Chromosome Xq23 is associated with lower atherogenic lipid concentrations and favorable cardiometabolic indices

Autosomal genetic analyses of blood lipids have yielded key insights for coronary heart disease (CHD). However, X chromosome genetic variation is understudied for blood lipids in large sample sizes. We now analyze genetic and blood lipid data in a high-coverage whole X chromosome sequencing study of 65,322 multi-ancestry participants and perform replication among 456,893 European participants. Common alleles on chromosome Xq23 are strongly associated with reduced total cholesterol, LDL cholesterol, and triglycerides (min $P = 8.5 \times 10^{-72}$), with similar effects for males and females. Chromosome Xq23 lipid-lowering alleles are associated with reduced odds for CHD among 42,545 cases and 591,247 controls ($P = 1.7 \times 10^{-4}$), and reduced odds for diabetes mellitus type 2 among 54,095 cases and 573,885 controls ($P = 1.4 \times 10^{-5}$). Although we observe an association with increased BMI, waist-to-hip ratio adjusted for BMI is reduced, bioimpedance analyses indicate increased gluteofemoral fat, and abdominal MRI analyses indicate reduced visceral adiposity. Co-localization analyses strongly correlate increased $CHRDL1$ gene expression, particularly in adipose tissue, with reduced concentrations of blood lipids.
Mendelian, population, and functional genetic analyses of blood lipids (total cholesterol, low-density lipoprotein cholesterol [LDL-C], high-density lipoprotein cholesterol [HDL-C], and triglycerides) have yielded important fundamental insights regarding the root causes of coronary heart disease (CHD)\textsuperscript{12}. For example, rare and common autosomal genomic variation influencing LDL-C, correspondingly influence CHD risk\textsuperscript{–6}. Such observations buttress clinical recommendations and bolster efforts to discover and validate lipid-related drug targets for CHD risk reduction\textsuperscript{7–9}.

Although the X chromosome comprises 5% of the genome, it has only been studied in a few genome-wide association analyses for blood lipids and coronary disease\textsuperscript{10–13}. Major reasons for exclusion include incomplete coverage on genotyping arrays, potential discrepancies in genotyping quality on arrays due to haplotype insufficiency in men, imputation and analytic challenges, and somatic X inactivation across tissues in women. Deep-coverage whole-genome sequencing (WGS) and analysis of the X chromosome now offers the promise for uniform coverage and high-fidelity genotyping for both sexes\textsuperscript{14}.

While differences in lipid levels and CHD risk by sex are well established\textsuperscript{15,16}, X chromosome dosage is also linked to lipid differences. Monosomy X (45X, Turner syndrome) is linked to dyslipidemia and premature CHD\textsuperscript{17–19}. While obesity and gonadal deficiency was long believed to be the primary contributor to these phenotypes, women with Turner syndrome have higher total cholesterol, LDL-C, and triglyceride concentrations than age- and body composition-matched 46XX women with premature ovarian failure\textsuperscript{15,20}. Men with an additional X chromosome (47XXX, Klinefelter syndrome) also suffer from infertility with higher rates of obesity, dyslipidemia, and CHD\textsuperscript{21,22}. Furthermore, adult gonadectomized mice with XY, XX, and XXY chromosomes, regardless of gonadal sex, demonstrate dose-dependent changes in lipid levels\textsuperscript{23}. Such observations suggest that apparent sexual dimorphism in lipid levels may be explained by the sex chromosomes themselves.

Our study aims to discover X chromosome genomic variation associated with blood lipid levels among 65,322 multi-ancestry individuals with high-coverage whole X chromosome sequencing and available lipids in the NHBLI Trans-Omics for Precision Medicine (TOPMed) program\textsuperscript{24}. Independent serial replication is performed in up to 390,606 and 66,287 individuals with GWAS array and lipids available in the UK Biobank and Nord-Trøndelag Health (HUNT) study, respectively\textsuperscript{25,26}. We further evaluate the phenotypic consequences of lipid-associated variation in the UK Biobank, HUNT, and 176,899 additional participants of FinnGen\textsuperscript{27}. Lastly, we perform colocalization analyses to pinpoint the possible causal gene in association regions. Here, we characterize an X chromosome locus associated with lipids and related cardiometabolic traits and prioritize \textit{CHRDL1} as the causal gene.

**Results**

**Baseline characteristics, blood lipids, and chromosome X genotypes.** TOPMed sequences were aggregated and aligned, and variants were called by the TOPMed Informatics Research Center. A total of 65,367 out of 140,000 individuals in TOPMed freeze 8 with WGS data, including X chromosome sequence data had harmonized lipid levels available (Supplementary Fig. 1). Forty-five individuals with anomalous X chromosome copy number were excluded, leaving 65,322 individuals for analysis. 40,577 (62.1%) individuals were female and mean (standard deviation [SD]) age was 52.4 (14.9) years. Across all 21 included cohorts, 29,513 (45.2%) were white, 16,431 (25.2%) black, 13,432 (20.6%) Hispanic, 4714 (7.2%) Asian, 1182 (1.8%) Samoan, and 50 (0.1%) Native American (Supplementary Table 1; Supplementary Fig. 2). The included studies were largely observational cohorts with some variations in ascertainment schemes as described in the Supplementary Note. Blood lipid distributions were generally similar across cohorts with some differences due to differences in study design and ancestry (Supplementary Table 2 and Supplementary Fig. 3). After adjusting for lipid-lowering medicines within each cohort and ancestry, we generated residuals within each cohort and race group adjusted for age, age\textsuperscript{2}, sex, 11 principal components of ancestry, and cohort-specific covariates. These residuals were inverse rank normalized and multiplied by the standard deviation within each cohort and race group to obtain effects in mg/dl units (see Methods) (Supplementary Fig. 4).

Among 65,322 TOPMed participants with lipid levels and WGS, we identified 19,898,222 total variants on the X chromosome by WGS. Of these variants, 88,008 (0.4%) were nonsynonymous variants and 4632 (0.02%) were rare (MAF <1%) predicted protein-truncating variants. As expected, participants of African ancestry had the most X chromosome variants (Fig. 1a). Likely due to sample size differences, there were overall more total variants observed in our dataset among white participants compared to other ancestries (Fig. 1b). Within the X chromosome, females had a greater average [SD] number of variants per individual (133,255 [22,455]) than males (90,117 [12,166]), as expected (Supplementary Table 3). Generally, most of the variation observed across individuals was uncommon (i.e., 98.8% of variants had MAF <5%) (Supplementary Table 4).

**X chromosome single-variant association with lipid levels.** In single-variant discovery analyses in TOPMed, we performed X chromosome-wide association analyses for genetic variants with minor allele count >20 that are not in the pseudautosomal region, yielding 2.2 million analyzed of the 19.8 million detected. To maximize power, all samples (i.e., males and females) were included in the linear mixed model association analyses with SD-adjusted residuals of lipid levels as the outcome, where adjustments included sex (Supplementary Fig. 5).

Across variants assessed, we found 21 variants showing suggestive evidence ($P <1 \times 10^{-6}$) of association with lipids in TOPMed (Supplementary Table 5 and Supplementary Fig. 6). We evaluated these associations for replication, serially, in the UK Biobank ($N=390,606$) (Supplementary Table 6) and HUNT (66,635) (Supplementary Table 7). Three variants showed evidence of replication ($P <0.05/21 = 0.002$) in UK Biobank and in HUNT and additionally met a stringent threshold for statistical significance in the meta-analysis (alpha = 0.05/2.2 M variants/4 traits $= 5.7 \times 10^{-9}$) (Table 1).

The three variants occurred on chrXq23 and were all in at least moderate linkage disequilibrium across all included TOPMed participants (Supplementary Fig. 7 and Supplementary Table 8). They were also in moderate linkage disequilibrium with a previously described nearby variant, rs59854712\textsuperscript{12}, ($r^2 = 0.61$–0.76). All three associated variants in our dataset have similar nonreference allele frequency (0.34–0.43), which was also similar between males and females. We observed similar associations for both males and females within TOPMed except male rs5985504-T carriers had greater decrease in triglycerides compared to female rs5985504-T carriers ($P_{interaction} = 0.001$) (Supplementary Table 9).

