The association between mitochondrial DNA copy number, low-density lipoprotein cholesterol, and cardiovascular disease risk

Xue Liu1, Xianbang Sun1, Yuankai Zhang1, Wenzing Jiang1, Lai Meng1, Kerri L. Wiggins2, Laura M. Rafffield3, Lawrence F. Bielak4, Wei Zhao5, Achilles Pitsillides6, Jeffrey Haessler7, Yinan Zheng8, Thomas W. Blackwell9, Jie Yao8, Xiuqing Guo8, Yong Qian8, Bharat Thyagarajan10, Nathan Pankratz11, Stephen S. Rich12, Kent D. Taylor12, Patricia A. Peyser4, Susan R. Heckbert13, Sudha Seshadri14,15,16, Eric Boerwinkle17,18, Megan L. Grove17, Nicholas B. Larson19, Jennifer A. Smith4, Ramachandran S. Vasan15,20, Annette L. Fitzpatrick21, Myriam Fornage22, Jun Ding23, April P. Carson23, Goncalo Abecasis7, Josée Dupuis1,24, Alexander Reiner5, Charles Kooperberg5, Lifang Hou6, Bruce M. Psaty2,25, James G. Wilson26, Daniel Levy15,27, Jerome I. Rotter8, Joshua C. Bis5, TOPMed mtDNA Working Group in NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Claudia L. Satizabal14,15,16,*, Dan E. Arking28,*, and Chunyu Liu1,15,*,##.@.

1. Department of Biostatistics, School of Public Health, Boston University, Boston, MA 02118, USA;.
2. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA;.
3. Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA;.
4. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA;.
5. Fred Hutchinson Cancer Research Center, Division of Public Health Science, Seattle, WA 98109, USA;.
6. Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA;.
7. TOPMed Informatics Research Center, University of Michigan, Ann Arbor, MI 48109, USA;.
8. The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA 90502, USA;.
9. Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, NIH, Baltimore, MD 21224, USA;.
10. Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA;.
11. Department of Computational Pathology, University of Minnesota, Minneapolis, MN 55455, USA;.
12. Department of Public Health Services, Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA;.
13. Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, WA 98101, USA;.
14. Glenn Biggs Institute for Alzheimer’s and Neurodegenerative Diseases, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA;.
15. Framingham Heart Study, NHLBI, 73 Mt. Wayte Avenue, Suite 2, Framingham, MA 01702, USA;.
16. Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA;.
17. Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA;.
18. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030, USA;.
19. Division of Clinical Trials and Biostatistics, Department of Quantitative Health Sciences, Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA;.
20. Sections of Preventive Medicine and Epidemiology, and Cardiovascular Medicine, Boston University School of Medicine, Boston, MA, 02118, USA;.
21. Departments of Family Medicine, Epidemiology, and Global Health, University of Washington, Seattle, WA 98195, USA;.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

CC-BY 4.0 International license
It is made available under a CC-BY 4.0 International license.
22. Center for Human Genetics, University of Texas Health Science Center at Houston, Houston, TX 77030, USA;
23. Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA;
24. Department of Epidemiology, Biostatistics and Occupational Health, School of Population and Global Health, McGill University Faculty of Medicine and Health Sciences, Montréal, QC H3G 2M1, Canada.
25. Departments of Epidemiology, and Health Systems and Population Health, University of Washington, Seattle, WA 98101, USA;
26. Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA;
27. Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA.
28. McKusick-Nathans Institute, Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, MD 21205, USA;

# These authors contributed equally to the paper.
@Correspondence: liuc@bu.edu
Abstract

Mitochondria are the primary organelle to generate cellular energy. Our group and others have reported that lower mitochondrial DNA copy number (mtDNA CN) is associated with higher risk of cardiovascular disease outcomes (CVD) and higher LDL levels. However, the causal relationship between mtDNA CN and CVD remains to be studied. Here we performed cross-sectional and prospective association analyses of blood-derived mtDNA CN and CVD outcomes in up to 27,316 participants from different racial/ethnic groups with whole genome sequencing. We validated most of the previously reported associations but effect sizes were smaller in this study. For example, one SD unit decrease in mtDNA CN was significantly associated with 1.08-fold (95% CI, 1.04, 1.12; \( P = 1.7 \times 10^{-4} \)) hazard for developing incident coronary heart disease (CHD) adjusting for age, sex and race/ethnicity. We conducted Mendelian randomization (MR) to explore causal relationships between mtDNA CN, LDL, and CHD. Bi-directional univariable MR analyses provided strong evidence indicating higher LDL level is causally associated with lower mtDNA CN, and CHD was weakly associated with lower mtDNA CN. We found no evidence supporting a causal association for lower mtDNA CN with higher CHD risk or higher LDL. In multivariable MR, no associations were observed between mtDNA CN and CHD controlling for LDL level (\( P = 0.92 \)), whereas strong evidence for a direct causal effect was found for higher LDL on lower mtDNA CN, adjusting for CHD status (\( P = 8.3 \times 10^{-10} \)). Findings from this study indicate high LDL underlies the complex relationships between vascular atherosclerosis and lower mtDNA CN.
Introduction

