* | *CMIP* | - |

**Table 3.** Sex-specific variants associated with blood lipid levels

| ID | VEP_mapped_gene | BETA* | P | PHENOTYPE | GWAS Run |
|---|---|---|---|---|---|
| chr1:109274570:A:G | *CELSR2* | -0.0220579 | 0.000130701 | HDL | Female-only |
| chr1:109274570:A:G | *CELSR2* | -0.0448627 | 1.92E-10 | HDL | Male-only |
| chr1:109274570:A:G | *CELSR2* | -0.0314057 | 2.36E-12 | HDL | Pooled |
| chr10:112150963:A:G | *GPAM* | 0.0321485 | 3.02E-15 | HDL | Pooled |
| chr10:112153464:T:C | *GPAM* | 0.0422569 | 3.34E-16 | HDL | Female-only |
| chr10:112153464:T:C | *GPAM* | 0.0129896 | 0.0392735 | HDL | Male-only |
| chr11:116778201:G:C | *ZPR1* | 0.0494462 | 7.68E-14 | LDL | Female-only |
| chr11:116778201:G:C | *ZPR1* | 0.0140275 | 0.105052 | LDL | Male-only |
| chr16:81486161:A:G | *CMIP* | 0.0373586 | 2.74E-11 | TG | Female-only |
| chr16:81486161:A:G | *CMIP* | 0.0186053 | 0.00833067 | TG | Male-only |
| chr16:81486161:A:G | *CMIP* | 0.0297154 | 1.49E-11 | TG | Pooled |
| chr20:45916308:T:C | *PLTP* | 0.0562298 | 4.39E-19 | HDL | Female-only |
| chr20:45916308:T:C | *PLTP* | 0.0240854 | 0.00163817 | HDL | Male-only |
| chr20:45916308:T:C | *PLTP* | 0.0429295 | 1.43E-18 | HDL | Pooled |

*Beta coefficients were estimated using the reference allele as the effect allele.

**Table 4.** Sex-specific genes associated with blood lipid levels

| Gene | Locus | P | NSNPs | Phenotype | GWAS Run |
|---|---|---|---|---|---|
| *CELSR2* | chr1:109249539-109275751 | 0.00062272 | 36 | HDL-C | Female-only |
| *CELSR2* | chr1:109249539-109275751 | 4.87E-07 | 40 | HDL-C | Male-only |
| *CELSR2* | chr1:109249539-109275751 | 1.65E-09 | 23 | HDL-C | Pooled |
| *GPAM* | chr10:112149865-112215416 | 1.21E-13 | 50 | HDL-C | Female-only |
| *GPAM* | chr10:112149865-112215416 | 0.20378 | 54 | HDL-C | Male-only |
| *GPAM* | chr10:112149865-112215416 | 4.54E-10 | 36 | HDL-C | Pooled |
| *PLTP* | chr20:45898621-45912235 | 3.33E-15 | 18 | HDL-C | Female-only |
| *PLTP* | chr20:45898621-45912235 | 0.00033633 | 19 | HDL-C | Male-only |
| *PLTP* | chr20:45898621-45912235 | 0.0068117 | 9 | HDL-C | Pooled |
| *ZPR1* | chr11:116773799-116788046 | 2.27E-12 | 9 | LDL-C | Female-only |
| *ZPR1* | chr11:116773799-116788046 | 0.44892 | 11 | LDL-C | Male-only |
| *ZPR1* | chr11:116773799-116788046 | 0.48202 | 2 | LDL-C | Pooled |
| *CMIP* | chr16:81444808-81711762 | 4.20E-15 | 518 | TG | Female-only |
| *CMIP* | chr16:81444808-81711762 | 0.080907 | 575 | TG | Male-only |
| *CMIP* | chr16:81444808-81711762 | 9.40E-11 | 291 | TG | Pooled |

# Figures

## Figure 1. Flowchart of cohort selection

The figure outlines the sequential filtering of participants with whole-genome sequencing data, lipid measurements, and demographic information. After removing individuals with incomplete or invalid lipid values and those without complete paired lipid measurements, a final cohort of 124,920 participants (75,127 female; 49,793 male) was retained for genome-wide association analysis (GWAS) (srWGS: short-read whole genome sequencing data; HDL-C: high density lipoprotein cholesterol; LDL-C: low density lipoprotein cholesterol; TG: triglyceride; TC: total cholesterol; mg/dL: milligrams per deciliter; Others: participants whose sex-at-birth field contained "I prefer not to answer," "pmi: skip," "sex at birth none of these," or "intersex").