The minor alleles for these variants are common in all TOPMed ancestries except for Asian Americans (MAF 0.02) and Samoans (MAF 0.01). Nevertheless, effect estimates were largely of similar magnitude across ancestries in TOPMed for total **ARTICLE**
cholesterol (Supplementary Table 10) with no evidence of heterogeneity ($P_{\text{heterogeneity}} > 0.05$).

The three chrXq23 variants were associated with reduced atherogenic lipoproteins (i.e., total cholesterol, triglycerides, and LDL-C) (Table 1). The rs5942634-T allele is an intergenic variant and is 8 kb downstream from RTL9 (also referred to as RGAG1 in the literature), and was the top variant for total cholesterol, associated with 1.95 mg/dl lower concentration ($P = 2 \times 10^{-4}$). The rs5942648-A allele occurs 81 kb downstream, is intergenic between RTL9 and CHRD1, and was the top variant for LDL-C, associated with 1.53 mg/dl lower concentration ($P = 1 \times 10^{-12}$). The rs5985504-T allele resides 60 kb further downstream and is 68 kb from CHRD1 and was the top variant for log(triglycerides) leading to 2% lower triglycerides concentration ($P = 4 \times 10^{-11}$). Overall, the associated variants reside within a ~0.22 Mb linkage disequilibrium block spanning RTL9 and CHRD1 (Supplementary Fig. 7). Within this block, variants with predicted active adult liver enhancers are in proximity to both the RTL9 and CHRD1 genes (Supplementary Fig. 8). Only two variants reside within both an adult liver enhancer and DNase hypersensitivity site—rs2883091 in an intron of RTL9, and rs2143760 residing 4 kb from CHRD1 but 214 kb from RTL9. These variants are in at least moderate linkage disequilibrium ($r^2 > 0.60$) with the top associated variants in the locus. Virtual 4C data additionally demonstrate a contact between the rs5985504 site and upstream of CHRD1 (Supplementary Fig. 9).

To determine whether our signal was independent of previously reported variants in the region, we performed conditional analysis for the associated between total cholesterol and rs5942634 for the individual previously reported variants in the region (results not shown). This indicates that rs5942634 is only partially explained by the three reported variants.

Phenome-wide association Of Chrxq23 variants. Given prior genetic associations of LDL-C-lowering and triglyceride-lowering autosomal variants with lower risk for CHD, we hypothesized that sex chromosome variants lowering LDL-C or triglycerides would also lower risk for CHD. In HUNT, UK Biobank, and FinnGen (Supplementary Table 12), we observed that the top lipid-lowering alleles at this locus showed a reduced risk for CHD (Fig. 2). We found a $0.98$ (95% CI 0.96, 0.99; $P = 1.7 \times 10^{-4}$) odds of CHD for each rs5942634-T allele, the lead cholesterol-lowering variant (alpha = 0.05 for the single haplotype assessment), and a correlation between the effect sizes of variants on total cholesterol in the chrXq23 locus and the effect sizes of these variants on CAD ($r = 0.25$), T2D ($r = 0.33$), and BMI ($r = -0.34$) (Supplementary Fig. 10).

To explore the range of phenotypes associated with the chrXq23 locus, we evaluated the associations of each of these three variants with 80 manually curated diverse clinical traits and conditions in the UK Biobank (Supplementary Table 13). Given the high degree of correlation among these variants, phenome-wide association results were similar (Supplementary Tables 14-16). As expected, the strongest associations were for reduced odds of hypercholesterolemia. Associations reaching a $P < 6.3 \times 10^{-4}$ ($P < 0.05/80$ traits) included reduced odds for diabetes mellitus type 2 (T2D), hypertension, and glaucoma, but increased odds for ever smoking as well as increased body-mass index (BMI) and body fat percentage. Notably, we observed lower odds of T2D for rs5942648 (OR = 0.97; 95% CI 0.96, 0.99; $P = 1.4 \times 10^{-5}$) (Fig. 2).

We additionally explored the association between each of these three variants with lipoprotein subspecies identified through nuclear magnetic resonance spectroscopy (NMR) within the Framingham Heart Study and Multi-Ethnic Study of Atherosclerosis cohorts (up to 6356 individuals). While we did not find any associations that passed a Bonferroni-corrected significance threshold ($0.05/(3 \text{ SNPs} \times 16 \text{ lipoprotein subspecies}) = 0.001$; Supplementary Table 17), we found two lipoprotein subspecies...
| rsID Minor allele | TOPMed (N = 390,606) UK Biobank non-Whites | HUNT (N = 66,287) |
|-----------------|------------------------------------------|---------------------|
| rs5942634 T     | 34.4%                                    | 80%                 |
| rs5985504 T     | 43.3%                                    | 95%                 |
| rs5942648 T     | 43.8%                                    | 95%                 |

**Table 1: Discovery and replication of chromosome Xq23 variants associated with lipid levels in TOPMed, UK Biobank, and HUNT.**

Gene expression analyses at chromosome Xq23. We leveraged the GTEx eQTL data to better understand the gene or genes in the region that are influencing atherogenic lipid levels. Our most significant SNP, rs5942634, was associated with reduced expression of CHRD1 in skeletal muscle (beta = -0.17, P = 1.2 x 10^-11), subcutaneous adipose (beta = -0.16, P = 8.6 x 10^-5), visceral adipose (beta = -0.17, P = 4.3 x 10^-5), and liver (beta = -0.25, P = 5.9 x 10^-5). Additionally, rs5942634 was associated with increased expression of RTL9 in skeletal muscle (beta = 0.18, P = 2.7 x 10^-3; Supplementary Table 20).

Interrogating eQTL data for a single variant may lead to biased interpretations for causal gene inference. Therefore, we colocalized eQTL results for 8 genes (i.e., AC51A, TMEM164, AMMECR1, RTL9, CHRD1, PAK3, CAPN6, DCX) within the ChrXq23 region across prespecified lipid-related tissues (i.e., subcutaneous adipose, terminal ileum, visceral omentum adipose, whole blood, liver, and skeletal muscle) to relate aggregate blood lipid-association data with gene expression data. We observe that increased gene expression of CHRD1 shows consistent colocalization with decreased cholesterol associated with suggestive evidence (P < 0.05), including greater concentration of medium HDL particles (but no effect on small or large HDL particles) and greater LDL size. We assessed for evidence of replication for indices related to LDL size (alpha 0.05) since the chrX variants associated with LDL-C. Among 6443 participants of the Atherosclerosis Risk in Communities cohort, we concordantly observed a -0.034 SD (P = 0.022) lower concentration of small dense LDL for rs5942648-A. Among 365,365 participants of the UK Biobank, when using LDL-C/apolipoprotein B ratio as a proxy for LDL particle size, we observed a nominal increase in LDL size even with adjusting for both LDL-C and apolipoprotein B (Beta = 1.1 x 10^-5, P = 0.048).

To better characterize effects on adiposity given the aforementioned clinical phenotype associations, we evaluated the association between rs5942634-T and body composition measurements in the UK Biobank. Although rs5942634-T was associated with increased BMI, it was associated with slightly reduced waist-to-hip ratio adjusted for BMI (Beta = -6.3 x 10^-4, SE = 1.1 x 10^-4, P = 1.3 x 10^-8). rs5942634-T is associated with both increased truncal fat mass (Beta = 63 g, SE = 10 g, P = 4.0 x 10^-10) as well as increased total peripheral fat mass, with increase of 21 g (P = 3.6 x 10^-12) of the right leg, 20 g (P = 3.4 x 10^-12) of the left leg, 7 g (P = 4.1 x 10^-7) of the right arm, and 8 g (P = 1.7 x 10^-9) of the left arm (Supplementary Table 18). Additionally, among 4750 unrelated UK Biobank participants with abdominal MRI measures available, rs5942634-T was associated with log-transformed inverse rank standardized increased abdominal subcutaneous adipose tissue (Beta = +0.43, SE = 0.15, P = 5.9 x 10^-5) but decreased visceral adipose fat (Beta = -1.12, SE = 0.14, P = 1.1 x 10^-15) to a greater degree. Given nine adiposity traits assessed, Bonferroni-corrected significance was assigned at 0.05/9 = 5.6 x 10^-3.