Cardiovascular diseases (CVDs) are a leading cause of death globally.\textsuperscript{1} A large proportion of CVDs is due to atherosclerosis, an inflammation process in the blood vessels.\textsuperscript{2,3} Mitochondria are primary sites for oxidative phosphorylation machinery that generates adenosine triphosphate (ATP).\textsuperscript{4} Mitochondria have their own DNA (mtDNA), a circular 16.6-kb molecule encoding essential proteins for ATP production and energy homeostasis.\textsuperscript{5} The role of mtDNA integrity and mitochondrial dysfunction in the pathogenesis of atherosclerosis has been studied in genetic knockout mice with defects in the apolipoprotein E (ApoE) and the catalytic subunit of the mtDNA polymerase gamma subunit (POLG).\textsuperscript{6,7} In these knockout mice, mtDNA damage was an early event during the initiation of atherogenesis and may result from reactive oxygen species.\textsuperscript{8} Accumulation of mtDNA damage promoted atherosclerosis and was associated with the formation of vulnerable plaques.\textsuperscript{9} Moreover, trapping and oxidization of low-density lipoprotein (LDL) cholesterol at arterial walls also plays a central role in the formation of the atherosclerotic lesion and the progression of CVD.\textsuperscript{3,10,11} Accumulation in the subendothelial space of the arterial wall leads LDL to undergo oxidation to become oxidized LDL (oxLDL). This can lead to increased mitochondrial permeability, which may cause subsequent damage to mtDNA.\textsuperscript{12,13}

Each human cell contains hundreds (e.g., in a blood cell) or even thousands (e.g., in a cardiac muscle cell) of mitochondria, and multiple copies of mtDNA are present per mitochondrion.\textsuperscript{14} The copy number of mtDNA (i.e., mtDNA CN) is strictly regulated for energy homeostasis and may serves as a surrogate marker of mitochondrial function.\textsuperscript{15,16} Thus, reduced mtDNA CN may serve as a biomarker of
mitochondrial dysfunction. A lower level of mtDNA CN has also been associated with a general decline in health, all-cause mortality, and associated with multiple cardiometabolic traits including higher level of LDL and hyperlipidemia that are major risk factors for CVD after controlling for other risk factors. Recent prospective studies have also reported significant associations between lower mtDNA levels and CVD outcomes. However, the causal relationship among mtDNA CN, LDL and CVD remains to be determined.

To that end, this study had two aims. The first aim was to validate the associations of mtDNA CN with CVD outcomes and total mortality using blood-derived mtDNA CN measured from whole genome sequencing (WGS) in eight cohorts of diverse populations. The previous mtDNA CN - CVD association studies used mtDNA CN measured by array-based methods or by qPCR in fewer cohorts. The second aim was to explore the causal relationships between mtDNA CN, LDL, and coronary heart disease (CHD) using Mendelian randomization (MR), a method that has been increasingly used to minimize issues of confounding and reverse causation with genetic variants as an instrumental variable (Figure 1).

Methods

Study population

This study included participants with WGS from eight prospective cohort studies, some with participants from different ethnic/racial groups (Supplemental Table 1): the Atherosclerosis Risk in Communities study (ARIC) (n=3,585), the Coronary Artery
Risk Development in Young Adults Study (CARDIA)\textsuperscript{27} (n=3,473), the Cardiovascular Health Study (CHS)\textsuperscript{28} (n=3,546), the Framingham Heart Study (FHS)\textsuperscript{29,30,31} (n=4,133), the Genetic Epidemiology Network of Arteriopathy (GENOA)\textsuperscript{32} (n=1,253), the Jackson Heart Study (JHS)\textsuperscript{33} (n=3,286), the Multi-Ethnic Study of Atherosclerosis Study (MESA)\textsuperscript{34} (n=4,596), and the Women’s Health initiative Study (WHI)\textsuperscript{35} (n=7,197). Except for WHI, in which only female participants were included, the other seven cohorts included both men and women. The CHS only included participants 65 and older (mean age 74 years) and the other cohorts included mostly middle-aged participants at blood-draw for this study (mean age ranging from 58 to 69). MESA excluded participants with any clinically recognized CVD at the baseline visit,\textsuperscript{34} while the other cohorts contained prevalent CVD cases at baseline. Several of the cohorts contained a small number of duplicate participants (n =136) due to study design and data collection\textsuperscript{26,32,33}. We removed these duplicate participants from subsequent association analyses.

**mtDNA CN estimation in whole genome sequencing**

WGS was performed from one of the TOPMed sequencing centers using blood-derived DNA for all participants in the eight cohorts. The average genome-wide coverage was ~39-fold across samples in TOPMed.\textsuperscript{36} The TOPMed Information Research Center conducted analyses to estimate mtDNA CN across all participants using the program \textit{fastMitoCalc} of the software package \textit{mitoAnalyzer}\.\textsuperscript{37} Because nuclear DNA (nDNA) is diploid while mitochondrial DNA is haploid, the average mtDNA CN per cell was
estimated as twice the ratio of the average coverage of mtDNA to the average coverage of the nuclear DNA (nDNA).³⁷

**Cardiovascular disease traits and total mortality**

The eight longitudinal cohorts in this study have been established to investigate risk factors contributing to CVD, morbidity and mortality. Each cohort used standardized definitions to adjudicate CVD outcomes. CHD was defined as the first incident myocardial infarction (MI) or death owing to CHD and cardiac procedures (typically revascularization).³⁸ Stroke was defined as the first nonfatal stroke or death owing to stroke.³⁹ Heart failure is a complex clinical syndrome resulting from a structural or functional cardiac disorder that impairs the ability of one or both ventricles to fill with or eject blood sufficiently to meet the needs of the body.⁴⁰ CVD included CHD, stroke, and heart failure, and death due to CHD, stroke and heart failure. All-cause mortality included all deaths. We analyzed associations of mtDNA CN with three prevalent and incident CVD outcomes (CHD, stroke, and CVD) and with all-cause mortality.