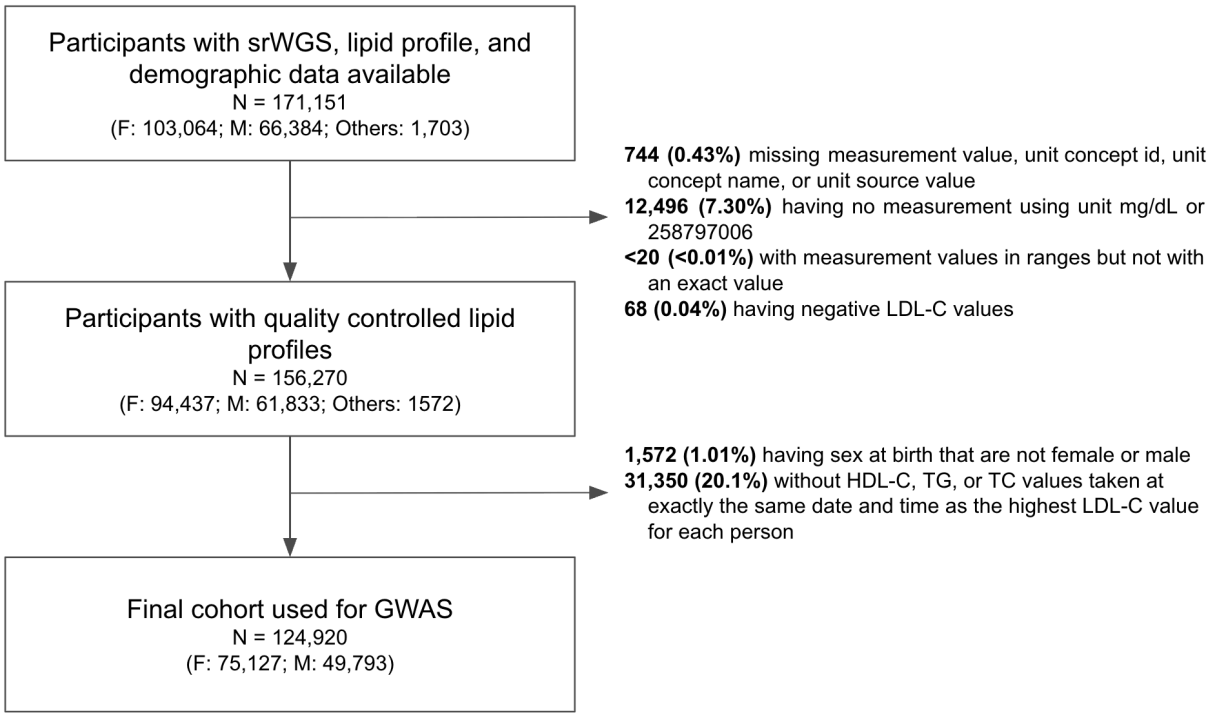
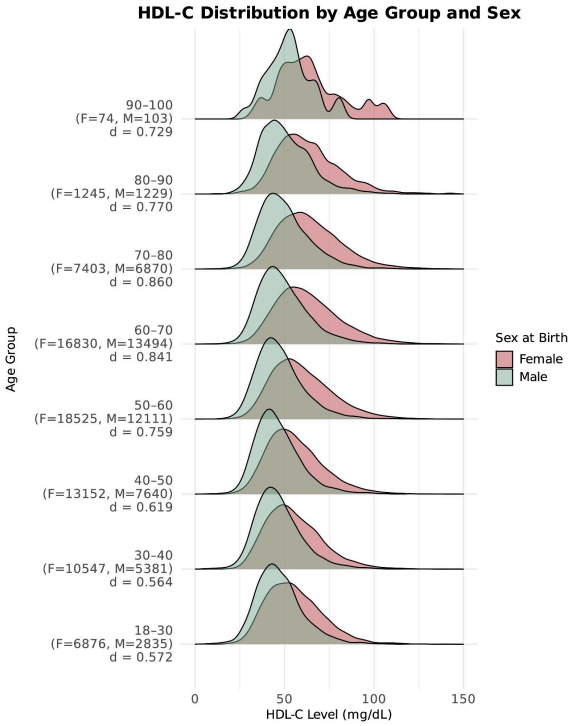
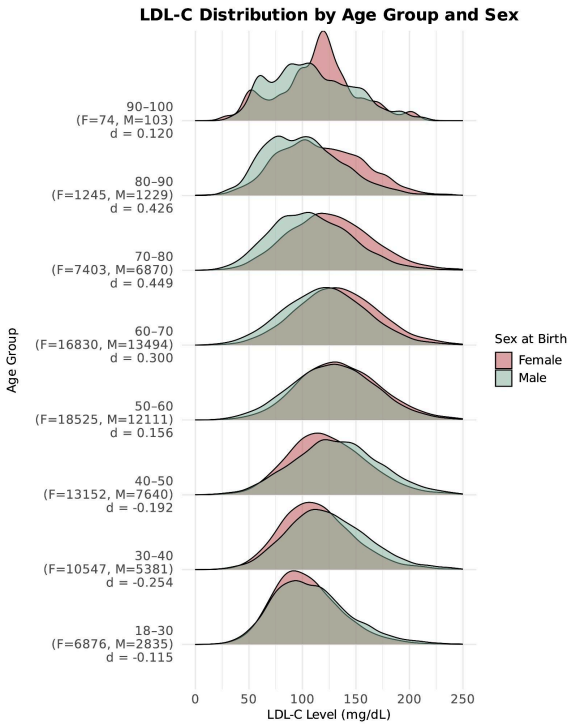
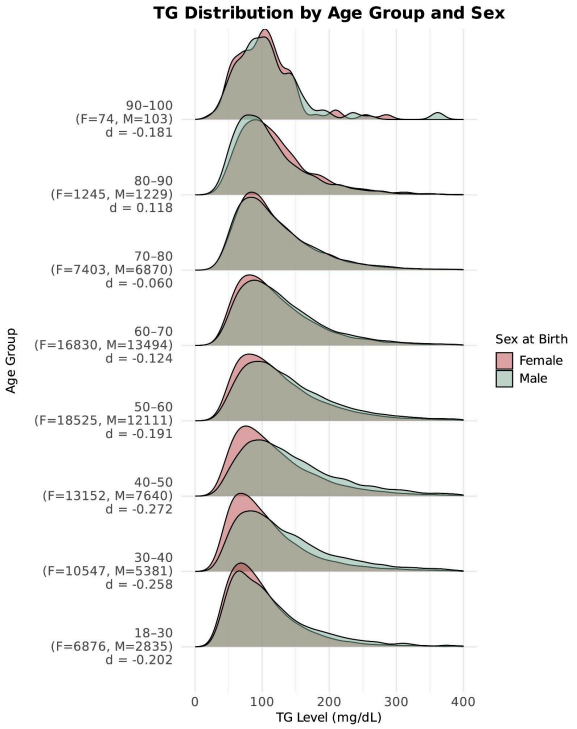

Participants with srWGS, lipid profile, and demographic data available
N = 171,151
(F: 103,064; M: 66,384; Others: 1,703)
744 (0.43%) missing measurement value, unit concept id, unit concept name, or unit source value
12,496 (7.30%) having no measurement using unit mg/dL or 258797006
<20 (<0.01%) with measurement values in ranges but not with an exact value
68 (0.04%) having negative LDL-C values
Participants with quality controlled lipid profiles
N = 156,270
(F: 94,437; M: 61,833; Others: 1572)
1,572 (1.01%) having sex at birth that are not female or male
31,350 (20.1%) without HDL-C, TG, or TC values taken at exactly the same date and time as the highest LDL-C value for each person
Final cohort used for GWAS
N = 124,920
(F: 75,127; M: 49,793)

**Figure 2.** Age and sex differential distribution of HDL-C, LDL-C, TG, and Total Cholesterol

Density plots show HDL-C and LDL-C distributions across age groups for females and males. For each age stratum, sample sizes and Cohen's d are reported where Cohen's d quantifies the standardized mean difference between sexes using the following equation (positive d means higher mean in females).


HDL-C Distribution by Age Group and Sex
90–100
(F=74, M=103)
d = 0.729
80–90
(F=1245, M=1229)
d = 0.770
70–80
(F=7403, M=6870)
d = 0.860
60–70
(F=16830, M=13494)
d = 0.841
50–60
(F=18525, M=12111)
d = 0.759
40–50
(F=13152, M=7640)
d = 0.619
30–40
(F=10547, M=5381)
d = 0.564
18–30
(F=6876, M=2835)
d = 0.572
Age Group
0
50
100
150
HDL-C Level (mg/dL)
Sex at Birth
Female
Male


LDL-C Distribution by Age Group and Sex
90–100
(F=74, M=103)
d = 0.120
80–90
(F=1245, M=1229)
d = 0.426
70–80
(F=7403, M=6870)
d = 0.449
60–70
(F=16830, M=13494)
d = 0.300
50–60
(F=18525, M=12111)
d = 0.156
40–50
(F=13152, M=7640)
d = -0.192
30–40
(F=10547, M=5381)
d = -0.254
18–30
(F=6876, M=2835)
d = -0.115
Age Group
0
50
100
150
200
250
LDL-C Level (mg/dL)
Sex at Birth
Female
Male


TG Distribution by Age Group and Sex
90–100
(F=74, M=103)
d = -0.181
80–90
(F=1245, M=1229)
d = 0.118
70–80
(F=7403, M=6870)
d = -0.060
60–70
(F=16830, M=13494)
d = -0.124
50–60
(F=18525, M=12111)
d = -0.191
40–50
(F=13152, M=7640)
d = -0.272
30–40
(F=10547, M=5381)
d = -0.258
18–30
(F=6876, M=2835)
d = -0.202
Age Group
0
100
200
300
400
TG Level (mg/dL)
Sex at Birth
Female
Male

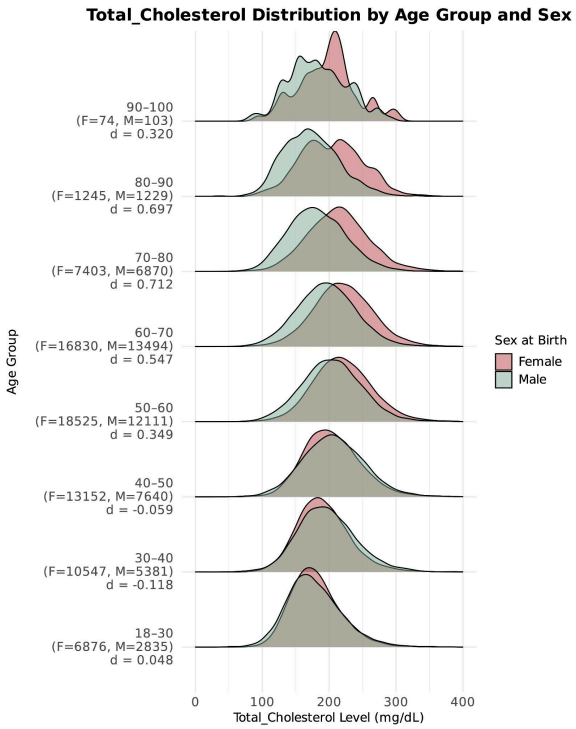

Total_Cholesterol Distribution by Age Group and Sex
90–100
(F=74, M=103)
d = 0.320
80–90
(F=1245, M=1229)
d = 0.697
70–80
(F=7403, M=6870)
d = 0.712
60–70
(F=16830, M=13494)
d = 0.547
50–60
(F=18525, M=12111)
d = 0.349
40–50
(F=13152, M=7640)
d = -0.059
30–40
(F=10547, M=5381)
d = -0.118
18–30
(F=6876, M=2835)
d = 0.048
Age Group
0
100
200
300
400
Total_Cholesterol Level (mg/dL)
Sex at Birth
Female
Male