Rare pathogenic variants in CHRD1 were previously linked to X-linked recessive megalocornea, a condition characterized by enlarged corneal diameters with associated complications, including reduced visual acuity. Given these prior observations, we asked whether common variants associated with cholesterol at the CHRD1 locus were associated with differences in visual acuity. Among 112,842 UK Biobank participants (46.5% women; median age at assessment 58.5 years), we observed no association of lipid-associated chrXq23 alleles with altered visual acuity (P > 0.05; Supplementary Table 19). Given our sample size of 112,842 and SNP frequency of 34.4%, we had >99% power to detect effects >1/10th of a standard deviation unit of visual acuity at an alpha of 0.05.
across tissues, indicating that CHRDL1 is the likely causal gene in the region (Fig. 3, Supplementary Fig. 11, and Supplementary Table 21).

Rare chromosome X variant association analyses. We performed SKAT gene-based tests of rare variants (MAF < 1%) across the X chromosome implemented by the GENESIS package within the TOPMed samples. We tested a maximum of 746 genes with more than 1 rare variant and at least 10 individuals carrying a minor allele with each lipid trait. No genes reached a Bonferroni-corrected significance threshold (0.05/746 = 6.7 × 10−5; Supplementary Table 22).

Discussion

Using deep-coverage next-generation sequencing of the X chromosome in the NHLBI TOPMed program, we identified a locus that was previously associated with lipids in primarily European ancestry datasets, which we now extend to diverse ancestries along with strong replication in two independent studies. In addition to replicating the associations of chrXq23 lipid-lowering alleles with reduced odds for CHD, we also observe associations with reduced T2D odds and favorable adiposity indices. These observations allow us to draw several conclusions about X chromosome genetic variation with blood lipid levels, as well as related cardiometabolic effects.

First, bioinformatic analyses implicate CHRDL1 as a candidate causal gene for the association of chrXq23 variants with lipids. Based on genomic proximity of the strongest signal, RTL9 was assigned as the likely causal gene in prior work. However, colocalization analyses strongly prioritize increased CHRDL1 gene expression in lipid-related, particularly adipose, tissues with reduced lipoprotein measures over other genes in the region. CHRDL1 is not a previously known Mendelian lipid gene. In our study, disruptive rare coding variants in CHRDL1 were not significantly associated with lipids, nor was any gene in the region.

Second, despite observing an association with increased BMI, chrXq23 lipid-associated alleles may lead to favorable effects on adiposity. We observed that chrXq23 lipid-lowering alleles were associated with increased gluteofemoral adipose tissue. Autosomal alleles similarly linked to expansion of gluteofemoral adipose tissue are associated with favorable risk for CHD and T2D. Autosomal alleles associated with body fat distribution are also associated with various functional adipose measures, including morphology, lipolysis, and lipogenesis. Recent gene expression analyses of human adipose-derived stromal cells showed persistent upregulation of CHRDL1 and other genes in the region. CHRDL1 is believed to influence adipogenic differentiation in human isolated preadipocytes. Comparative gene expression analyses suggest relatively greater CHRDL1 expression in subcutaneous versus visceral fat. In our analyses, the chrXq23 lipid-lowering alleles were associated with an increase in abdominal subcutaneous adipose tissue but a decrease in visceral adipose tissue. A proteomic discovery analysis showed that increasing circulating CHRDL1 concentrations were associated with increased birth weight but decreased triglycerides and homeostatic model assessment of insulin resistance.

Third, chrXq23 lipid-lowering alleles have favorable cardiometabolic effects that appear to reduce risk for CHD and T2D. Our results for CHD at chrXq23 are consistent with prior work at this locus, and extend to prior observational epidemiology, genetic, functional, and clinical trial evidence implying a causal relationship between reduced LDL-C and reduced CHD risk. In aggregate, autosomal LDL-C-reducing alleles are associated with increased T2D risk but the effects are inconsistent across individual variants. In meta-analyses of randomized controlled trials, statins are associated with a modestly increased risk of
incident T2D38,39. The effects of triglyceride-lowering alleles and T2D risk are generally inconsistent40–42. Common triglyceride-lowering variants at LPL and ANGPTL4. E40K in the lipoprotein lipase pathway are associated with reduced triglyceride concentrations, CHD odds, and T2D risk40,43,44. Rare loss-of-function variants in ANGPTL3, also implicated in the lipoprotein lipase pathway, are associated with reduced LDL-C and triglyceride concentrations as well as reduced CHD odds but favorable effects on T2D risk have not yet been observed40,45,46. However, we uniquely describe a genetic locus associated with reduced LDL-C concentrations, reduced triglyceride concentrations, lower CHD risk, and lower T2D risk. Similarly, we observed that lipid-lowering chrXq23 alleles were independently associated with increased LDL-C/apoB ratio, which has been independently associated with reduced CHD and T2D risk in observational epidemiologic studies47–50. These data imply that therapeutic modulation of the causal pathway may lead diverse favorable cardiometabolic indices. Whether implicated variants influence the lipoprotein lipase pathway or represent a novel lipid-related pathway for combined CHD and T2D should be addressed by future research.

Fourth, common lipid-associated variants linked to increased CHRD1 expression are not associated with visual acuity measures. Ventropin, the product of CHRD1, was first described as a bone morphogenetic protein 4 inhibitor and a regulator of retinal development51. Pathogenic disruptive variants in CHRD1 are implied in X-linked megalocornea52. However, common lipid-associated variants linked to increased CHRD1 expression in the present study do not associate with measures of visual acuity in the UK Biobank. These data imply that therapeutic modulation to recapitulate the protective effects associated with chrXq23 lipid-lowering alleles is not anticipated to lead to on-target adverse visual acuity effects.

Important limitations should be considered in the interpretation of our findings. First, our genetic association analyses of the X chromosome do not account for random X inactivation. Accounting for random X inactivation is expected to modestly improve power and thus our approach biases our findings toward the null53,54. We found that there was slightly higher variance of total cholesterol in heterozygous females (sd = 44.1 mg/dl) compared to homozygous females (homozygous ref sd = 43.0, homozygous alt sd = 43.9) of rs5942634 using the Levene’s Test for Homogeneity of variance (P = 0.002), indicating this locus may be subject to random X inactivation. Second, while our in silico analyses and prior literature strongly implicate CHRD1 as the causal gene, additional analyses including perturbational experiments in model systems are necessary to confirm our hypotheses. Furthermore, whether additional cis-acting or trans-acting gene expression for other genes are additionally relevant for the observed effects on blood lipids are currently unknown. Third, chrXq23 lipid-lowering alleles were associated with both increased truncal and gluteofemoral adiposity indices, and whether the former associations result in adverse clinical consequences is not known. Nevertheless, phenotype-wide association analyses did not reveal concerning clinical phenotype associations related to modest increases in BMI. Notably, chrXq23 lipid-lowering alleles were associated with decreased visceral adipose tissue.

In conclusion, we observe a consistent association of chrXq23 alleles with reduced total cholesterol, LDL-C, triglycerides, CHD, and T2D. Despite an increase in BMI, these alleles were favorably associated with increased high density lipoprotein and reduced abdominal subcutaneous adiposity, decreased visceral adiposity, and increased LDL-C/apoB ratio. Colocalization analyses strongly implicate increased CHRD1 expression in adipose tissues with these favorable cardiometabolic indices, pointing to CHRD1 as the leading candidate gene in the region.

### Methods

#### Study participants

For discovery, 65,322 individuals from 21 studies in the freeze release of the NHLBI TOPMed Program with WGS passing central quality control by the TOPMed Informatics Research Core and blood lipid data available were included for analysis (Supplementary Fig. 1). The included studies are Atherosclerosis Risk in Communities study (ARIC, 7991), Old Order Amish (Amish, 1083), Mt Sinai (BSU, 8557), Canadian Study of Child Health Outcomes (CSH, 577), Cardiovascular Health Study (CHS, 2773), Diabetes Heart Study (DHS, 365), Framingham Heart Study (FHS, 3990), Genetic Epidemiology Network of Arteriopathy (GENOA, 1,044), Genetics of Lipid-Lowering Drugs and Diet Network (GOLDN, 924), Genetic Epidemiology Network of Salt Sensitivity (GetSalt, 1770), Genomics of Atherosclerosis Risk (GeneSTAR, 1755), Hispanic Community Health Study—Study of Latinos (HCHS/SOL, 7391), Hypertension Genetic Epidemiology Network and Genetic Epidemiology Network of Arteriopathy (HyperGEN, 1853), Jackson Heart Study (JHS, 2846), Multi-Ethnic Study of Atherosclerosis (MESA, 5283), Massachusetts General Hospital Atrial Fibrillation Study (MGH_AF, 683), San Antonio Familial Studies (SAFS, 617), Samoan Aiposity Study (Samoan, 1182), Taiwan Study of Hypertension using Rare Variants (THRV, 1976), and Women’s Health Initiative (WHI, 8305) (Please refer to the Supplementary Note for additional details). Study participants provided consent for each study’s IRB approved protocol. These data were secondarily analyzed through a protocol approved by the Partners Healthcare IRB and Boston University IRB.