**Covariates**

In the primary analysis, age at blood draw, sex, study center (if applicable), and self-reported racial/ethnic group were adjusted for in the base model. Additional variables included body mass index (BMI, kg/m²), fasting plasma lipid measures including total cholesterol (TC, mg/dL) and high density lipoprotein cholesterol (HDL, mg/dL), systolic blood pressure (SBP, mmHg), treatment for high blood pressure or hypertension (HRX), current smoking status, and diabetes status. Diabetes was defined as fasting blood
glucose level of ≥126 mg/dL or currently receiving medications to lower blood glucose levels to treat diabetes. This study used mtDNA CN calculated using WGS of blood-derived DNA. Different blood cell types (e.g. neutrophils, lymphocytes, etc…) contain different levels of mtDNA CN. To minimize potential confounding, we accounted for white blood cell count and differential components (the proportions of neutrophils, lymphocytes, monocytes, eosinophils, and basophils) and platelet count in association analyses in cohorts in which these cell count variables were available.

Association analyses of mtDNA CN with CVD outcomes and total mortality

For primary analyses, we generated mtDNA CN residuals by regressing mtDNA CN on age, age-squared, sex and blood collection year (as a categorical variable) in each cohort. For age-stratified analysis, we generated mtDNA CN residuals by regressing mtDNA on sex and blood collection year in each cohort. For sex-stratified analysis, we generated mtDNA CN residuals by regressing mtDNA on age, age-squared, and blood collection year in each cohort. The residuals were standardized to a mean of zero and standard deviation (s.d.) of one. The standardized residuals were used as the main predictor in all regression models. We removed participants whose mtDNA CN standardized residuals were greater than 4 s.d. from the mean.

We performed cohort-specific association analyses between mtDNA CN and outcomes. We used logistic regression to quantify the associations of mtDNA CN with prevalent CVD outcomes. We used a Cox proportional hazards regression model to quantify the association of mtDNA CN with incident CVD outcomes and total mortality in
all cohort-specific analyses. Due to a special study design in selecting participants for WGS in WHI, we applied a weighted logistic regression for cross-sectional outcome or a weighted Cox proportional hazards regression for incident outcomes in WHI (Supplemental Methods). We performed three models for association analyses of mtDNA CN with both prevalent and incident outcomes. Model 1 included age, sex, study center (if applicable), and race/ethnicity. In Model 2, we additionally adjusted for several traditional covariates including BMI, TC, HDL, SBP, HRX, current smoking and diabetes for CVD outcomes. For analyzing total mortality as the outcome, we excluded participants who have prevalent CHD or diabetes and adjusted for BMI, TC, HDL, SBP, HRX and current smoking in Model 2. In Model 3, white blood cell and differential counts as well as platelet counts were further adjusted in addition to covariates in Model 2. We used an inverse variance meta-analysis with a fixed effects model to summarize cohort-specific association analyses. An odds ratio (OR) or a hazard ratio (HR) was reported corresponding to one s.d. decrease in the mtDNA CN level.

In secondary analyses, we performed association analyses between mtDNA CN and outcomes in: 1) male and female only samples, and 2) in participants who were at least aged 60 years at blood draw for WGS. We also performed several sensitivity analyses in FHS to investigate if different cardiometabolic disease status (hypertension, diabetes and hyperlipidemia) may result in different directionalities or effect sizes in associations of mtDNA CN with CVD.

Mendelian randomization
LDL plays a central role in the pathogenesis of CVD outcomes including CHD. To thoroughly evaluate the causal relationship between CHD and mtDNA CN, we first conducted univariable bi-directional two sample MR analysis to assess the causal relationship between LDL and mtDNA CN, and between CHD and mtDNA CN. We used single nucleotide polymorphisms (SNPs) (LD $r^2 < 0.001$ based on EUR population reference panel) with no ambiguous allele information (i.e., palindromic SNPs). We further excluded SNPs that are known to be pleiotropic in MR analysis (i.e., the missense mutations rs7412, rs429358, rs768374191 and rs367866106 in ApoE). We used the inverse variance weighted (IVW) method to combine the causal effects of independent SNPs. Leave-one-out plots were used to detect influential outliers and MR-PRESSO was used to detect and correct for potential outliers. We also conducted several sensitivity analyses, including MR-Egger regression, Cochran’s Q statistic and funnel plots, and obtained median and mode estimates to test the validity of MR estimators. In secondary analysis to test causality of mtDNA CN to CHD, we conducted MR analysis using SNPs identified by Gene Ontology analysis. These selected SNPs are directly involved in mitochondrial functions. (Supplemental Table 2). TwoSampleMR package (version 0.5.0) in R (version 0.5.6) was used for univariable MR analyses.