**Figure 3.** Manhattan plots for pooled GWAS

Manhattan plots display pooled genome-wide association results for HDL-C (Figure 3A), LDL-C (Figure 3B), TG (Figure 3C), and TC (Figure 3D) in the pooled cohort. Each point represents a genetic variant plotted by chromosomal position and $-\log_{10}(P)$. The horizontal dashed line denotes the genome-wide significance threshold after Bonferroni correction.

**Figure 3A.** Manhattan plots for pooled GWAS for HDL-C.

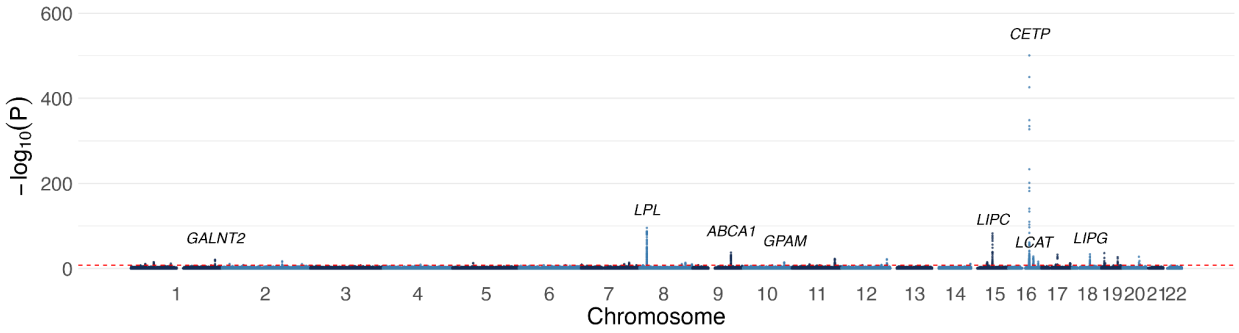



**Figure 3B.** Manhattan plots for pooled GWAS for LDL-C.

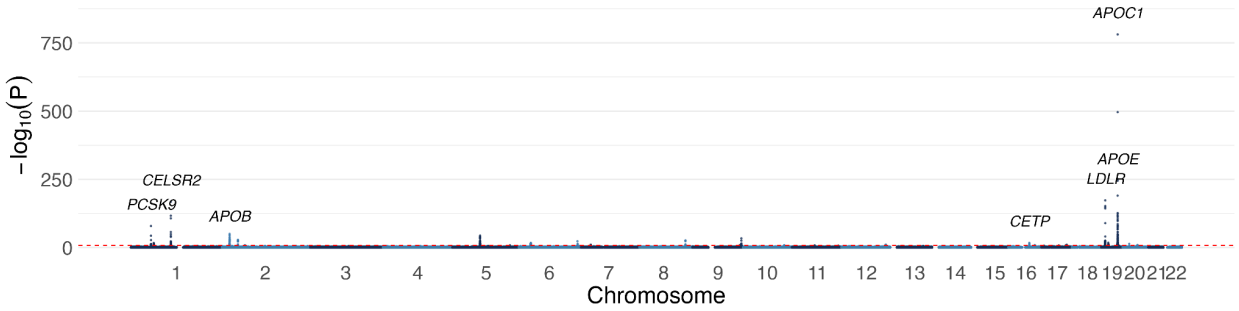



**Figure 3C.** Manhattan plots for pooled GWAS for TG.

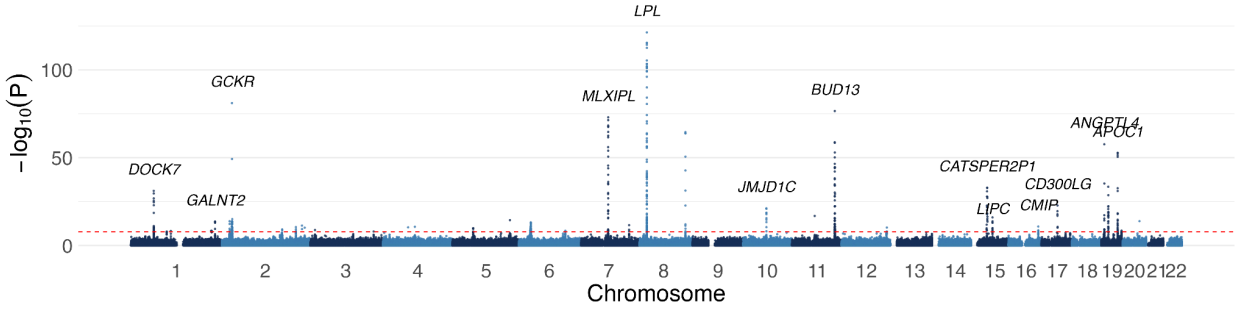
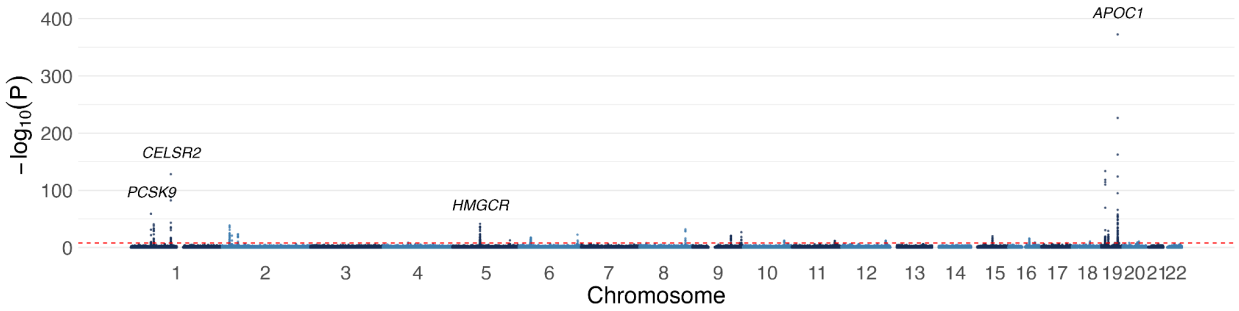


**Figure 3D.** Manhattan plots for pooled GWAS for TC.




## **Figure 4.** Regional association plots showing sex-specific associations.

Regional association plots illustrate sex-specific associations. Variants are shown with $-\log_{10}(P)$ values plotted against genomic position, and are colored according to linkage disequilibrium ($r^2$) with the lead variant. Gene models and transcript orientations are displayed below each plot to contextualize the association signals using the locus zoom package.