For lipid replication, 69,635 participants from the Nord-Trondelag Health (HUNT) study and 390,606 participants of the UK Biobank with genome-wide array data and lipid data were included. The HUNT study is a longitudinal, repetitive population-based health survey conducted in the county of Nord-Trondelag, Norway55. Since 1966, the adult population in the county has been examined three times, through HUNT1 (1984–86), HUNT2 (1995–97), and HUNT3 (2006–08). A fourth survey, HUNT4 (2017–2019), is ongoing. HUNT was approved by the Data Inspectorate and the Regional Ethics Committee for Medical Research in Norway (REK: 2014/144). All HUNT participants have been informed. Approximately 120,000 individuals have participated in HUNT1–HUNT3 with extensive phenotypic measurements and biological samples. The subset of these participants that have been genotyped (~70,000) using Illumina HumanCoreExome v1.0 and 1.1 and imputed with Minimac3 using a combined HRC and HUNT-specific WGS reference panel are included in the current study. The UK Biobank is a large, prospective cohort of ~500,000 British adults aged 40–69 years56,57. Patients provided answers to questions regarding socio-demographic, lifestyle, and health-related factors; additionally, participants provided blood, urine, and saliva for genetic and other future assays. Genotyping was performed on a custom Affymetrix array followed by imputation. Various additional measurements were performed on all recruited participants (e.g., electrocardiography, etc) and some measurements in subsets (e.g., cardiac magnetic resonance imaging, etc). Study participants provided consent for the UK Biobank’s IRB approved protocol. We excluded UK Biobank individuals that met the following criteria: (1) Individuals whose submitted gender is not same as inferred gender; (2) individuals with partial sex chromosome; (3) individuals with second degree or higher degree relatives; and (4) Individuals who withdrawn consent. We analyzed individuals who were British white separately from those that were not considered British White. These data were secondarily analyzed through a protocol approved by the Partners Healthcare IRB.

The effects of lipid-associated X chromosome variants on CHD risk were estimated in 69,635 participants of HUNT, 390,606 British White participants of the UK Biobank, and 176,899 participants of FinnGen. The effects on risk of diabetes mellitus were estimated in 69,635 participants of HUNT, 390,606 participants of UK Biobank, and 171,987 participants of FinnGen. FinnGen (https://www.finnngen.fi/en) is a large biobank study that aims to genotype 500,000 Finns on a FinnGen ThermoFisher Axiom custom array, with the current data freeze comprising 181,820 Finnish individuals7. FinnGen includes prospective epidemiological and disease-based cohorts, and hospital biobank samples. The data were linked by the unique national personal identification numbers to national hospital discharge (available from 1968), death (1953–), and cancer registries (1995–) registries and disease endpoints were defined by harmonizing the International Classification of Diseases (ICD) revisions 8, 9, and 10, cancer-specific ICD-O-3 and ATC-codes. The FinnGen project is approved by Finnish Institute for Health and Welfare (THL).

#### Sequencing, genotyping, and quality control of TOPMed

Whole-genome sequencing of at least 30x was performed across six sequencing centers using PCR-free library preparation kits for TOPMed samples48. In most cases, all samples for a given study were sequenced at the same center. Samples were excluded if estimated coverage by verifyBamId was <3x [Jun, 2012 #278] or <50% of the genome attained >10x coverage. The reads were centrally realigned at the TOPMed Informatics Research Center (IRC) to human genome build GRCh38 at each center using BWA-MEM49,50. Joint variant discovery was subsequently performed with the ‘GotCloud’ pipeline by the IRC51. The variant calling software tools are under software development; updated versions can be accessed at https://github.com/hiyuki/irc-protocol, https://github.com/hyunminkang/apigenome, and https://github.com/statgen/hrc.

#### NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-22339-1 | www.nature.com/naturecommunications

6 NATURE COMMUNICATIONS | (2021) 12:2182 | https://doi.org/10.1038/s41467-021-22339-1 | www.nature.com/naturecommunications
previously discovered variants, and variants with Mendelian inconsistencies from included families, variant-level quality controlled was performed using a support vector machine-based classifier on samples with available genotyping data. TOPMed IRC and Data Coordinating Center (DCC) to remove samples genotypic/report inconsistencies for pedigree and sex, and substantial discordance with prior genome-wide array genotyping. Only variants and samples that passed quality control were included in the call set.

One individual from duplicate pairs identified by the DCC was removed, retaining the individual with lipid levels available when one did not have lipid levels. If both individuals had lipid levels, one individual was randomly selected. Individuals were excluded when their genotype determine sex did not match phenotype reported sex (n = 6) and individuals <18 years old where excluded (n = 865).

Blood lipid measurements and phenotypic modeling. Conventionally measured fasting blood lipids, including total cholesterol, LDL-C, HDL-C, and triglycerides, were included for analysis. Harmonization of the lipid values, lipid-lowering medication status, and fasting status at lipid blood draw was performed by the TOPMed Data Coordinating Center. LDL-C was either calculated by the Friedewald equation when triglycerides were <400 mg/dl or directly measured. Given the availability of specific covariates (study site or known founder mutations) where created within each study by self-reported race. Effect sizes are reported in mg/dl or log(mg/dl) for TG.

Coronary heart disease and diabetes mellitus type 2 phenotyping. In the UK Biobank, we used the National Health System OPTIC-4 (Office of Population, Censuses and Surveys: Classification of Interventions and Procedures, version 4) codes K40.1-4.0, K41.1-4.4, K45.5-45.5, K49.1-4.9, K49.8-4.9, K50.2, K57.1-7.4, or K75.8-7.9 to indicate the presence of coronary heart disease. For diabetes mellitus type 2 classification in the UK Biobank, we used the presence of OPTIC-4 codes E11.0-E11.9 or ICD9 code 223.

Coronary heart disease (CHD) in HUNT was defined as individuals with self-reported coronary artery bypass, angioplasty, or stent placement or with diagnosis of ischemic heart disease based on ECG characteristics and/or a history of myocardial infarction or chronic ischemic heart disease based on at least one occurrence of the following codes: ICD9: 410, 411, 412, 414.0, 414.8, 414.9 or ICD10: 121, 122, 123, 124, 125.1, 125.2, 125.5, 125.6, 125.7, 125.8, 125.9. Individuals with angiography were excluded from controls. Type 2 diabetes was defined based on at least one occurrence of the following diagnosis codes: ICD9: 250.00, 250.02, 250.10, 250.12, 250.20, 250.30, 250.32, 250.40, 250.52, 250.55, 250.60, 250.62, 250.70, 250.72, 250.80, 250.82, 250.90, 250.92 ICD10: E11 or by diagnosis of diabetes during HUNT clinical examinations (nonfasting serum or glucose blood > 11.1 mmol/l or Hemoglobin A1c > 6.5%).

In FinnGen, CHD cases were defined as subjects with either an underlying or direct effect on lipid levels. For each of the three variants with lipid-lowering medicines (largely calculated by the Friedewald equation when triglycerides were <400 mg/dl or directly measured). Given the availability of specific covariates (study site or known founder mutations) where created within each study by self-reported race. Effect sizes are reported in mg/dl or log(mg/dl) for TG.

UK Biobank phenotype ascertainment. Our approach to phenome-wide association analyses are similar to prior efforts.59,60. Briefly, a phenome-wide association analysis was performed to evaluate the associations of X chromosome lipid-associated variants with a broad range of clinical phenotypes. A total of 80 manually curated traits were classified according to a combination of self-report and billing codes, except for the following which were based on genotyping data: the UK Biobank data fields: death (40000), ever smoked (20610), BMI (23104), and percentage body fat (23099) (Supplementary Table 13). Lipid-associated variants were associated with each trait, using linear regression for continuous traits and logistic regression for dichotomous traits, adjusting for age, sex, array type, and the first five PCs in the model.