Because common genetic variants are associated with both LDL and CHD, horizontal pleiotropy effect may exist, which violates the 3rd assumption of MR. To account for this potential pleiotropic effect and to simultaneously investigate the direct effect of either LDL or CHD on mtDNA CN, we conducted multivariable MR
We performed the extended framework of IVW and MR-Egger methods to estimate causal effects in multivariable MR analysis. We used GWAS results from non-overlapping participants for each exposure (see below) and thus the covariance between the effects of the genetic variants on each exposure was fixed at zero. We used generalized Cochran’s Q test to assess instrumental variable (IV) validity in the two-sample summary data setting. The MVMR package (version 0.2.0) in R (version 0.5.6) was used for MVMR analyses.

In all MR analyses, we used significant SNPs ($p < 5e-8$) as IVs identified from genome-wide association studies (GWAS). The significant SNPs of mtDNA CN GWAS were recently reported from the Cohorts of Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium and the UK Biobank data. In the present study, we used the mtDNA CN GWAS results for the selected SNPs from the ~300,000 UK Biobank only participants analysis. We identified the CHD-associated SNPs list and extracted the summary GWAS results for the identified SNPs from UK Biobank participants (10,157 cases and 351,037 controls, 2018, the Neale lab, https://gwas.mrcieu.ac.uk/datasets/ukb-d-I9_CHD/). We also identified SNPs from the Global Lipids Genetics Consortium summary GWAS results of LDL from analyses of 1,166,583 multi-ancestries participants (Graham, Kanoni, Ramdas, et al., 2021 http://csg.sph.umich.edu/willer/public/glgc-lipids2021/).

Results

Participant characteristics
This study included up to 27,316 participants (mean age 62, age range of 20-98 years, and 68% women) with 16636 (60.9%) European Americans (EA), 8709 (31.9%) African Americans (AA), 1229 (4.5%) Hispanic/Latino Americans (HA) and 728 (2.7%) East Asian Americans (Supplemental Table 1) (EAS). The prevalence of any CVD outcomes was higher in African Americans (13.4%) than other ethnic and racial groups (EA: 9.4%, HA: 9.8%, EAS: 10.2%). During a median of 6 to 16 years (across the cohorts) of follow-up, the prevalence and incidence rates of CVD outcomes varied across cohorts (Supplemental Table 1).

Association analyses of mtDNA CN with prevalent CVD outcomes

In total, 2,162 (7.9%) participants had prevalent CHD, 1154 (4.2%) had prevalent stroke, and 3397 (12.4%) had prevalent CVD at baseline (Supplemental Table 1). Meta-analysis showed that one s.d. decrease in mtDNA CN was significantly associated with a 1.11-fold odds of CHD (95% CI, 1.07, 1.16; \( P = 1.3 \times 10^{-7} \)), 1.13-fold odds of stroke (95% CI, 1.05, 1.22; \( P = 0.002 \)) and 1.14-fold odds of CVD (95% CI, 1.11, 1.16; \( P = 1.9 \times 10^{-33} \)), adjusting for age, sex and race/ethnicity (Figure 2, Model 1). The associations were slightly attenuated after further adjusting for traditional CVD risk factors (Supplemental Figure 1, Model 2) and white blood cell count in addition to traditional CVD risk factors (Supplemental Figure 2, Model 3). The association directions were consistent across 6 of the 7 cohorts with one null association for CVD outcomes (Figure 2).
Association analyses of incident CVD outcomes and all-cause mortality

A total of 24,019 participants free of CVD at baseline was followed up for a median of 12 years. During the follow-up, 3975 (16.5%) had incident CHD, 5208 (21.7%) had incident stroke, and 8590 (35.4%) had incident CVD. Meta-analysis showed that one s.d. decrease in mtDNA CN was significantly associated with a 1.08-fold of hazards for developing incident CHD (95% CI, 1.04, 1.12; \( P=1.7\text{E}-04 \)), 1.07-fold of hazards for developing incident CVD (95% CI, 1.03, 1.10; \( P=3.9\text{E}-05 \)) when we adjusted for age, sex and race/ethnicity in association analyses (Figure 3). The associations were slightly attenuated after further adjusting for traditional CVD risk factors in Model 2 (Incident CHD: HR =1.05; 95% CI, 1.01, 1.09; \( P= 0.023 \); Incident CVD: HR =1.05; 95% CI, 1.02, 1.09; \( P= 2.3\text{E}-03 \)) (Supplemental Figure 3). The associations were also changed little after additionally adjusting for white blood cell count/differential count and platelet count in Model 3 (Incident CHD: HR =1.07; 95% CI, 1.02, 1.12; \( P= 4.3\text{E}-03 \); Incident CVD: HR =1.06; 95% CI, 1.03, 1.10; \( P= 9\text{E}-04 \)) (Supplemental Figure 4). Incident stroke was not significantly associated with mtDNA CN in meta-analyses of the three models (Supplemental Figure 3, Supplemental Figure 4).

