Low leukocyte mitochondrial DNA abundance drives atherosclerotic cardiovascular diseases: a cohort and Mendelian randomization study

Luo, J.; Noordam, R.; Jukema, J.W.; Dijk, K.W. van; Hagg, S.; Grassmann, F.; ... ; Heemst, D. van

ESC European Society of Cardiology

Cardiovascular Research (2023) **119**, 998–1007
https://doi.org/10.1093/cvr/cvac182

Low leukocyte mitochondrial DNA abundance drives atherosclerotic cardiovascular diseases: a cohort and Mendelian randomization study

Jiao Luo [1,2]*, Raymond Noordam [1], J. Wouter Jukema[3,4], Ko Willems van Dijk[5,6], Sara Hägg[7], Felix Grassmann[7,8], Saskia le Cessie [1,9], and Diana van Heemst[1]

[1]Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands; [2]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; [3]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands; [4]Netherlands Heart Institute, Utrecht, The Netherlands; [5]Department of Internal Medicine, Division of Endocrinology, Leiden University Medical Center, Leiden, The Netherlands; [6]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; [7]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [8]Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK; and [9]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

Received 31 January 2022; revised 7 October 2022; accepted 21 October 2022; online publish-ahead-of-print 20 December 2022

Aim

Mitochondrial DNA dysfunction has been implicated in the pathogenesis of cardiovascular diseases. We aimed to investigate the associations between leukocyte mitochondrial DNA (mtDNA) abundance, as a proxy of mitochondrial function, and coronary artery disease (CAD) and heart failure (HF) in a cohort study and approximate the causal nature of these relationships using Mendelian randomization (MR) in genetic studies.

Methods and results

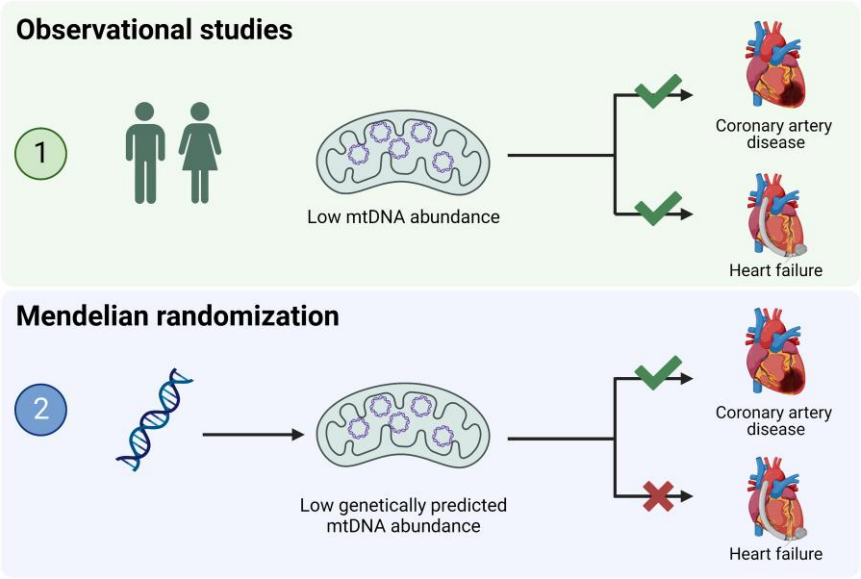
Multivariable-adjusted Cox regression analyses were conducted in 273 619 unrelated participants of European ancestry from the UK Biobank (UKB). For genetic studies, we first performed MR analyses with individual-level data from the UKB using a weighted genetic risk score (GRS); two-sample MR analyses were subsequently performed using summary-level data from the publicly available three consortia/biobank for CAD and two for HF. MR analyses were performed per database separately and results were subsequently meta-analysed using fixed-effects models. During a median follow-up of 11.8 years, restricted cubic spline Cox regression analyses showed associations between lower mtDNA abundance and higher risk of CAD and HF. Hazard ratios for participants in the lowest quintile of mtDNA abundance compared with those in the highest quintile were 1.08 (95% confidence interval: 1.03, 1.14) and 1.15 (1.05, 1.24) for CAD and HF. Genetically, no evidence was observed for a possible non-linear causal effect using individual-level weighted genetic risk scores calculated in the UKB on the study outcomes; the pooled odds ratios (95% confidence interval) from two-sample MR of genetically predicted per one-SD decrease in mtDNA abundance were 1.09 (1.03, 1.16) for CAD and 0.99 (0.92, 1.08) for HF, respectively.