For adiposity analyses, body composition values were obtained using bioelectrical impedance measurement (Category 10009) with a Tanita BC141MA body composition analyzer. Separate readings for fat percentage, mass, and free mass as well as predicted muscle mass are generated for the whole body, trunk, each extremity, and each segment. The total regression coefficients of each bilateral fat-fat association were associated with fat mass (in kilograms) for each of the aforementioned covariates adjusting for age, sex, and the first five PCs in the model. We also separately analyzed the association of lipid-associated variants with abdominal subcutaneous adipose fat (22488) and visceral adipose fat (22487) among unrelated UK Biobank individuals, in which samples from a starting set was performed on each of these phenotypes was natural log-transformed and inverse rank standardized; linear regression models were then adjusted for age, sex, array type, and the first 21 PCs.

For visceal lipid analyses, we included data where baseline visual acuity was available for at least one eye and genotyping data was available at the UKBB.70. Methods from visual acuity assessment in UKBB were previously described. Briefly, visual acuity was measured using the logarithm of the minimum angle of resolution (‘LogMAR’) chart (Precision Vision, LaSalle, Illinois, USA) at a distance of four meters. Using one eye at a time, participants were tasked with identifying the five letters at the top row; they proceeded to read letters from each successive row, which had progressively smaller text. The test was terminated when two or more of the five letters on a given row were read incorrectly. Visual acuity was computed based on the number of successfully read rows; the visual acuity result was provided by the UK Biobank as Field ID 5208 (left eye) and Field ID 5201 (right eye). Data from the right eye was available in 60.421 women and 51.245 men, while data from the left eye was available in 60.402 women and 51.280 men. Linear mixed models from the lme4 package in R (v3.6.1) were used to estimate the association if any of the three variants with visual acuity, adjusting for age, sex, the first five PCs in the model, as well as a random effect accounting for which eye was tested (left or right).

Single-variant association analyses. For discovery, each single variant on the X chromosome of at least 200 unrelated individuals had a suggestive level of association with each adjusted blood lipid residual across all TOPMed samples with lipid levels available (see Blood lipids measurements and phenotypic modeling) using a fast linear mixed model with kinship adjustment (SAIGE-QT, version 0.29.4.48) since a large proportion of TOPMed participants are related in ENCORE (https://encore. nph.ucl.ac.uk). Additionally adjusted for the first 11 PCs in the model. SAIGE-QT was specifically used to maximize computational efficiency given hosting and kinship precompilation by the TOPMed Informatics Research Core. Heterozygous and homozygous were coded as having 1 or 2 nonreference alleles, respectively. Hemizygous males were coded as having two reference alleles. This model was consistent with random X inactivation of one of two X chromosomes in females yet consistent expression of the single X chromosome in males.

For SNPs with suggestive evidence of association in TOPMed (P < 1 × 10−6), we sought replication in UKBB. For replication in UKBB unrelated individuals, we performed linear regression associations using R version 3.6.0. Covariates included age, sex, and the first 10 PCs in the model, based on study characteristics and recommendations from study investigators. We took loci reaching a suggestive association with lipid levels (P < 1 × 10−4) in TOPMed onto replication in UKBB and HUNT. For replication, we used a genome-wide significance level of 0.05/2.2 M variants/4 traits (P < 5.7 × 10−9) to determine if a suggestive level of association in TOPMed. We used a fixed effects meta-analysis to combine the association results from TOPMed, UKBB, and HUNT. We set the statistical significance for our meta-analysis to be alpha = 5.7 × 10−9 (0.05/2.2 M variants/4 traits) which is more stringent than a standard genome-wide significance level of 5 × 10−8. Heterogeneity from the meta-analysis was determined through Cochran’s Q and I² is reported. Additionally, we performed the interaction for rs9585504-T on log triglycerides adjusting for the same covariates as the main analysis.

To determine the correlation of the effect sizes of variants on total cholesterol in the chrXq23 locus and the effect sizes of these variants on CAD, T2D, and BMI, we performed analysis of chrXq23 variation on these three outcomes adjusting for age, sex, genotyping array, and PCs in the UK Biobank, using the main effects model assuming X inactivation. We correlated effect sizes of total cholesterol–variants with effect sizes of variants on CAD, T2D, or BMI limiting to total cholesterol–variants with a MAF > 0.05 and P < 0.05.

Expression quantitative trait analyses. We downloaded the v7 SNP gene association results in tissue-specific files from the GTex portal (https://gtexportal.org/home/datasets). We limited to six tissues that have been implicated in lipid biology or CHD (Adipose_Subcutaneous, Small_Intestine_Terminal_Ileum, Adipose_, Visceral_Omentum, Whole_Blood, Liver, Muscle_Skeletal) and looked at expression of eight genes within the ChrXq23 region (ACSL4, TMEFF1, AMMCR1, RYR5, CHRD1L, PAK3, CAPN6, DCK). We set our significance threshold to 0.001 (0.05/6 tissues x 8 genes). First, we determined eQTls of our top association with lipids. Second, we performed correlation of our lipid–variant associations with the association of each of the eight genes expression on the variants using the gene transcript z-scores K 100 KB. Lastly, we used the Imkem function in the R package kinship2 to reproduce the expression–variant test statistic. We additionally used the expression–variant test statistic to adjust for the correlation between the variants.
Lipoprotein concentrations were measured at LipidomeX (Raleigh, NC) using NMR spectroscopy on plasma EDTA specimens. LipidomeX has developed validated software for analysis of NMR measured LipidomeX spectra that uses an optimized deconvolution algorithm to quantify lipoprotein subtypes. MESA was measured with the Lipoprotein-III assay while FHS samples were measured with the Lipoprotein-III assay, which provides less accuracy for some measurements but is similar to Lipoprotein-III. We associated lipoprotein profiles with top associated SNPs within up to 1,802 FHS and 4,551 MESA participants adjusting for age, sex, and lipid-lowering therapy.

For individuals who participated in ARIC study visit 4 (1996–1998), a homogenous DNA assay method was used for the direct measurement of LDL-C in plasma (sd-LDL-EX “Seiken”, Denka Seiken, Tokyo, Japan) on a Hitachi 917 automated chemist analyzer. We associated top associated SNPs with ARIC participants adjusting for age, sex, lipid-lowering therapy, race, study center, and the first 11 principal components of ancestry.

**Rare variant association analyses.** We performed the SKAT test to associate aggregates of rare coding variants with blood lipids within TOPMed as implemented by GENESIS v2.14.4 in the CHARGE Analysis Commons. This gene-based test, high confidence loss-of-function (HC LOF by LOFTEE) and missense metaSVM damaging variants with MAF ≤ 1% were collapsed into regions based on the gene annotations generated by snpEff 4.3t (https://snpeff.sourceforge.net/) using the gRCh38.86 database.

**Code availability**

The variant calling software tools are under active development; updated versions can be accessed at https://github.com/atks/vt, http://github.com/hyunminkang/apigenome, and https://github.com/statgen/topmed_variant_calling.