Examining the individual cohorts, we found that lower mtDNA CN was associated with higher hazards for developing incident CHD and incident CVD in five cohorts, with ARIC displaying the strongest associations while FHS and WHI showed weak inverse associations or no association (Figure 3). Sensitivity analysis removing ARIC showed non-statistically significant result. Additional sensitivity analyses in FHS demonstrated that several factors, including age, sex, hypertension status, diabetes status and hyperlipidemia status were not the cause for the inverse effect or null association.
observed in FHS compared to other cohorts although the magnitude of associations seemed different in stratified analyses (Supplemental Table 3).

A total of 8018 (33.3%) participants died due to any cause during the follow-up. One s.d. decrease in mtDNA CN was significantly associated with 1.06-fold of hazards for all-cause mortality (95% CI, 1.03-1.09; \( P=2.5\times10^{-5} \)) adjusting for age, sex and race/ethnicity. All the cohorts showed consistent directionality between mtDNA CN and total mortality in Model 1, that is, lower mtDNA CN was associated with higher rates of all-cause mortality (Figure 4). The associations were very similar after further adjusting for multiple clinical covariates in Model 2 (HR =1.05, 95% CI, 1.02-1.08; \( P= 5.9\times10^{-4} \)) and additionally adjusting for cell counts/differential components and platelet count in Model 3 (HR =1.06, 95% CI, 1.02-1.10; \( P= 1.1\times10^{-3} \)) (Supplemental Figure 5).

Mendelian randomization analyses to test causality

We performed bi-directional univariate MR analyses to test the causal relationships between mtDNA CN, LCL, and CHD. We used 77 (primary analysis) independent SNPs (LD \( r^2 < 0.001 \)) as IVs to test the causal relationship of mtDNA CN on CHD (Supplemental Table 4). The IVW MR analyses yielded insufficient evidence (OR = 1.0, 95% CI, 0.997, 1.010; \( P=0.25 \)) to support that lower mtDNA CN was casually associated with higher odds of CHD. Sensitivity analyses and the secondary analysis with 23 SNPs that are directly involved in mitochondrial function further confirmed no causal relationship from mtDNA CN to CHD (Supplemental Table 5, Supplemental Figure 6). To test for the causal relationship of CHD on mtDNA CN, we used 10
significant GWAS SNPs (LD $r^2 < 0.001$) (Supplemental Table 7) as IVs in MR analyses. The IVW MR analysis supported that having CHD displayed a borderline significant causal relationship with a lower mtDNA CN level ($\beta = -0.86; 95\% \text{ CI, } -1.80, -0.08; P=0.072; \text{Egger } P=0.067$) (Supplemental Table 6, Supplemental Figure 7). For all the above MR analyses, heterogeneity was observed although results remained unchanged after correction by MR-PRESSO. In addition, the MR-Egger test did not provide significant evidence of directional pleiotropy (Supplemental Table 5).

To account for the observed heterogeneity with the 10 SNPs ($P = 9.6E-09$), we applied multiple sensitivity analyses to test for validity of MR estimators. Three of six sensitivity analyses gave rise to significant causal effects of CHD on mtDNA CN (Supplemental Table 6). For example, weighted median test ($\beta = -1.06; P = 0.0036$) and weighted mode ($\beta = -1.48; P = 0.0041$).

We selected 75 independent GWAS SNPs associated with mtDNA CN (Supplemental Table 8) to test the causal relationship of mtDNA CN on LDL. The IVW MR analysis gave rise to weak evidence that lower mtDNA CN was borderline significantly associated with a higher LDL ($\beta=0.067, 95\% \text{ CI, } -0.0036, 0.14; P=0.059; \text{Egger } P=0.22$) (Supplemental Table 9, Supplemental Figure 8). The corrected effect of mtDNA CN on LDL is not statistically significant ($\beta=0.025, 95\% \text{ CI=} -0.0024, 0.052, P\text{-value=} 0.08$) after removing outliers identified by the MR-PRESSO test. No directional pleiotropy was detected using the MR-Egger method (Supplemental Table 9). All sensitivity analyses yielded non-statistically significant causal effect of mtDNA CN on LDL-C ($P > 0.05$, Supplemental Table 9).
To test the causal relationship of LDL-C on mtDNA CN, 345 SNPs were selected as IVs (Supplemental Table 10). The IVW MR analysis showed a strong causal relationship between lower LDL and higher mtDNA CN (beta=0.084, 95% CI=0.062, 0.11, P-value=1.1E-14). High heterogeneity was observed in the IVW analysis and thus MR-PRESSO was performed. After outliers were removed by MR-PRESSO, the corrected causal effect became more statistically significant (beta=0.077, 95% CI=0.062, 0.092, P-value=1.0E-21) than the IVW MR result. All the sensitivity analyses presented statistically significant causal relationship of LDL on mtDNA CN (P<0.0001) (Supplemental Table 9).

Given the findings from univariable MR analyses and the previous findings that LDL is a primary cause for CHD development,\textsuperscript{10,11} we conducted MVMR to estimate the direct effect of either CHD or LDL on mtDNA CN controlling for each other. After excluding SNPs with a pairwise $r^2>0.001$, 346 SNPs were used in the analysis. We observed strong evidence for a direct causal effect of LDL on mtDNA CN adjusting for CHD, that is, one s.d. decrease in genetically predicted LDL level was causally associated with 0.083 s.d. higher mtDNA CN level (IVW beta=0.083, 95% CI= 0.059, 0.11, \(P =1.2E-12\)) (Table 1, Figure 4). In contrast, the direct causal effect of CHD on mtDNA CN was not significant controlling for LDL level (IVW beta=-0.1, 95% CI= -0.57, 0.77, \(P =0.76\)) (Table 1). The MVMR-Egger test yielded consistent results as those from IVW MVMR analysis for both LDL and CHD.