Conclusion

Our findings support a possible causal role of lower leukocyte mtDNA abundance in higher CAD risk, but not in HF.

* Corresponding author. Tel: +31 71 526 6640, Fax: +31 71 526 6881, E-mail: j.luo@lumc.nl, jiao.luo@regionh.dk

Graphical Abstract



Keywords Mitochondrial DNA • Coronary artery disease • Heart failure • Mendelian randomization

1. Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide. The heart is high oxygen-consuming, with large amounts of mitochondria constituting up to one-quarter of cardiomyocytes volume.[1] Mitochondrial dysfunction, a hallmark of the aging process,[2] leads to reduced bioenergetic capacity and disrupted redox homeostasis, and is therefore hypothesized to be a critical component in the pathogenesis of CVD.[3,4] Mitochondria have their own circular genome, the mitochondrial DNA (mtDNA), consisting of 37 genes, 13 of which encode proteins on the electron transport chain. Individual mitochondrion may contain several copies (CN) of the mitochondrial genome, known as mtDNA copy numbers. The mtDNA abundance is associated with bioenergetics, mitochondrial membrane potential, and oxidative stress,[5] and therefore could serve as a surrogate biomarker of mitochondrial dysfunction.[6] A better understanding on the role of mtDNA abundance may provide early opportunities in the prevention and treatment of CVD.

Recent epidemiological studies have assessed the associations between leukocyte mtDNA abundance using mtDNA-CN and multiple cardiovascular endpoints. These studies so far unequivocally indicated lower mtDNA-CN as an independent risk factor of prevalent CVD in case-control and retrospective cohort studies[7–10] and of incident CVD and risk of sudden cardiac death in the prospective Atherosclerosis Risk in Communities (ARIC) study.[11–14] Nevertheless, apart from the ARIC study, other studies comprised a small sample size and/or a limited number of cases, which might have resulted in insufficient statistical power. In a recent cross-sectional study integrating multiple studies, mtDNA abundance was associated with a cluster of cardiometabolic traits that increase the risk of CVD, including obesity, hypertension, and hyperlipidaemia.[15] However, due to the vague onset and long-term progression of CVD pathogenesis, it is not possible to fully eliminate reverse causation and residual confounding in studies with observational study designs. Whether these associations are causal, therefore, remains unclear.

Triangulation of causal inference in aetiological epidemiology has been proposed, which integrates results from different methodological approaches to enhance the reliability of a research study.[16] The confidence in the findings will be strengthened if results from different approaches are consistent with each other. Based on earlier studies, we hypothesized that a lower mtDNA abundance is associated with an increased risk of incident CVD. Consequently, we first examined the associations between mtDNA abundance and incidence of coronary artery disease (CAD) and heart failure (HF) in participants of European ancestry in the UK Biobank (UKB) using Cox proportional hazards regression models. Second, to test for possible non-linear effect of the exposure on the outcome, we conducted Mendelian randomization (MR) using a weighted genetic risk score from individual-level data in the UKB; subsequently, publicly available data were exploited to perform two-sample MR to investigate whether genetically predicted low mtDNA abundance were causally associated with increased risk of diseases.

2. Methods

2.1 Prospective study

2.1.1 Study population

The UKB cohort is a prospective cohort with 502 628 participants between the age of 40 and 69 years recruited from the general population at multiple assessment centres across the UK between 2006 and 2010.[17] More detailed information about the recruitment of participants is available in Supplementary methods. The UKB study was approved by the North-West Multi-center Research Ethics Committee (MREC) and conducted according to the Declaration of Helsinki. All participants provided written informed consent. We used genotype data from 488 377 individuals in the full genetic data release (July 2017) in the present study.

Participants who did not pass the sample quality control were initially excluded according to the parameters presented in the sample quality control file of the UKB, including participants (i) who were not used to compute principal components (PCs); (ii) who were identified as outliers in heterozygosity and missing rates, which is indicative of poor-quality genotype data for these samples; (iii) who were identified as putatively sex chromosome aneuploidy; (iv) whose sex inferred from genotype data did not match their self-reported

sex; and (v) who had an excessive number (more than 10) of relatives in the database. To minimize the variation resulting from population substructures, we restricted the study population to unrelated white British individuals. Participant relatedness was available in the UKB by estimating kinship coefficients for all pairs. White British ancestry was identified based on self-reported ethnic background, and the population definition was further refined in a principal component analysis of the genotype data that were tightly clustered as performed and provided by the UKB. This resulted in a primary study cohort comprising 302 685 unrelated European ancestry participants. A flowchart of the exclusions steps is provided in Supplementary material online, *Figure S1*.