**Received:** 28 August 2020; **Accepted:** 2 March 2021; **Published online:** 12 April 2021.

References

1. Peloso, G. M. & Natarajan, P. Insights from population-based analyses of plasma lipids across the allele frequency spectrum. *Curr. Opin. Genet. Dev.* 50, 1–8 (2018).
2. Natarajan, P. et al. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nat. Commun.* 9, 3391 (2018).
3. Abifadel, M. et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* 34, 154–156 (2003).
4. Brown, M. S. & Goldstein, J. L. Familial hypercholesterolemia: defective binding of lipoproteins to cultured fibroblasts associated with impaired regulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity. *Proc. Natl Acad. Sci. USA* 71, 788–792 (1974).
5. Cohen, J. C., Boerwinkle, E., Chakravarti, A., Pepe, T. H. Jr & Hobs, H. H. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* 354, 1264–1272 (2006).
6. Holmes, M. V. et al. Mendelian randomization of blood lipids for coronary heart disease. *Euro. Heart J.* 36, 539–550 (2015).
7. Ferrer, B. A. et al. Variants in CDKAL1 and HMGCR and risk of cardiovascular disease and diabetes. *N. Engl. J. Med.* 375, 2144–2153 (2016).
8. Myocardial Infarction Genetics Consortium, I. et al. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N. Engl. J. Med.* 370, 2072–2082 (2014).
9. Sabatine, M. S. et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa1615664 (2017).
10. Wise, A. L., Gyí, L. & Manolio, T. A. Exclusion: toward integrating the X chromosome in genome-wide association analyses. *Am. J. Hum. Genet.* 92, 643–647 (2013).
11. Coronary Artery Disease Genetics Consortium, A. et al. Genome-wide association studies in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* 43, 339–344 (2011).
12. Consortium, U. K. et al. The UK10K project identifies rare variants in health and disease. *Nature* 526, 82–90 (2015).
13. Hoffmann, T. J. et al. A large electronic-health-record-based genome-wide association study of serum lipids. *Nat. Genet.* 50, 401–413 (2018).
14. Bonas-Guarch, S. et al. Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. *Nat. Commun.* 9, 321 (2018).
15. McNamara, J. R. et al. Effect of gender, age, and lipid status on low density lipoprotein subfraction distribution. Results from the Framingham Offspring Study. *Arterioscler. Thromb.* 8, 490 (1987).
16. Schafer, E. J. et al. Effects of age, gender, and menopausal status on plasma low density lipoprotein cholesterol and apolipoprotein B levels in the Framingham Offspring Study. *J. Lipid Res.* 35, 779–792 (1994).
17. Silberbach, M. et al. Cardiovascular Health in Turner Syndrome: a scientific statement from the American Heart Association. *Circ. Genom. Precis. Med.* 11, e000048 (2018).
18. Stohlholm, K., Juul, S., Juul, K., Naeraa, R. W. & Gravholt, C. H. Prevalence, incidence, diagnostic delay, and mortality in Turner syndrome. *J. Clin. Endocrinol. Metab.* 91, 3897–3902 (2006).
19. Van, P. L., Bakalov, V. K. & Bondy, C. A. Monosomy X is associated with an atherogenic lipid profile. *J. Clin. Endocrinol. Metab.* 91, 2867–2870 (2006).
20. Cooley, M., Bakalov, V. & Bondy, C. A. Lipid profiles in women with 45,XY vs 46,XX primary ovarian failure. *JAMA* 290, 2127–2128 (2003).
21. Belling, K. et al. Klinefelter syndrome comorbidities linked to increased X chromosome gene dosage and altered protein interaction activity. *Hum. Mol. Genet.* 26, 1219–1229 (2017).
22. Zore, T., Palafoux, M. & Reue, K. Sex differences in obesity, lipid metabolism, and inflammation–A role for the sex chromosomes? *Mol. Metab.* 15, 35–44 (2018).
23. Link, J. C. et al. Increased high-density lipoprotein cholesterol levels in mice with XX versus XY sex chromosomes. *Arterioscler. Thromb. Vasc. Biol.* 35, 1778–1786 (2015).
24. Talou, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nat. Genet.* 50, 290–299 (2018).
25. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562, 203–209 (2018).
26. Krokstad, S. et al. Cohort Profile: the HUNT Study, Norway. *Int. J. Epidemiol.* 42, 968–977 (2013).
27. Tabassum, R. et al. Genetic architecture of human plasma lipoproteins and their link to cardiovascular disease. *Nat. Commun.* 10, 4329 (2019).
28. Emdin, C. A. et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *JAMA* 317, 626–634 (2017).
29. Lotta, L. A. et al. Association of genetic variants related to glucosefemoral vs abdominal fat distribution with type 2 diabetes, coronary disease, and cardiovascular risk factors. *JAMA* 320, 2553–2563 (2018).
30. Dahlman, I. et al. Numerous genes in loci associated with body fat distribution are linked to adipose function. *Diabetes* 65, 433–437 (2016).
31. Amabile, M. A., Dorewolds, C., Durand, C. & Popper, M. S. Genome-wide analysis of gene expression during adipogenesis in human adipose-derived stromal cells reveals novel patterns of gene expression during adipocyte differentiation. *Stem Cell Res.* 16, 725–734 (2016).
32. Gustafson, B. et al. BMP4 and BMP antagonists regulate human white and beige adipogenesis. *Diabetes* 64, 1670–1681 (2015).
33. Dahlman, I. et al. Functional annotation of the human fat cell secretome. *Arct. Physiol. Biochem.* 118, 84–91 (2012).
34. Menni, C. et al. Circulating proteomic signatures of chronological age. J. Gerontol. A Biol. Sci. Med. Sci. 70, 809–816 (2015).
35. Ferrucci, L. et al. Low-density lipoprotein and atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Euro. Heart J. 38, 2349–2472 (2017).
36. Liu, D. J. et al. Exome-wide association study of plasma lipids in >300,000 individuals. Am. J. Hum. Genet. 98, 1758–1766 (2016).
37. Lotta, L. A. et al. Association between low-density lipoprotein cholesterol-lowering genetic variants and risk of type 2 diabetes: a meta-analysis. JAMA Cardiol. 3, 1383–1391 (2016).
38. Sattar, N. et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. Lancet 375, 735–742 (2010).
39. Swerdlow, D. I. et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. Lancet 385, 351–361 (2015).
40. 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).
41. White, J. et al. Association of lipin 2 variation with risk for coronary artery disease and diabetes. JAMA Cardiol. 1, 692–699 (2016).
42. Klimentidis, Y. C., Chougue, A., Arora, A., Frazier-Wood, A. C. & Hsu, C. H. Triglyceride-increasing alleles associated with protection against type 2 diabetes. PLoS Genet. 11, e1005204 (2015).
43. Dewey, F. E. et al. ANGPTL4-mediated lipolysis and low-density lipoprotein cholesterol-lowering alleles with risk of coronary disease and type 2 diabetes. JAMA Cardiol. 3, 957–966 (2018).
44. Klimentidis, Y. C., Chougule, A., Arora, A., Frazier-Wood, A. C. & Hsu, C. H. Triglyceride-increasing alleles associated with protection against type 2 diabetes. JAMA Cardiol. 1, 692–699 (2016).
45. Gao, F. et al. XWAS: A software toolset for genetic data analysis and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract 18CDA34110116). B.E.C. is also supported by K01HL135405. S.L. is supported by NIH grant R01HL139731 and American Heart Association 18SPR34250007. Whole-genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung, and Blood Institute (NHLBI). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-0251; contract HHSN268201800002J). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center (3R01HL-120393-0251; contract HHSN268201800001D). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. Please refer to Supplementary Notes 1–4 for study-specific acknowledgements. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

### Conceptual contributions

Conceptualization of work and wrote original draft of manuscript: P.N. and G.M.P. Data curation and provided resources for study: P.N., A.Pampana, S.E.G., S.E.R., P.S.d.V., M.A., L.F.B., R.D., L.S.E., M.R.I., L.A.L., R.N.L., N.D.P., P.A.P., L.R.Y., B.M.P., A.V.S., C.W., and G.M.P. All authors approved the final manuscript.

### Competing interests

P.N. reports grants from Amgen, Apple, Boston Scientific, and Novartis, and consulting income from Apple, Blackstone Life Sciences, Genentech, and Novartis. S.L. reports grants from Bristol Myers Squibb / Pfizer, Bayer AG, and Boehringer Ingelheim, and consulting income from Bristol Myers Squibb / Pfizer and Bayer AG. P.E. is supported by a grant from Bayer AG to the Broad Institute focused on the genetics and therapeutics

### Acknowledgements

This work was supported by grants from the National Heart, Lung, and Blood Institute (NHLBI): R01HL142711 (to P.N. and G.M.P.), K01HL125751 (to G.M.P.). P.N. is also supported by a Hassenfeld Scholar Award from the Massachusetts General Hospital, Fondation Leducq (TNE-18CD04), and additional grants from the National Heart, Lung, and Blood Institute (R01HL148565 and R01HL148560). P.S.d.V. is supported by American Heart Association grant number 18CDA34110116. B.E.C. and J.L. are supported by R35HL135818, H1113338, and HL46380. B.E.C. is also supported by R01HL135405. S.L. is supported by NIH grant R01HL139731 and American Heart Association 18SPR34250007. Whole-genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung, and Blood Institute (NHLBI). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-0251; contract HHSN268201800002J). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center (3R01HL-120393-0251; contract HHSN268201800001D). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. Please refer to Supplementary Notes 1–4 for study-specific acknowledgements. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