**Figure 4A.** Regional association plots for *CELSR2* showing male-specific association with HDL-C.

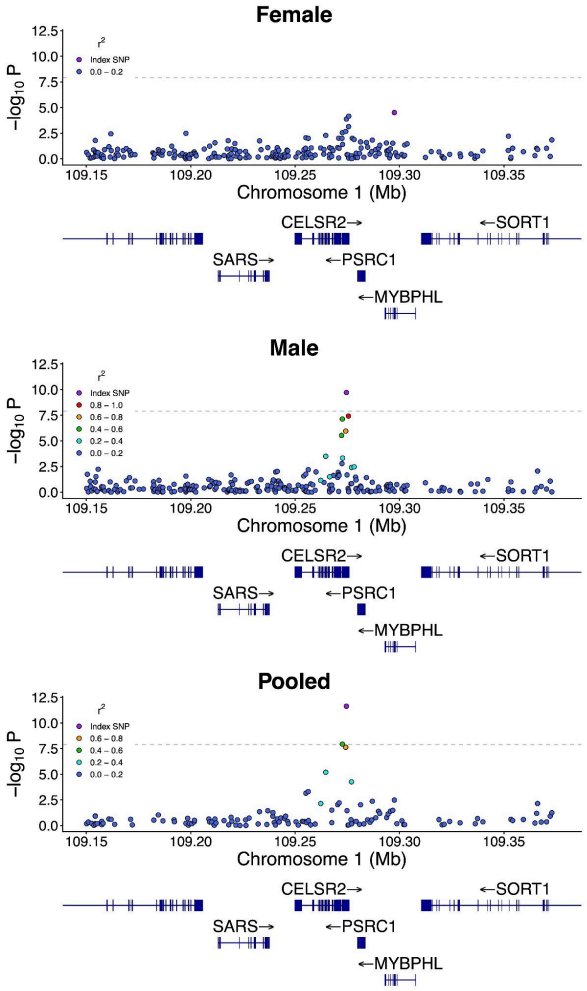

Female
Male
Pooled
$r^2$
Index SNP
0.8 – 1.0
0.6 – 0.8
0.4 – 0.6
0.2 – 0.4
0.0 – 0.2
$-\log_{10}$ P
12.5
10.0
7.5
5.0
2.5
0.0
109.15
109.20
109.25
109.30
109.35
Chromosome 1 (Mb)
CELSR2→
←SORT1
SARS→
←PSRC1
←MYBPHL

**Figure 4B.** Regional association plots for *CMIP* showing female-specific association with TG.

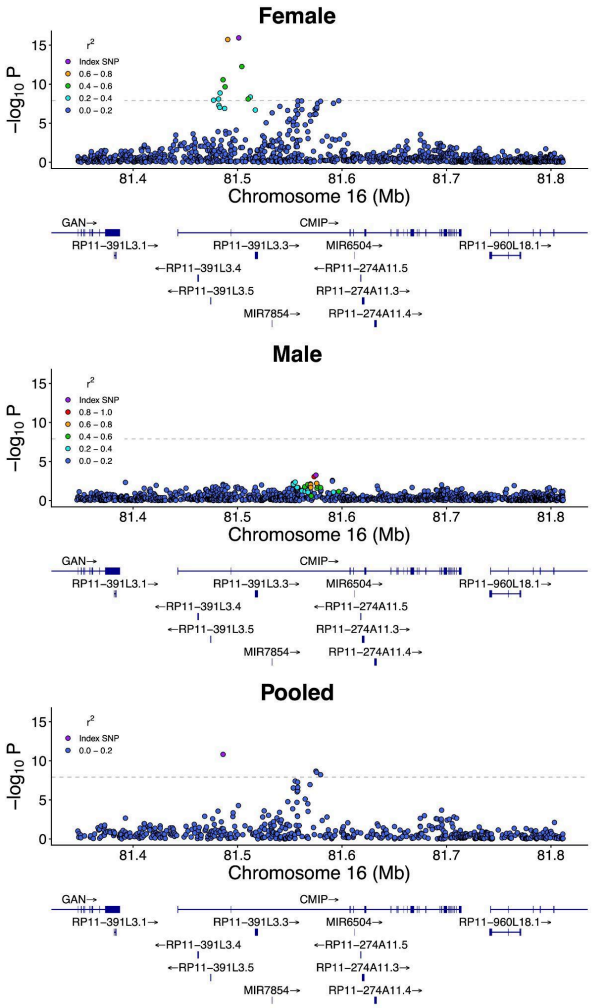
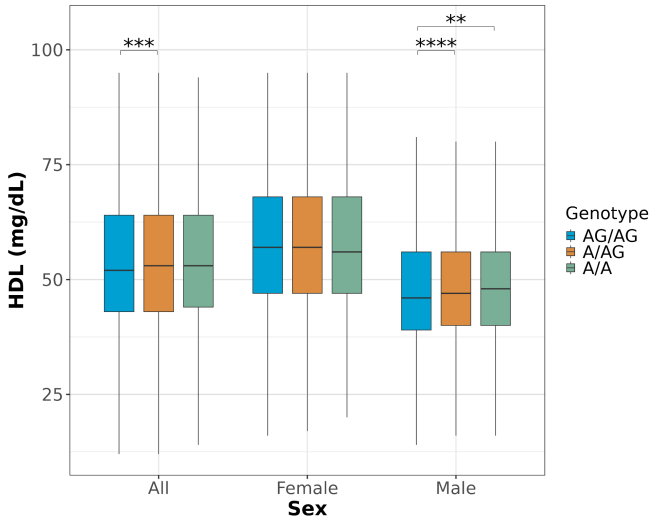


**Figure 5. rs7528419 at chr1:109274570:A:G shows male-specific significance for HDL-C (CELSR2).**

Boxplots show lipid profile levels stratified by genotype and sex for variants exhibiting sex-specific associations.Values outside of 1.5 inter-quartile range from quartiles were not shown. Significance threshold are noted in asterisks using the following convention:

****: Tuckey adjusted p-value < 0.0001
***: Tuckey adjusted p-value < 0.001
**: Tuckey adjusted p-value < 0.01
*: Tuckey adjusted p-value < 0.05