2.1.2 mtDNA abundance computation

We computed somatic mtDNA abundance from the UKB participants from the intensities of genotyping probes on the mitochondrial chromosome on the Affymetrix Array. The method for computing mtDNA abundance has been described in detail earlier.[18] We followed the same pipeline for calculation in the available data of UKB (https://github.com/GrassmannLab/MT_UKB). In brief, the relative amount of mtDNA hybridized to the array at each probe was the log2 transformed ratio (L2R) of the observed genotyping probe intensity divided by the intensity at the same probe observed in a set of reference samples. The median L2R values across all 265 variants passing quality control on the MT chromosome were used as an initial raw measure of mtDNA abundance. To correct for the confounding induced by poorly performing probes, we weighted the L2R values of each probe multiplied by the weight of the probe that was generated from a multivariate linear regression model in which those intensities statistically significantly predicted normalized mitochondrial coverage from exome sequencing data, resulting in a single mtDNA abundance estimate for each individual. To eliminate the plate effect, we subsequently standardized the abundance to a mean of zero and a standard deviation (SD) of one within each genotyping plate comprising 96 wells. An additional quality control step was performed by eliminating individuals with high SD (two SD from the mean) of autosomal probes log2 ratio (L2R). Consequently, 293 245 individuals remained in the cohort and are further used in the subsequent observational analyses and individual-level data genetic study. Importantly, the mtDNA abundance has been shown to be highly consistent with the number of mtDNA-CN.[19,20]

2.1.3 Outcome definition

Outcomes in the analysis were incident cardiovascular diseases during the period from recruitment to 1 January 2021. Incident disease status was ascertained by linkage with hospital admissions data and national death register data to identify the date of the first known CVD or CVD-related death after the date of baseline assessment. The linkage details are presented in the original study protocol (https://www.ukbiobank.ac.uk/media/gnkeyh2q/study-rationale.pdf, accessed April 2021). The outcomes were incident CAD and HF, separately. Incident disease diagnoses are coded according to the International Classification of Diseases edition 10 (ICD-10); CAD cases were defined as angina pectoris (I20), myocardial infarction (MI) (I21 and I22), and acute and chronic ischaemic heart disease (IHD) (I24 and I25); Incident HF cases were defined as I50. In addition, we analysed acute MI and chronic IHD as separate outcomes in sensitivity analyses. Follow-up time is computed from the baseline visit to the diagnosis of incident disease, death, or censoring, whichever occurred first.

2.1.4 Covariates

Covariates were from baseline measurements, including demographic parameters (age at recruitment, sex, deprivation index); the first 10 PCs to correct for possible remaining population stratification; genotyping batch; cell numbers (white blood cell counts and platelet counts); anthropometric measure of body mass index (BMI) in kg/m^2; self-reported lifestyle factors [smoking status (never, past and current), alcohol consumption frequency (twice or less per week/more than three times per week), physical activity (MET hours per week for moderate-vigorous activity), sleep duration in hours and insomnia symptoms (yes/no)]; familial CVD history (yes/no), lipid levels (mmol/l) [total and LDL (low-density lipoprotein) cholesterol] lipid-lowering medication, blood pressure (mmHg, average of the two measurements taken a few moments apart when applicable), and blood pressure-lowering medication, as well as baseline type 2 diabetes mellitus (T2DM, yes/no) from the medical records.

2.1.5 Statistical analysis

After further exclusion of participants with any prevalent CVD or withdrawn informed consent, the study cohort comprised an analytic sample of 273 619 individuals (see Supplementary material online, *Figure S1*). The baseline characteristics of the study population were described in quintiles of mtDNA abundance and presented as mean (SD) or median (interquartile range, IQR) for continuous variables and frequency (proportion) for categoric variables. Cumulative incidence for competing risks (CICR) was used to plot the cumulative incidence of both CAD and HF against follow-up time by mtDNA abundance quintiles, accounting for death as a competing event.