### Author contributions

Conceptualization of work and wrote original draft of manuscript: P.N. and G.M.P. Data curation and provided resources for study: P.N., A.Pampana, S.E.G., S.E.R., P.S.d.V., J.G.B., B.W., I.A.R., L.A.F., K.C.H., J.P.M., M.A., S.A., T.L.A., C.M.B., L.F.B., J.C.B., B.E.C., R.D., J.H., L.E., L.Y., J.I.H., A.T.K., I.A.L., J.L., R.N.L., L.W.M., G.M., J.Y.M., J.R.O., N.D.P., J.M.P., A.M.S., M.T., F.F.W., D.E.W., L.R.Y., I.B., E.B., D.W.B., D.M.L.-J., Y.-C.C., Y.I.C., W.J.C., E.C.J., M.D., S.K.D., P.T.E., S.A.L., M.F., B.S.F., S.G., R.A.G., J.H., G.P.I., R.K.C., S.L.K., S.K., I.A.C., L.C.C., D.M.L., S.A.L., K.A.V.M., S.T.M., A.C.M., K.E.N., A-Palote, C.P.J., B.M.P., D.C.R., S.P.R.d.E., M.P.J., D.S.-J., S.S.V., J.L.R., S.S.R., S.R., A.G., and G.M.P. Statistical Analysis: P.N., A.Pampana, S.E.G., J.P.M., B.W., I.A.R., and G.M.P. A.S.R. and supervision of computing studies: P.N., S.E.R., T.L.A., I.A.R., J.C.B., Y.-J.H., R.N.L., M.E.O., J.R.O., P.A.P., J.G.W., G.A., D.K.A., L.C.B., J.B., E.B., Y.-C.C., A.C., E.J.C., K.H., S.L.K., E.K., R.W.K., R.J.L., R.A.M., B.D.M., D.A.N., K.E.N., B.M.P., A.V.S., C.W., and G.M.P. Critically revised manuscript: P.N., A.Pampana, S.E.G., S.E.R., P.S.d.V., M.A., L.F.B., R.D., L.S.E., M.R.I., L.A.L., N.D.P., P.A.P., L.R.Y., J.G.W., I.A.R., J.B., A.C., J.E.C., J. L., C.L.K., E.W., I.J.D., R.A.M., A.C.M., K.E.N., and G.M.P. All authors approved the final manuscript.
cardiovascular diseases. P.E. reports consulting income from Bayer AG, Quest Diagnostics, and Novartis. All others declare no competing interests for the present work.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-22339-1.

Correspondence and requests for materials should be addressed to P.N. or G.M.P.

Peer review information Nature Communications thanks Christopher Nelson and the other, anonymous, reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Pradeep Natarajan, Akhil Pampana, Sarah E. Graham, Sanni E. Ruotsalainen, James A. Perry, Paul S. de Vries, Jai G. Broome, James P. Pirruccello, Michael C. Honigberg, Krishna Aragam, Brooke Wolford, Jennifer A. Brody, Lucinda Antonacci-Fulton, Moscati Arden, Stella Aslibekyan, Themistocles L. Assimes, Christie M. Ballantyne, Lawrence F. Bielak, Joshua C. Bis, Brian E. Cade, Ron Do, Harsha Doddapaneni, Leslie S. Emery, Yi-Jen Hung, Marguerite R. Irvin, Alyna T. Khan, Leslie Lange, Jiwon Lee, Rozenn N. Lemaitre, Lisa W. Martin, Ginger Metcalf, May E. Montasser, Don M. Lloyd-Jones, Yi-Cheng Chang, Yi-Der I. Chen, Won Jung Choi, Adolfo Correa, James P. Pirruccello, Eric Boerwinkle, Michael Tsai, Daniel E. Weeks, Sunil K. Parikh, Michael T. Zinman, John Blangero, Eric Boerwinkle, Donald W. Bowden, Yi-Cheng Chang, Yii-Der I. Chen, Won Jung Choi, Adolfo Correa, Joanne E. Curran, Mark J. Daly, Susan K. Dutcher, Patrick T. Ellinor, Myriam Fornage, Barry I. Freedman, Stacey Gabrielson, Soren Germer, Richard A. Gibbs, Jiang He, Kristian Hveem, Gail P. Jarvik, Robert C. Kaplan, Sharon L. R. Kardia, Emear Kenny, Ryan W. Kim, Charles Kooperberg, Cathy C. Laurie, Seonwook Lee, Don M. Lloyd-Jones, Ruth J. F. Loos, Steven A. Lubitz, Rasika A. Mathias, Karine A. Viaud Martinez, Stephen T. McGarvey, Braxton D. Mitchell, Deborah A. Nickerson, Kari E. North, Aarno Palotie, Cheol Joo Park, Bruce M. Psaty, D. C. Rao, Susan Redline, Alexander P. Reiner, Daekwan Seo, Jeong-Sun Seo, Albert V. Smith, Ramachandran S. Vasan, Sekar Kathiresan, L. Adrienne Cupples, Alanna C. Morrison, Stephen S. Rich, Samuli Ripatti, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, FinnGen* & Gina M. Peloso.

1Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. 2Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. 3Department of Medicine, Harvard Medical School, Boston, MA, USA. 4Department of Internal Medicine: Cardiology, University of Michigan, Ann Arbor, MI, USA. 5Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. 6University of Maryland School of Medicine, Division of Endocrinology, Diabetes and Nutrition and Program for Personalized and Genomic Medicine, Baltimore, MD, USA. 7Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA. 8Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. 9Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. 10Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. 11Department of Engineering Sciences, University of California, Santa Barbara, CA, USA. 12Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, USA. 13Department of Medicine, Stanford University.
School of Medicine, Stanford, CA, USA. 16 VA Palo Alto Health Care System, Palo Alto, CA, USA. 17 Section of Cardiovascular Research, Baylor College of Medicine, Houston, TX, USA. 18 Houston Methodist DeBakey Heart and Vascular Center, Houston, TX, USA. 19 Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. 20 Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. 21 Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. 22 Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan branch, Taipei, Taiwan. 23 Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. 24 Division of Cardiology, George Washington University School of Medicine and Healthcare Sciences, Washington, DC, USA. 25 Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA. 26 Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. 27 Department of Human Genetics and South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA. 28 Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA. 29 Departments of Human Genetics and Biostatistics, University of Pittsburgh, Pittsburgh, PA, USA. 30 Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA. 31 Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. 32 Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA. 33 Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA. 34 Deans office, School of Public Health, University of Kentucky, Lexington, KY, USA. 35 Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan. 36 The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. 37 Psmogen, Inc. (formerly Macrogen USA), Rockville, MD, USA. 38 Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. 39 Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. 40 Cardiac Arrhythmia Service and Cardiovascular Research Center Massachusetts General Hospital, Boston, MA, USA. 41 Institute of Molecular Medicine, the University of Texas Health Science Center at Houston, Houston, TX, USA. 42 Department of Internal Medicine, Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC, USA. 43 Genomics Platform, Broad Institute of Harvard and MIT, Cambridge, MA, USA. 44 New York Genome Center, New York, NY, USA. 45 Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. 46 Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, and Tulane University Translational Science Institute, Tulane University, New Orleans, LA, USA. 47 Department of Public Health and General Practice, HUNT Research Centre, Norwegian University of Science and Technology, Levanger, Norway. 48 K. G. Jebsen Center for Genetic Epidemiology, Dept of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. 49 Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington, Seattle, WA, USA. 50 Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. 51 Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA. 52 The Minda Child Health and Development Institute, Ichan School of Medicine at Mount Sinai, New York, NY, USA. 53 Illumina Laboratory Services, Illumina Inc., San Diego, CA, USA. 54 Department of Epidemiology and International Health Institute, Brown University, Providence, RI, USA. 55 Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA. 56 Department of Genome Sciences, University of Washington, Seattle, WA, USA. 57 University of Washington Center for Mendelian Genomics, Seattle, WA, USA. 58 Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. 59 Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. 60 Departments of Epidemiology and Health Services, University of Washington, Seattle, WA, USA. 61 Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA. 62 The Icelandic Heart Association, Kopavogur, Iceland. 63 Departments of Pathology & Laboratory Medicine and Biochemistry, Larner College of Medicine, University of Vermont, Colchester, VT, USA. 64 Sections of Preventive Medicine and Epidemiology and Cardiology, Department of Medicine, Boston University School of Medicine, Boston, MA, USA. 65 Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA. 66 NHLBI Framingham Heart Study, Framingham, MA, USA. 67 Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. 68 Verve Therapeutics, Cambridge, MA, USA. 69 Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. 70 Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. 71 Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland. 72 Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. * Lists of authors and their affiliations appear at the end of the paper.