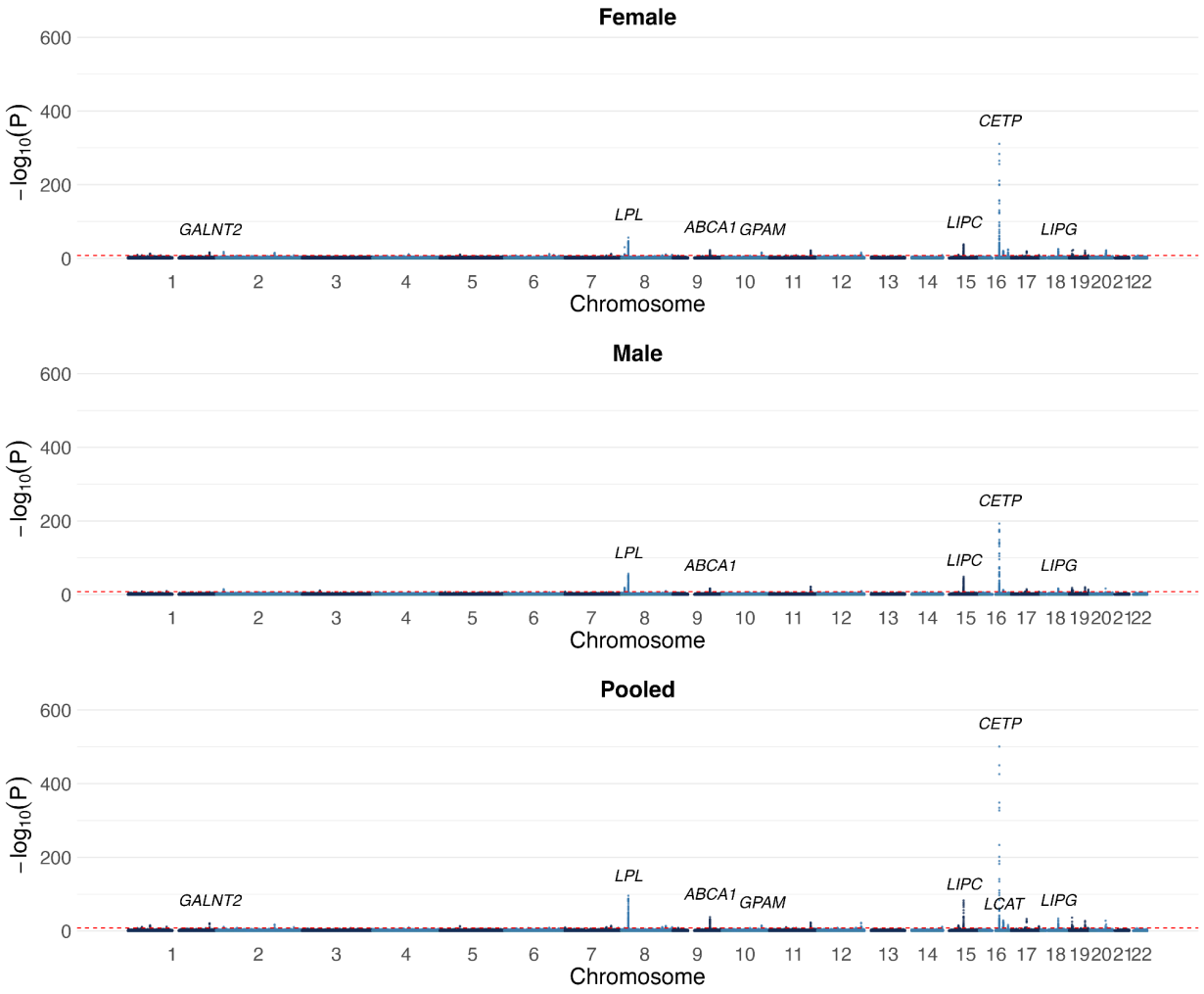



# Supplementary Figures

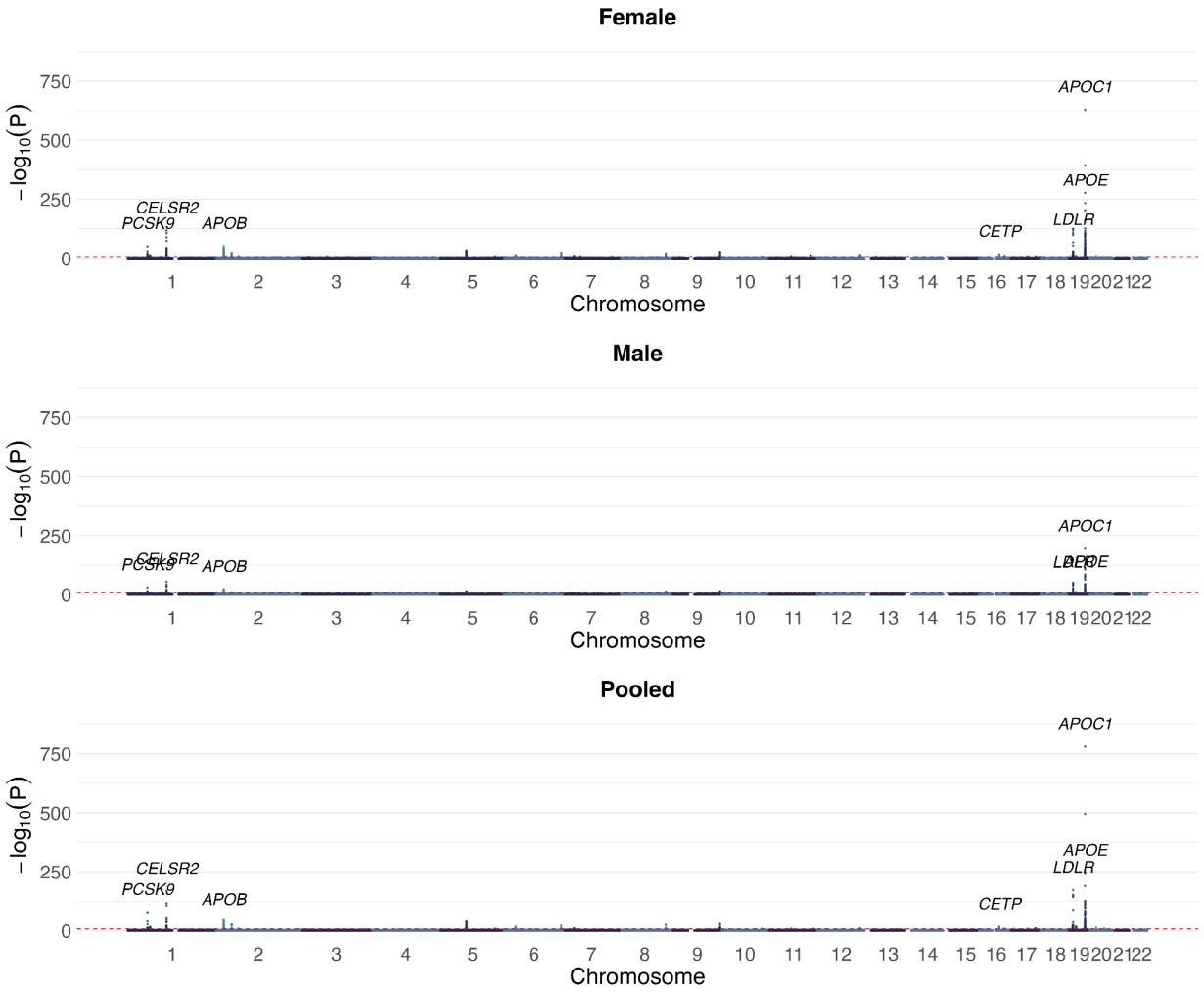
## **Supplementary Figure 1.** Manhattan plots for sex-specific and pooled GWAS

Manhattan plots display female-only, male-only, and pooled genome-wide association results for HDL-C, LDL-C, TG, and TC in the pooled cohort. Each point represents a genetic variant plotted by chromosomal position and $-\log_{10}(P)$. The horizontal dashed line denotes the genome-wide significance threshold after bonferroni correction.