Cox proportional hazards models were used to estimate hazard ratios (HRs) and corresponding 95% confidence intervals (CIs) for the association between mtDNA abundance and incident CAD and HF. Two multivariable-adjusted regression models were fitted: Model 1 was adjusted for age, sex, the first 10 PCs, genotyping batch, white blood cell count, and platelet count; Model 2 was additionally adjusted for BMI, smoking, alcohol consumption, sleep duration, insomnia, physical activity, familial CVD history, lipid levels and lipid-lowering medication, blood pressure, and blood pressure-lowering medication and T2DM. In the primary analysis, we treated mtDNA abundance as a continuous variable and assessed the risk of incident diseases associated with per one-SD decrease in mtDNA abundance using restricted cubic spline Cox regression, with knots located at 5th, 50th, and 95th percentiles. Subsequently, mtDNA abundance was categorized into quintiles, and HRs compared the 1st to 4th quintiles (lower mtDNA abundance) with the 5th quintile (reference, highest mtDNA abundance). The proportional hazard assumption was graphically assessed by plotting log(−log[survival]) vs. log(follow-up time) and was tested using Schoenfeld residuals.

Missing data were present in the covariates and were imputed using multiple imputation by chain equations (MICE),[21] setting the number of imputed datasets to 10. We used predictive mean matching for continuous variables, logistic regression for binary variables, and polytomous regression for categorical variables. The imputation model included mtDNA abundance, all covariates, the Nelson–Aalen estimator of cumulative hazard and incident disease status. Cox-proportional hazards models were fitted within each imputed dataset and were subsequently pooled according to Rubin's rules.

As sensitivity analyses, firstly, interaction terms between mtDNA abundance and age and sex were added to Model 2 to test for the presence of effect modification by sex or age. Subgroup analyses were also performed in each stratum of sex and age (<50 years, 50~60 years, > 60 years), respectively. Secondly, all analyses were performed for the CAD subtypes, i.e. MI and IHD. Thirdly, analyses were repeated restricting to participants without missing data on covariates, i.e. complete cases ($n = 162\,002$).

2.2 Mendelian randomization analyses

2.2.1 Instrumental variables

We retrieved 129 independent (linkage disequilibrium < 0.05) nuclear single-nucleotide polymorphisms (SNPs) on autosomes as genetic instruments that were associated with continuous mtDNA abundance at a genome-wide significance threshold ($P < 5 \times 10^{-08}$), as identified in a recent genome-wide association study (GWAS) by Longchamps *et al.*[22] The study was performed in 465 809 individuals of White European ancestry combining the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium and the UKB. Genetic associations were adjusted for age, sex, and covariates that were specific in each cohort, such as PCs, blood collection sites, family structure, and cell composition. To minimize horizontal pleiotropic effect, we thoroughly scrutinized the phenotypes associated with each SNP when available in Phenoscanner

V2 (see Supplementary material online, *Table S1*). We subsequently excluded 9 SNPs in total that were genome-wide significantly associated with CAD [rs7213347 (*SMG6*), rs142158911 (*LDLR*), rs7412 (*APOE*), rs2736100 (*TERT*)] or risk factors for CAD, predominantly blood pressure, lipids and lipoproteins [rs7800558 (*DENND2A*), rs7896518 (*JMJD1C*), rs261290 (*ALDH1A2*), rs289713 (*CETP*), rs4895441 (*HBS1L*), rs4841132 (*RP11-115J16.1*)] in European descendant participants. *F*-statistics [(β/S.E.)2] were computed to evaluate instrumental strength. Furthermore, we calculated the proportion of total variance in the exposure explained by each instrument (R^2) separately.[23]

2.2.2 Individual-level data from the UKB

We used 293 245 individuals with both prevalent and incident outcomes as illustrated in Supplementary material online, *Figure S1*. We calculated a genetic risk score (GRS) for each participant weighted by the associations of the genetic instrumental variables for mtDNA abundance identified from the previous step. To be aligned with the observational analyses, we created quintiles based on the GRS. MR estimates were obtained by dividing the GRS-outcome association by the GRS-mtDNA abundance, where logistic regression and linear regression were used for outcomes and cell counts, respectively. All regression models were adjusted for age at recruitment, sex, genotyping batch, and the first 10 PCs. Moreover, we also investigated the non-linearity of the exposure–outcome relationship using a piecewise linear method,[24] in which the population was divided into strata by instrumental variable-free exposure (i.e. residuals from mtDNA regressing on GRS), and a causal effect in each stratum was estimated, referred to as a localized average causal effect (LACE). We reported the *P*-values from two tests for non-linearity, including the quadratic test and the Cochran's *Q* test.