Discussion
In this study, we validated the association of mtDNA CN with prevalent and incident CVD outcomes (except for incident stroke), as well as all-cause mortality during a median 12 years of follow-up in up to 27,316 participants from eight cohort studies including self-identified European Americans, African Americans, Hispanic/Latino Americans, and East Asian Americans. The associations of mtDNA CN with the outcome variable remained statistically significant after further adjustment for traditional clinical variables (i.e. total cholesterol and HDL, etc) and blood cell counts. More importantly, we performed comprehensive univariable and multivariable MR analyses, using SNPs identified from the latest GWAS for CHD\textsuperscript{44,62,63,64}, LDL\textsuperscript{65,66} and mtDNA CN\textsuperscript{61} to explore the causal relationships between mtDNA CN, LDL and CHD. We applied the MR-PRESSO and MR-Egger tests as well as several sensitivity analyses to minimize the bias and to test validity of MR analyses. The MR analyses implicate that elevated LDL levels as the primary driver for the observed significant association of mtDNA CN with CHD.

It has always been challenging to assess causality in epidemiological association analyses. The bi-directional univariable MR analyses found that having CHD seemed to have a causal effect on lower mtDNA CN level rather than an opposite direction that lower mtDNA CN had causal effect on CHD. CHD is a multifactorial end-point disease that is characterized as the reduction of blood flow to the heart muscle due to build-up of atherosclerotic plaque.\textsuperscript{3} Studies have consistently shown that atherosclerosis is initiated by excess LDL levels in the plasma. In addition, our recent study reported that higher LDL levels are associated with lower mtDNA CN in blood.\textsuperscript{21} Thus, it was necessary to consider LDL- possibly play a major role in assessing the causal
relationship between CHD and mtDNA CN. Bi-directional MR supported that LDL has a causal effect on mtDNA CN while mtDNA CN had no causal effect on LDL. Given it is well known that LDL is a causal factor for the development of CHD\(^6\) and that LDL and CHD share common genetic variants, we performed a MVMR analysis to assess the direct causal effect of CHD or LDL on mtDNA CN. We observed a significant, direct causal effect of LDL on mtDNA CN adjusting for CHD (Figure 5A). In contrast, the direct causal effect of CHD on mtDNA CN became insignificant controlling for LDL. Based on these findings, it is reasonable to conclude that the observed association between mtDNA CN and CVD outcomes (prevalent and incident) is a manifestation of the causal effect of higher LDL levels on lower mtDNA CN.

Our findings are supported by recent advances focusing on elucidating the role of oxidative stress and mitochondrial dysfunction in vascular inflammation and atherosclerosis in animal models.\(^6,7,8\) LDL in the plasma is the primary molecule that triggers a cascade of inflammation responses. The excess of LDL is oxidized into oxidized LDL (oxLDL), which attracts immune cells like monocytes into the arterial wall. Monocytes then differentiate into macrophages which swallow oxLDL and become foam cells, the most abundant immune cells within the atherosclerotic lesion.\(^6,7,8\) CHD occurs when the plague buildup is ruptured to form a large thrombus. On the other hand, the oxLDL and other factors can increase the permeability of the mitochondrial transition pore, lead to the swelling of the mitochondrial matrix, and the increase in ROS production in mitochondria.\(^68,69\) The elevated ROS production in mitochondria may damage the mtDNA and eventually result in mitochondrial dysfunction.\(^70\) The MR analyses in our study support that higher LDL is the driver for CHD and lower mtDNA
CN levels (Figure 5B). Nonetheless, the role of mtDNA CN in the atherosclerotic formation, the pathogenesis of CVD, and inflammation is complex and warrants further investigation.

**Limitations of the study**

Heterogeneity was observed in the association estimates of CVD outcomes across cohorts even though we harmonized phenotypes and accounted for confounders and known batch effects in association analyses with mtDNA CN. This remaining heterogeneity may partially be related to different distributions in age, sex and CVD phenotypes across study cohorts. Experiment conditions for blood draws, DNA extraction, storage and other unobserved confounding factors may also have contributed to the heterogeneity. For the MVMR analysis, we only included LDL which is known to associate with mtDNA CN and play a causal role in CHD to investigate the direct causal effects of LDL and CHD on mtDNA CN. Other risk factors, e.g., triglycerides and diabetes, were not considered in MVMR. Therefore, horizontal pleiotropy was not fully accounted for, which was also reflected by a large test statistic in testing heterogeneity in MVMR (Table 1). However, MR-PRESSO is currently not available to correct for pleiotropic effects in MVMR.