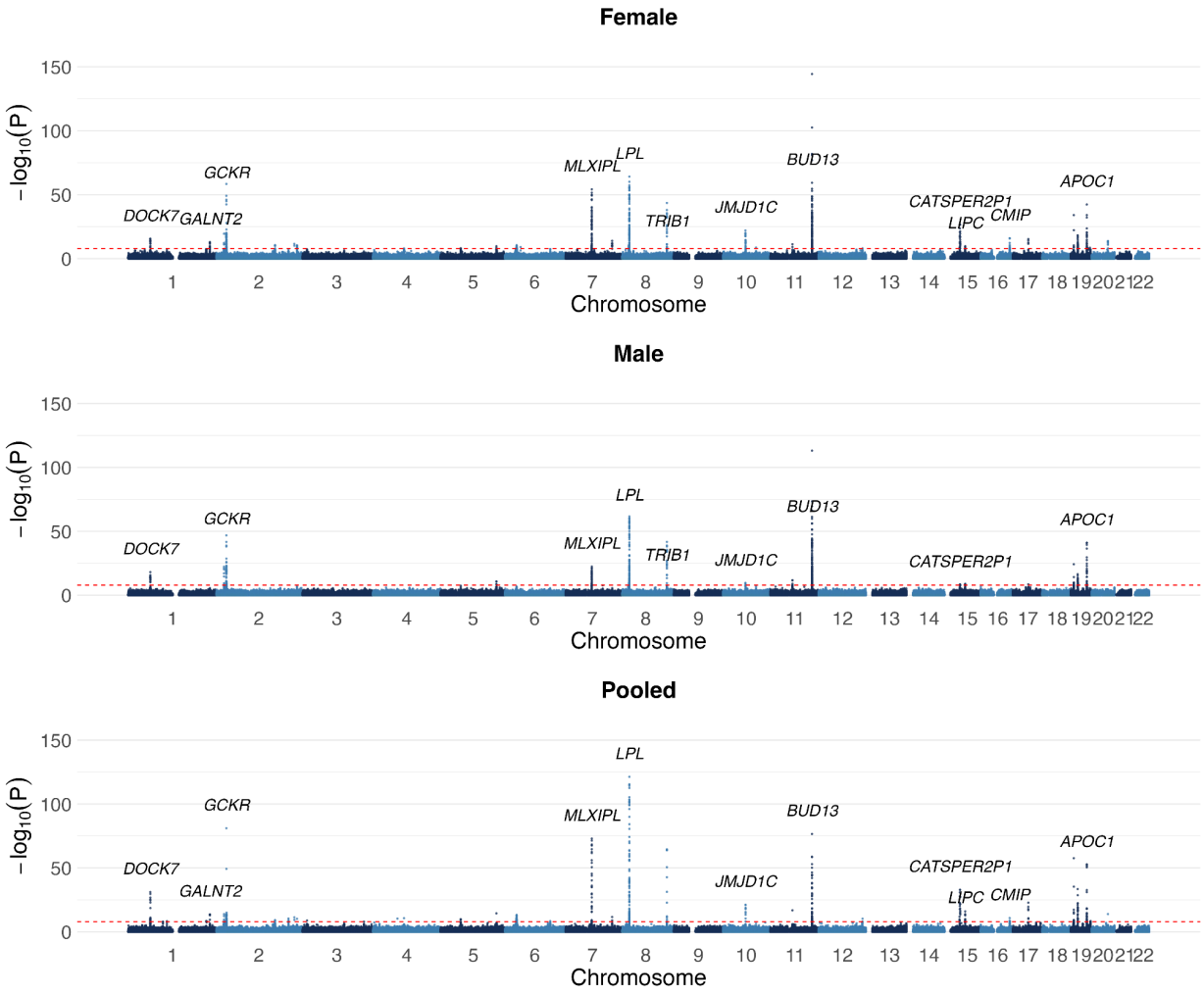
**Supplementary Figure 1A.** Manhattan plots for sex-specific and pooled GWAS for HDL-C



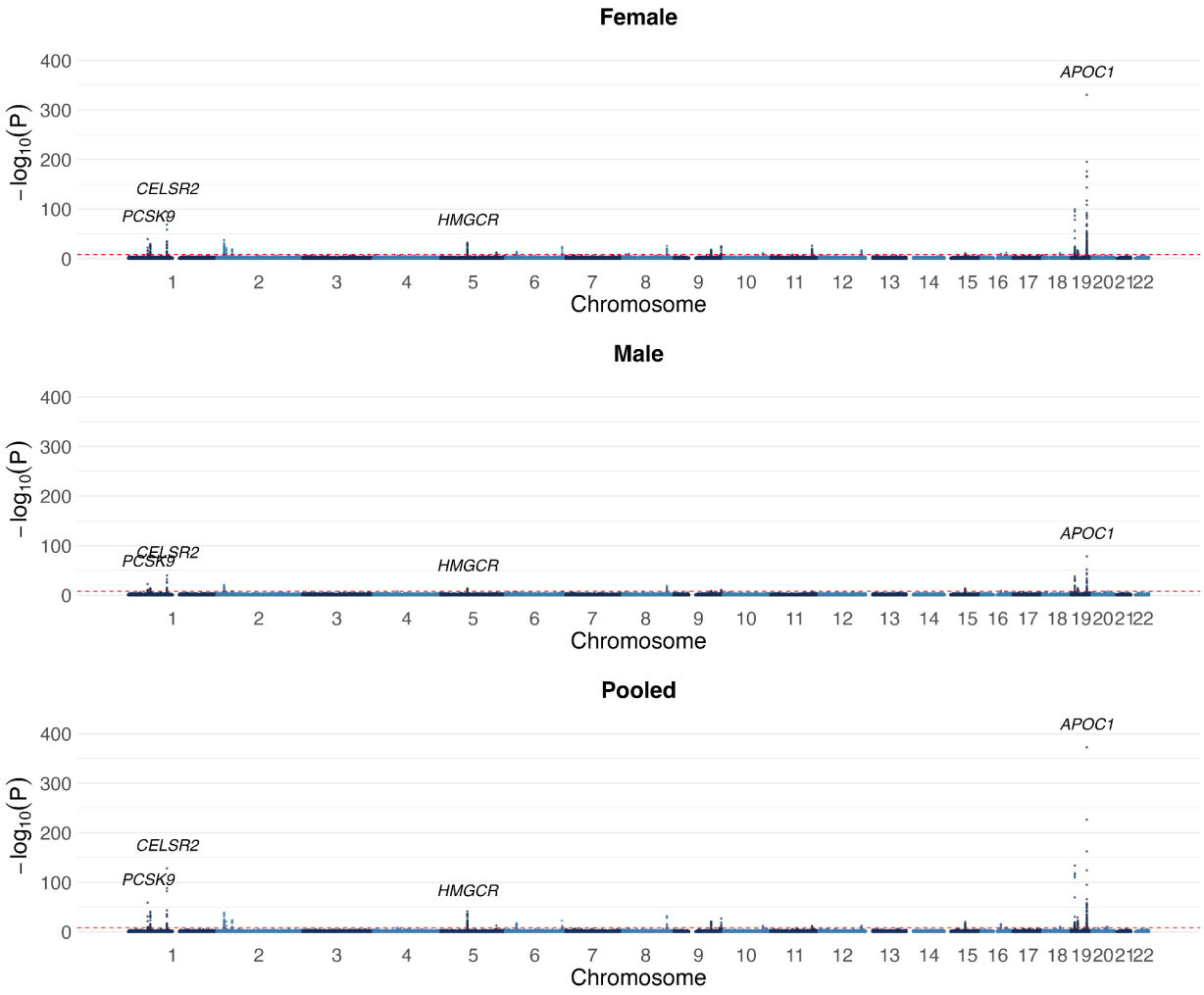
**Supplementary Figure 1B.** Manhattan plots for sex-specific and pooled GWAS for LDL-C