2.2.3 Summary-level outcome data source

Summary statistics for instrument-CAD associations were extracted from three large databases separately, the CARDIoGRAMplusC4D (Coronary Artery Disease Genome-Wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics from Nikpay *et al.*,[25] where UKB data were not included) consortium, UKB, and FinnGen study (freeze 5, released in May 2021). Similarly, summary statistics for SNP-HF associations were drawn from HERMES Consortium (Heart Failure Molecular Epidemiology for Therapeutic Targets Consortium, which included data from UKB in the meta-analysis) and the FinnGen study, respectively. The descriptions, number of cases and controls, cases definition, as well as covariates used for association tests of each of the databases, are presented in detail in Supplementary methods and Supplementary material online, *Table S2*.

2.2.4 Mendelian randomization analysis

SNP-exposure and SNP-outcome data were harmonized to make alignment on effect alleles. Palindromic SNPs were eliminated.[26] The primary MR analysis was performed using inverse-variance weighted (IVW) method to combine the SNP-specific estimates calculated using Wald ratios, assuming all instrumental variables are valid.[27] Results were expressed as an odds ratio (OR) on disease risk for a one-SD decrease in genetically predicted mtDNA abundance. When the MR assumptions were met, this OR approximated the causal effect of the exposure on the outcome. Sensitivity analyses accounting for pleiotropy were conducted, including weighted-median estimator (WME) and MR-Egger regression,[28,29] both of which assumed that at least half of the instrumental variable had to be valid. The intercept from MR-Egger represents the average pleiotropic effect; when the intercept deviates from zero, estimates from IVW might be biased. MR-PRESSO (MR Pleiotropy RESidual Sum and Outlier) was applied to detect and correct for horizontal pleiotropy through removing outliers.[30] Moreover, we examined the heterogeneity using Cochran's *Q* statistic among all SNPs within each outcome database.

For outcomes derived from the UKB, despite the gene-exposure associations being from the same population, it has been shown that two-sample MR methods can be reliably used for one-sample MR performed within large biobanks, such as UKB, except for the MR-Egger sensitivity analysis.[31]

2.2.5 Meta-analysis of estimates from different databases

The effects of mtDNA abundance on CAD/HF in MR analyses were separately estimated in different outcome databases, CARDIoGRAMplusC4D consortium (CAD) or HERMES (HF), UKB (CAD only), and FinnGen (both), and derived estimates were subsequently pooled using fixed-effects meta-analysis.

2.2.6 Sensitivity analysis

Furthermore, the selected SNPs from large GWAS may associate with other downstream traits of the trait of interest, such as effects directly on the outcome. The inclusion of SNPs with stronger associations to outcomes than to exposures could result in incorrect approximation of the causal effect due to reverse causation of those particular SNPs. Therefore, we applied MR Steiger filtering to test the direction of causality for each instrumental variable on exposure and outcome. Steiger filtering assumes that a valid instrumental variable should explain more variation in the exposure than in the outcome; if an instrumental variable meets the criterion, the causal direction of this instrument is 'TRUE', otherwise, it is 'FALSE',[32] indicating that this instrument may likely suffer from reverse causation. *P*-values for the inference of direction were also obtained. After removing those SNPs with the 'FALSE' causal direction, we repeated all MR analyses using the IVW method and meta-analysed using the fixed-effect model.

Despite the large sample size of the GWAS used for the selection of instrumental variables in the Longchamps *et al.* study, which increased the statistical power, the assessments of mtDNA abundance among cohorts that contributed data to the meta-analysis were very different. To account for this measurement heterogeneity, we additionally performed sensitivity analyses restricting to genetic instruments identified from the UKB only. Therefore, 66 independent (linkage disequilibrium < 0.1) SNPs were used that were associated with mtDNA abundance at a genome-wide significance threshold ($P < 5 \times 10^{-08}$) from 295 150 participants conducted by Hägg *et al.*[18] Genetic associations were adjusted for PCs, age, sex, genotyping batch, genotyping missingness/call rate, and cell composition. Similar to the main analyses, we retrieved all the phenotypes associated with the identified SNP (see Supplementary material online, *Table S3*), and 6 SNPs were excluded due to the association with either CAD or related risk factors. All MR analyses were repeated with the substitution of the 60 genetic instruments for mtDNA abundance.