**Strength of the study**

The main strength of this study is that we adopted bi-directional and multivariable MR analysis to disentangle the complex relationship between mtDNA CN and CHD in cohort
studies. Observational epidemiological studies are susceptible to confounding and subclinical disease stage may impact observed associations between CVD outcomes and mtDNA CN. Robust genetic variants have been identified in large GWAS with mtDNA CN (n = 465,809), CHD (n = 361,194) and LDL (n=1,166,583). To minimize bias in MR analyses, we removed known pleiotropic SNPs (e.g., APOE SNPs) that are associated with both mtDNA CN and CVD traits. We performed MR IVW as well as sensitivity analyses including MR-Egger, Median and Mode methods to provide evidence for validity of MR estimators. Benefiting from the widely available GWAS data, we were able to utilize non-overlapping samples for GWAS of each exposure, which strengthened the validity of the two-sample MVMR analysis. Overall, the multivariable MR results provide evidence that higher LDL level is the driving factor for the association between lower mtDNA CN and CHD. An additional strength is that, in most of these cohorts, the CVD outcomes and all-cause death data, have been regularly collected for adjudicated by a physician endpoints review committee. The well-characterized outcome and predictor variables and hierarchical association analyses with three models to reduce potential confounding.

In summary, this study validated the previously reported association of mtDNA CN with CVD outcomes and all-cause mortality. In addition, we used both univariable and multivariable MR analyses to demonstrate an independent causal effect of LDL underlying both mtDNA CN and CHD, even after accounting for other risk factors. Findings from this study add to an increasing volume of evidence surrounding the harmful effects of high LDL in the complex relationships between vascular inflammation, atherosclerosis, lower mtDNA CN, and mitochondrial dysfunction. Therefore, control for
LDL and inflammation may be a feasible therapeutic strategy to improve mitochondrial function and cardiovascular health.

Acknowledgments

We included detailed acknowledgment for each cohort in Supplemental Materials. In brief, the authors thank the staff and participants of the Atherosclerosis Risk in Communities study, Cardiovascular Health Study, Coronary Artery Risk Development in Young Adults Study, Framingham Heart Study, Jackson Heart Study, Genetic Epidemiology Network of Arteriopathy Study, Multi-Ethnic Study of Atherosclerosis Study, and Women’s Health Initiative. We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. 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 (R01HL-117626-02S1; contract HHSN268201800002I). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393-02S1; contract HHSN268201800001I). Additional phenotype harmonization was performed by the current study (R01AG059727). 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.
References

18. Mengel-From J, Thinggaard M, Dalgard C, Kyvik KO, Christensen K, Christiansen L. Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. Hum Genet 2014;133:1149-59.


Figure 1. Study design. Association analysis of mtDNA cop number (CN) with cardiovascular disease traits was performed in eight cohorts of total 27,316 participants of multiple races/ethnicities. Meta-analysis was performed using the fixed effects inverse variance method to summarize the cohort specific results. Bi-directional univariable Mendelian Randomization was performed to test causality between mtDNA CN, CHD and LDL. Multivariable MR was performed to test the direct causal effect of LDL and CHD on mtDNA CN.
Figure 2. Association and meta-analysis of mtDNA CN and prevalent CVD outcomes. We performed logistic regression with an outcome and mtDNA residuals as independent variable adjusting for age, sex, study center (if applicable), and race/ethnicity in Model1. The size of the square represent the weight of each cohort. ARIC, Atherosclerosis Risk in Communities study; CARDIA, Coronary Artery Risk Development in Young Adults Study; CHS, Cardiovascular Health Study; FHS, Framingham Heart Study; GENOA, Genetic Epidemiology Network of Arteriopathy Study; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; WHI, Women’s Health initiative.
### Figure 3. Association and meta-analysis of mtDNA CN and incident CVD outcomes.

We performed Cox proportional hazards regression with an outcome and mtDNA residuals as independent variable adjusting for age, sex, study center (if applicable), and race/ethnicity in Model 1. The size of the square represent the weight of each cohort. ARIC, Atherosclerosis Risk in Communities study; CARDIA, Coronary Artery Risk Development in Young Adults Study; CHS, Cardiovascular Health Study; FHS, Framingham Heart Study; GENOA, Genetic Epidemiology Network of Arteriopathy Study; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; WHI, Women’s Health initiative.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>N</th>
<th>Estimate [95% CI]</th>
<th>Cohort</th>
<th>N</th>
<th>Estimate [95% CI]</th>
<th>Cohort</th>
<th>N</th>
<th>Estimate [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD</td>
<td></td>
<td></td>
<td>Stroke</td>
<td></td>
<td></td>
<td>CVD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIC</td>
<td>6662042</td>
<td>1.60 [1.45, 1.76]</td>
<td>ARIC</td>
<td>2143014</td>
<td>1.60 [1.37, 1.87]</td>
<td>ARIC</td>
<td>10002982</td>
<td>1.48 [1.37, 1.60]</td>
</tr>
<tr>
<td>CHS</td>
<td>8623232</td>
<td>1.01 [0.94, 1.08]</td>
<td>CHS</td>
<td>7393890</td>
<td>1.07 [0.99, 1.15]</td>
<td>CHS</td>
<td>13623108</td>
<td>1.04 [0.96, 1.09]</td>
</tr>
<tr>
<td>FHS</td>
<td>2302616</td>
<td>0.90 [0.77, 1.04]</td>
<td>FHS</td>
<td>1832502</td>
<td>0.67 [0.56, 0.79]</td>
<td>FHS</td>
<td>4392583</td>
<td>0.89 [0.70, 1.09]</td>
</tr>
<tr>
<td>GENOA</td>
<td>15476</td>
<td>1.20 [0.90, 2.19]</td>
<td>GENOA</td>
<td>17459</td>
<td>0.65 [0.43, 0.96]</td>
<td>GENOA</td>
<td>17459</td>
<td>1.40 [1.20, 2.63]</td>
</tr>
<tr>
<td>JHS</td>
<td>1682952</td>
<td>0.99 [0.80, 1.27]</td>
<td>JHS</td>
<td>1473149</td>
<td>1.13 [0.99, 1.30]</td>
<td>JHS</td>
<td>4782271</td>
<td>1.11 [0.99, 1.24]</td>
</tr>
<tr>
<td>MESA</td>
<td>3164593</td>
<td>1.02 [0.91, 1.14]</td>
<td>MESA</td>
<td>2194509</td>
<td>1.29 [1.04, 1.60]</td>
<td>MESA</td>
<td>6604593</td>
<td>1.01 [0.94, 1.10]</td>
</tr>
<tr>
<td>WHI</td>
<td>17218651</td>
<td>1.00 [0.93, 1.07]</td>
<td>WHI</td>
<td>37007050</td>
<td>0.83 [0.79, 0.88]</td>
<td>WHI</td>
<td>46256687</td>
<td>0.97 [0.91, 1.03]</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>39752372</td>
<td>1.08 [0.91, 1.12]</td>
<td>SUMMARY</td>
<td>52092401</td>
<td>0.97 [0.94, 1.01]</td>
<td>SUMMARY</td>
<td>858023583</td>
<td>1.07 [1.03, 1.10]</td>
</tr>
</tbody>
</table>