**Supplementary Figure 1C.** Manhattan plots for sex-specific and pooled GWAS for TG



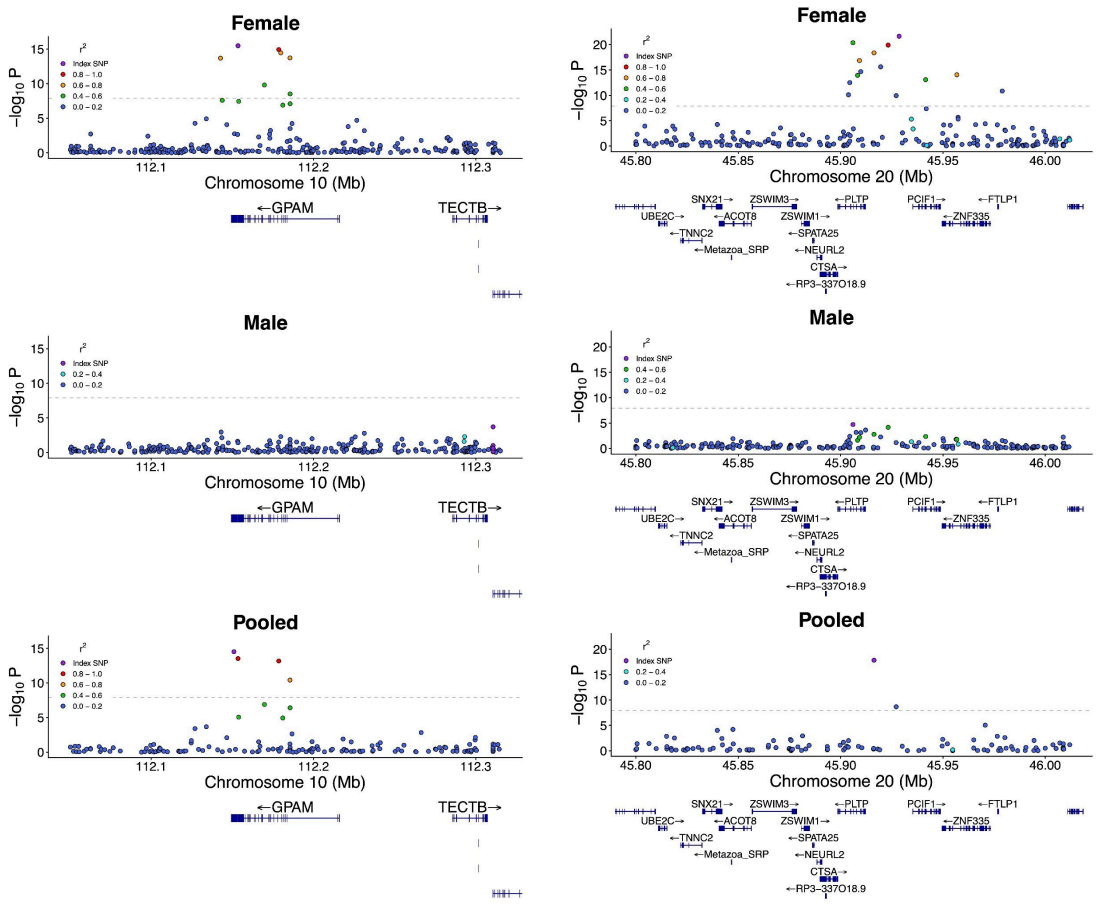
**Supplementary Figure 1D.** Manhattan plots for sex-specific and pooled GWAS for TC



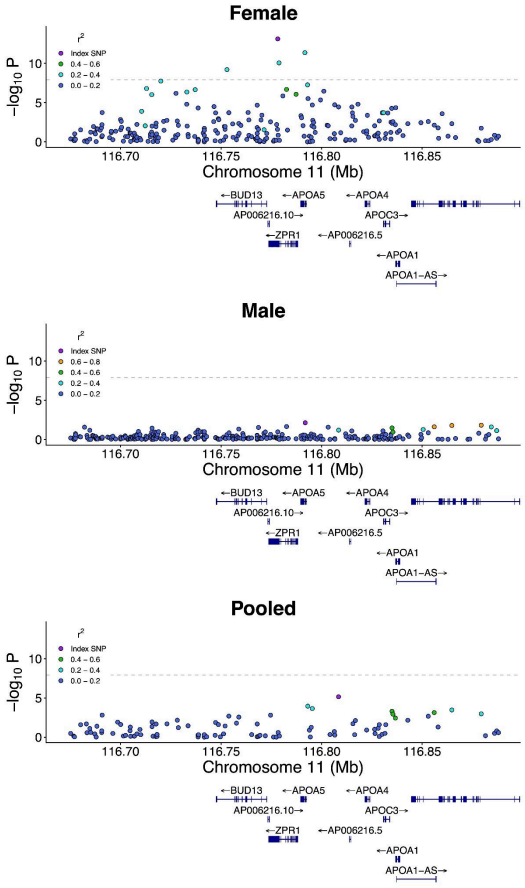
**Supplementary Figure 2.** Regional Association plots showing sex-specific associations

Regional association plots illustrate sex-specific associations. Variants are shown with $-\log_{10}(P)$ values plotted against genomic position, and are colored according to linkage disequilibrium ($r^2$) with the lead variant. Gene models and transcript orientations are displayed below each plot to contextualize the association signals using the locus zoom package.

**Supplementary Figure 2A.** Regional association plots for *GPAM* (left) and PLTP (right) showing female-specific associations with HDL-C



**Supplementary Figure 2B.** Regional association plots for *ZPR1* showing female-specific association with LDL-C.

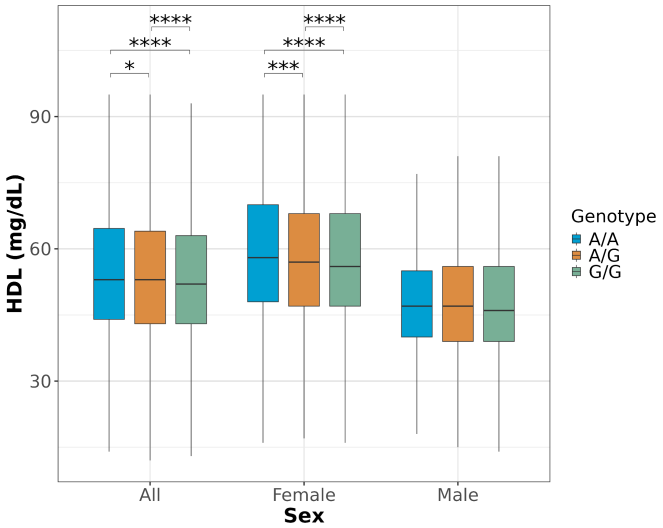



**Supplementary Figure 3.** Lipid panel values for individuals with different genotypes at sex-specific loci

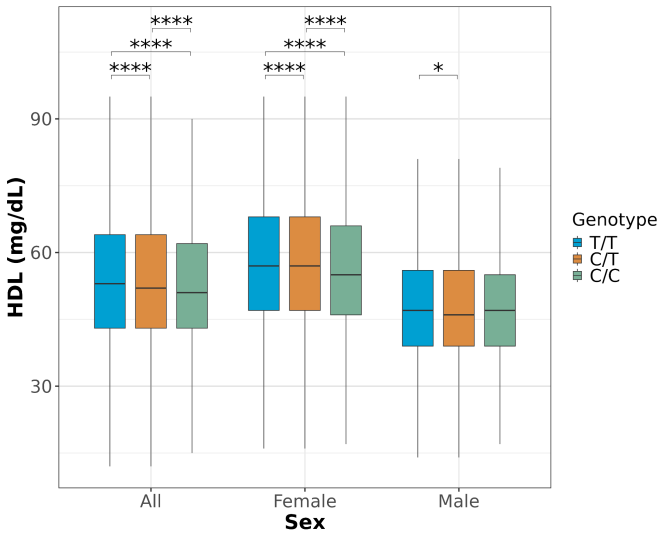
Boxplots show lipid profile levels stratified by genotype and sex for variants exhibiting sex-specific associations.Values outside of 1.5 inter-quartile range from quartiles were not shown. Significance threshold are noted in asterisks using the following convention:

****: Tuckey adjusted p-value < 0.0001
***: Tuckey adjusted p-value < 0.001
**: Tuckey adjusted p-value < 0.01
*: Tuckey adjusted p-value < 0.05

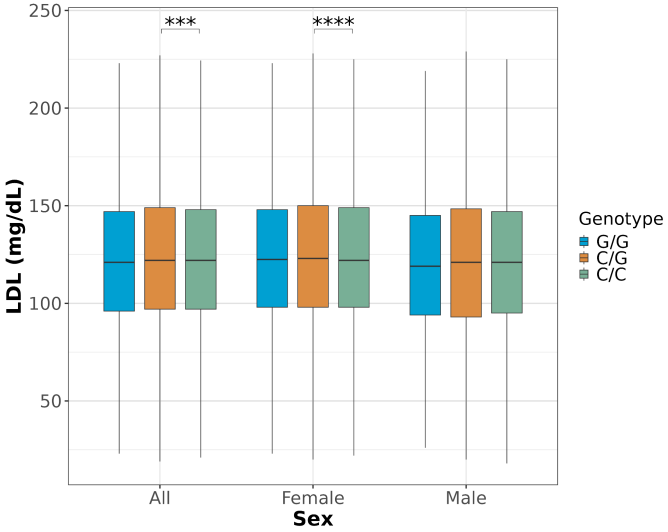
**Supplementary Figure 3A.** rs1129555 at chr10:112150963:A:G shows female-specific significance for HDL-C (*GPAM*).