email: pnatarajan@mhg.harvard.edu; gpeloso@bu.edu

**NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium**

Namiko Abe44, Christine Albert73, Laura Almasy74, Alvaro Alonso75, Seth Ament76, Peter Anderson77, Pramod Anugu78, Deborah Applebaum-Bowden79, Dan Arking80, Allison Ashley-Koch81, Paul Auer82, Dimitrios Avramopoulos80, John Barnard83, Kathleen Barnes84, R. Graham Barr85, Emily Barron-Casella80, Terri Beaty80, Diane Becker86, Rebecca Bee87, Ferdouse Begum80, Amber Beitelsheshe86, Emilma Benjamin88, Marcos Bezerra89, Larry Bielak90, Thomas Blackwell90, Russell Bowler91, Ulrich Broeckel92, Karen Bunting44, Esteban Burchard93, Erin Buth94, Jonathan Cardwell84, Cara Cart85, Richard Casaburi96, James Casella80, Mark Chaffin97, Christy Chang76, Daniel Chamson98, Sameer Chavan84, Bo-Juen Chen99, Wei-Min Chen99, Michael Cho100, Seung Hoan Choi97, Lee-Ming Chuang101, Mina Chung102, Matthew P. Conomos103, Elaine Cornell104, Carolyn Crandall96, James Crapo105, Jeffrey Curtis90, Brian Custer106, Coleen Damcott76, Elane Cornell101, Sayantan Das90, Sean David108, Colleen Davis77, Michelle Daya88, Mariza de Andrade109, Michael DeBaun110, Ranjan Deka111, Dawn DeMeo100, Scott Devine76, Qing Duan112, Ravi Duggirala113, Jon Peter Durda104, Susan Dutcher114, Charles Eaton115, Lynette Eknue78, Charles Farber99, Leanna Farnam100, Tasha Fingerlin116, Matthew Flickinger90, Nora Franceschini117, Mao Fu76, Stephanie M. Fullerton77, Lucinda Fulton114, Weiniu Gan87, Yan Gao78, Margery Gass118, Bruce Gelb119, Xiaqoi (Priscilla) Geng90,
Soren Germer, Chris Gignoux, Mark Gladwin, David Glahn, Stephanie Gogarten, Da-Wei Gong, Harald Goring, C. Charles Gu, Yue Guan, Xiuqing Guo, Jeff Haessler, Michael Hall, Daniel Harris, Nicola Hawley, Ben Heaver, Susan Heckbert, Ryan Hernandez, David Herrington, Craig Hersh, Bertha Hidalgo, James Hixson, John Hokanson, Elliott Hong, Karin Hoth, Chao (Agnes) Hsiung, Haley Huston, Chii Min Hwu, Rebecca Jackson, Deepi Jain, Cashell Jaquish, Min A. Jhun, Jill Johnsen, Andrew Johnson, Craig Johnson, Rich Johnston, Kimberly Jones, Hyun Min Kang, Laura Kaufman, Shannon Kelly, Michael Kessler, Greg Kinney, Barbara Konkle, Holly Kramer, Stephanie Krauter, Christoph Lange, Ethan Lange, Cecelia Laurie, Meryl LeBoff, Seunggeun Shawn Lee, Wen-Jane Lee, Jonathon LeFai, David Levine, Dan Levy, Joshua Lewis, Yun Lin, Honghuang Lin, Xihong Lin, Yongmei Liu, Kathryn Lunetta, James Luo, Michael Mahaney, Barry Make, Ani Manichaikul, JoAnn Manson, Lauren Margolin, Susan Mathai, Patrick McArdle, Merry-Lynn McDonald, Sean McFarland, Caitlin McHugh, Hao Mei, Deborah A. Meyers, Julie Mikulla, Nancy Min, Mollie Minear, Ryan L. Minster, Solomon Musani, Stanford Mwasongwe, Josyf C. Mychaleckyj, Girish Nadkarni, Rakhi Naik, Take Naseri, Sergei Nekhai, Sarah C. Nelson, Deborah Nickerson, Jeff O’Connell, Tim O’Connor, Heathcrot Balcom, James Pankow, George Papanicolaou, Margaret Parker, Afshin Parsa, Sara Penchev, Marco Perez, Ulrike Peters, Lawrence S. Phillips, Sam Phillips, Toni Pollin, Wendy Post, Julia Powers Becker, Meher Preethi Boorgula, Michael Preuss, Dmitry Prokopenko, Pankaj Qasba, Dandi Qiao, Nicholas Rafaels, Laura Rafaield, Laura Rasmussen-Torvik, Aakrosh Ratan, Robert Reed, Elizabeth Regan, Muagututi’a Sefuiva Reupena, Ken Rice, Dan Roden, Carolina Roselli, Ingo Ruczinski, Pamela Russe, Sarah Ruuska, Kathleen Ryan, Ester Cerdeira Sabino, Phwanat Sakornsakolpat, Shabnam Salimi, Steven Salzberg, Kevin Sandow, Vijay G. Sankaran, Christopher Scheller, Ellen Schmidt, Karen Schwander, David Schwartz, Frank Scurba, Christine Seidman, Jonathan Seidman, Vivien Sheehan, Amol Shetty, Wayne Hui-Heng Sheu, M. Benjamin Shoemaker, Brian Silver, Edwin Silverman, Jennifer Smith, Josh Smith, Nicholas Smith, Tanja Smith, Sylvia Smoller, Beverly Snively, Tamar Sofer, Nona Sotoodehnia, Elizabeth Streeten, Jessica Lasky Su, Yun Ju Sung, Jody Sylvest, Adam Szpiro, Carole Szalard, Daniel Taliun, Hua Tang, Margaret Taub, Kent D. Taylor, Simeon Taylor, Marilyn Telen, Timothy A. Thornton, Lesley Tinke, David Tirschwell, Heman Tiwari, Dhananjay Vaidya, Peter VandeHaar, Scott Vrieze, Tarik Walker, Robert Wallace, Avram Walts, Emily Wan, Heming Wang, Karol Watson, Bruce Weir, Scott Weiss, Lu-Chen Weng, Kayleen Williams, L. Keoki Williams, Carla Wilson, Quenna Wong, Huichun Xu, Ivana Yang, Rongze Yang, Norann Zaghloul, Maryam Zekavat, Yingze Zhang, Snow Xueyan Zhao, Wei Zhao, Degui Zhi, Xiang Zhou, Xiaofeng Zhu, Michael Zody, and Sebastian Zoellner.

73Brigham & Women’s Hospital, Cedars Sinai, Boston, Massachusetts, USA. 74Children’s Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA, USA. 75Emory University, Atlanta, GA, USA. 76University of Maryland, Baltimore, MD, USA. 77University of Washington, Seattle, WA, USA. 78University of Mississippi, Jackson, MS, USA. 79National Institutes of Health, Bethesda, MD, USA. 80Johns Hopkins University, Baltimore, MD, USA. 81Duke University, Durham, NC, USA. 82University of Wisconsin Milwaukee, Milwaukee, WI, USA. 83Cleveland Clinic, Cleveland, OH, USA. 84University of Colorado at Denver, Denver, CO, USA. 85Columbia University, New York, NY, USA. 86Johns Hopkins University, Medicine, Baltimore, MD, USA. 87National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA. 88Boston University, Massachusetts General Hospital, Boston University School of Medicine, Boston, MA, USA. 89Fundação de Hematologia e Hemoterapia de Pernambuco - Hemope, Recife, Brazil. 90University of Michigan, Ann Arbor, MI, USA. 91National Jewish Health, National Jewish Health, Denver, CO, USA. 92Medical College of Wisconsin, Milwaukee, WI, USA. 93University of California, San Francisco, San Francisco, CA, USA. 94University of Washington, Biostatistics, Seattle, WA, USA. 95Women’s Health Initiative, Seattle, WA, USA. 96University of California, Los Angeles, Los Angeles, CA, USA. 97Broad Institute, Cambridge, MA, USA. 98Brigham & Women’s Hospital, Division of Preventive Medicine, Boston, MA, USA. 99University of Virginia, Charlottesville, VA, USA. 100Brigham & Women’s Hospital, Boston, MA, USA. 101National Taiwan University, National