**Hazard ratio**
Figure 4. Association and meta-analysis of mtDNA CN and all-cause mortality. We performed Cox proportional hazards regression with an outcome and mtDNA residuals as independent variable adjusting for age, sex, study center (if applicable), and race/ethnicity in Model 1. The size of the square represent the weight of each cohort. ARIC, Atherosclerosis Risk in Communities study; CARDIA, Coronary Artery Risk Development in Young Adults Study; CHS, Cardiovascular Health Study; FHS, Framingham Heart Study; GENOA, Genetic Epidemiology Network of Arteriopathy Study; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; WHI, Women’s Health initiative.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>N</th>
<th>Estimate [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIC</td>
<td>897/2611</td>
<td>1.12 [1.03, 1.21]</td>
</tr>
<tr>
<td>CHS</td>
<td>2399/2891</td>
<td>1.09 [1.05, 1.14]</td>
</tr>
<tr>
<td>FHS</td>
<td>600/3057</td>
<td>1.01 [0.92, 1.10]</td>
</tr>
<tr>
<td>JHS</td>
<td>308/2360</td>
<td>1.01 [0.98, 1.15]</td>
</tr>
<tr>
<td>MESA</td>
<td>590/4085</td>
<td>1.01 [0.93, 1.10]</td>
</tr>
<tr>
<td>WHI</td>
<td>3229/6498</td>
<td>1.02 [0.97, 1.10]</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>8018/21272</td>
<td>1.06 [1.03, 1.09]</td>
</tr>
</tbody>
</table>
Figure 5. The Causal relationships between LDL, CHD and mtDNA CN. A. Mendelian randomization. B. Biological experiment (adapted from Lee at al.³). LDL: LDL cholesterol. CHD: coronary heart disease. mtDNA CN: mitochondrial DNA copy number. OxLDL: oxidized LDL. MPTP: mitochondrial permeability transition pore. ROS: reactive oxygen species. Univariable MR analysis demonstrated that one s.d. lower in genetically predicted LDL was causally associated with 0.99 hazard ratio of CHD. Multivariable MR analysis demonstrated the direct effect of one s.d. lower in genetically predicted LDL was causally associated with 0.092 s.d. lower in mtDNA CN (P=3.2E-09), and the direct effect of genetically predicted CHD condition was not associated with mtDNA CN (beta = -0.039, P=0.92).
Table 1. Comparison of results between Multivariable and univariable Mendelian randomization

<table>
<thead>
<tr>
<th>Exposure</th>
<th>N SNPs</th>
<th>Beta</th>
<th>Se</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariable MR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHD</td>
<td>5</td>
<td>-0.10</td>
<td>0.34</td>
<td>0.76</td>
</tr>
<tr>
<td>LDL</td>
<td>341</td>
<td>0.083</td>
<td>0.012</td>
<td>1.2E-12</td>
</tr>
<tr>
<td>Univariable MR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHD</td>
<td>10</td>
<td>-0.86</td>
<td>0.48</td>
<td>0.072</td>
</tr>
<tr>
<td>LDL</td>
<td>345</td>
<td>0.084</td>
<td>0.011</td>
<td>1.1E-14</td>
</tr>
</tbody>
</table>

Multivariable Mendelian randomization (MR) was performed for CHD and LDL on mtDNA CN. Univariable MR was performed for CHD on mtDNA CN, and LDL on mtDNA CN (IVW results were presented for univariable MR analyses). Q statistic for instrument strength of CHD is 1035, LDL is 3858. High heterogeneity was detected in multivariable MR IVW (Q value=1380, d.f.=340, p = 4.5e-125)