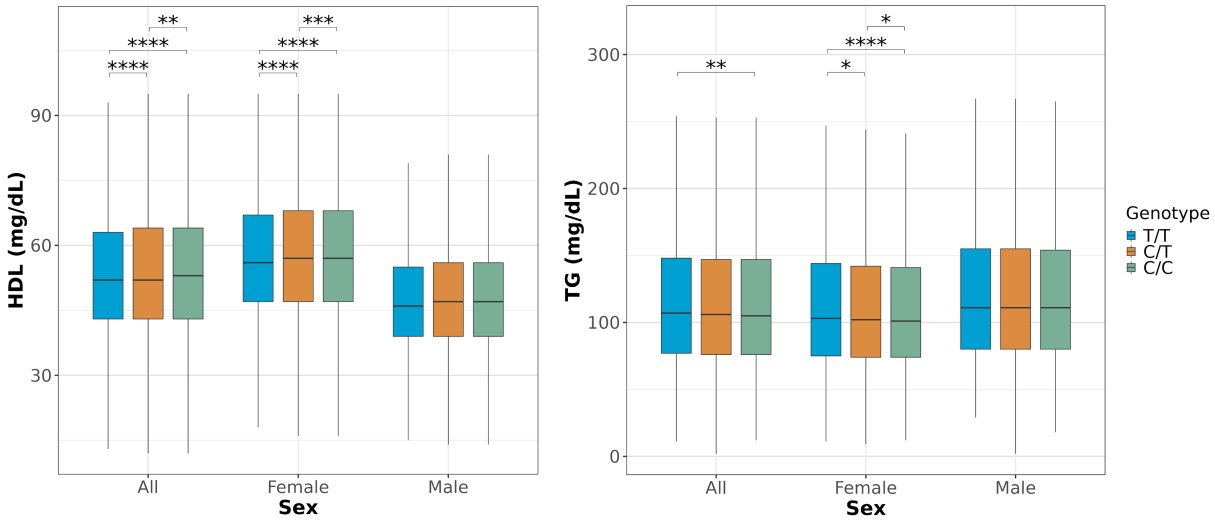

**Supplementary Figure 3B.** rs58847685 at chr20:45916308:T:C shows female-specific significance for HDL-C (*PLTP*).



**Supplementary Figure 3C.** rs964184 at chr11:116778201:G:C shows female-specific significance for HDL-C (*ZPR1*).

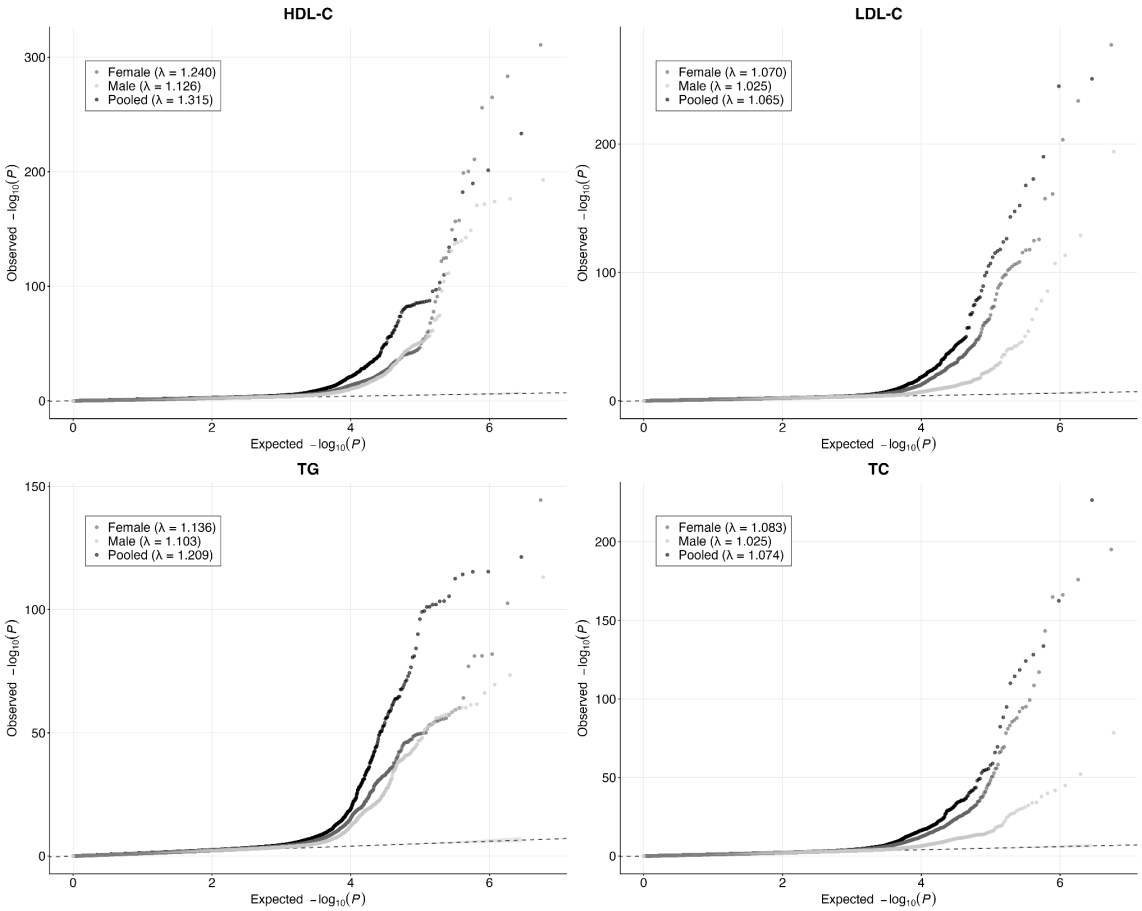



**Supplementary Figure 3D.** rs2925979 at chr16:81501185:A:G shows female-specific significance for HDL-C and TG (*CMIP*).




## **Supplementary Figure 4. Quantile-quantile (**QQ) plots for pooled and sex-specific GWAS for HDL-C, LDL-C, TG, and TC

For each trait, observed –log10(P) values from the pooled (all participants) GWAS and from sex-specific (female-only, male-only, and pooled) GWAS are plotted against the expected –log10(P) values under the null distribution. Lambda genomic inflation factors (λGC) for each analysis are indicated in the panels.


HDL-C
Female (λ = 1.240)
Male (λ = 1.126)
Pooled (λ = 1.315)
LDL-C
Female (λ = 1.070)
Male (λ = 1.025)
Pooled (λ = 1.065)
TG
Female (λ = 1.136)
Male (λ = 1.103)
Pooled (λ = 1.209)
TC
Female (λ = 1.083)
Male (λ = 1.025)
Pooled (λ = 1.074)
Observed $-\log_{10}(P)$
Expected $-\log_{10}(P)$

## Supplementary Tables

**Supplementary Table S1.** SNP level summary table of statistically significant variants of pooled GWAS run. Beta coefficients were estimated using the reference allele as the effect allele.

**Supplementary Table S2.** Gene level summary table of statistically significant genes of pooled GWAS run.

**Supplementary Table S3.** SNP level summary table of statistically significant variants of female-specific GWAS run. Beta coefficients were estimated using the reference allele as the effect allele.

**Supplementary Table S4.** SNP level summary table of statistically significant variants of male-specific GWAS run. Beta coefficients were estimated using the reference allele as the effect allele.

**Supplementary Table S5.** Gene level summary table of statistically significant variants of female-specific GWAS run.

**Supplementary Table S6.** Gene level summary table of statistically significant variants of male-specific GWAS run.