All the analyses were performed using R (v3.6.3) statistical software (The R Foundation for Statistical Computing, Vienna, Austria). Packages used in the analyses included 'cmprsk' for cumulative incidence for competing risk analyses, 'mice' for multiple imputations, 'survival' and 'survminer' for Cox-proportional hazard regression, 'rms' for non-linear dose-response associations, 'TwoSampleMR' for MR analyses, and 'meta' for meta-analyses. All results were reported as HRs in observational analyses and ORs in MR analyses with accompanied 95% CIs.

3. Results

3.1 Prospective results

3.1.1 Main analyses

A total of 273 619 participants were eligible for analyses after exclusion. Compared with the highest quintile of mtDNA abundance (*Table 1*), participants in the lower quintiles were more likely to have unfavourable CVD risk factors, including older age, male sex, higher BMI, higher blood pressure and more blood pressure-lowering medication, higher lipids (total and LDL cholesterol) and more cholesterol-lowering medication, less physical activity, more current smokers, and a higher percentage of familial history of CVD or prevalent T2DM.

During a median follow-up of 11.8 (IQR: 11.0, 12.5) years, 18 346 participants developed CAD and 5795 participants developed HF. Cumulative incidence of both CAD and HF increased stepwise with the decrease in

Table 1 Baseline characteristics of the study participants by quintiles of mtDNA abundance

Variable	Q1 (lowest abundance)	Q2	Q3	Q4	Q1 (highest abundance)	*P*-value
n	54 724	54 724	54 723	54 724	54 724	—
Age (years)	57.2 (8.0)	56.8 (8.0)	56.5 (8.0)	56.2 (8.0)	55.9 (7.9)	<0.001
Sex						<0.001
Male	25 413 (46.4%)	24 898 (45.5%)	24 478 (44.7%)	24 320 (44.4%)	23 622 (43.2%)	
Female	29 311 (53.6%)	29 826 (54.5%)	30 245 (55.3%)	30 404 (55.6%)	31 102 (56.8%)	
BMI (kg/m^2)	27.6 (5.0)	27.4 (4.8)	27.3 (4.6)	27.1 (4.6)	26.9 (4.5)	<0.001
Deprivation index	−1.5 (2.9)	−1.6 (2.9)	−1.7 (2.9)	−1.7 (2.9)	−1.7 (2.9)	<0.001
Diastolic blood pressure (mmHg)	82.9 (10.2)	82.6 (10.0)	82.4 (10.0)	82.3 (10.1)	81.8 (10.1)	<0.001
Systolic blood pressure (mmHg)	139.5 (18.9)	138.7 (18.7)	138.2 (18.7)	137.6 (18.4)	136.6 (18.3)	<0.001
Blood pressure-lowering medication						<0.001
Yes	10 826 (19.8%)	10 085 (18.4%)	9598 (17.5%)	9089 (16.6%)	8524 (15.6%)	
No	43 898 (80.2%)	44 637 (81.6%)	45 128 (82.5%)	45 635 (83.4%)	46 199 (84.4%)	
Total cholesterol (mmol/L)	5.8 (1.1)	5.8 (1.1)	5.8 (1.1)	5.8 (1.1)	5.7 (1.1)	<0.001
HDL (mmol/L)	1.5 (0.4)	1.5 (0.4)	1.5 (0.4)	1.5 (0.4)	1.5 (0.4)	<0.001
LDL (mmol/L)	3.7 (0.9)	3.6 (0.8)	3.6 (0.9)	3.6 (0.8)	3.6 (0.8)	<0.001
Triglycerides (mmol/L)	1.8 (1.0)	1.8 (1.0)	1.7 (1.0)	1.7 (1.0)	1.7 (1.0)	<0.001
Cholesterol-lowering medication						<0.001
Yes	7596 (13.9%)	7394 (13.5%)	7148 (13.1%)	6822 (12.5%)	6663 (12.2%)	
No	47 128 (86.1%)	47 330 (86.5%)	47 575 (86.9%)	47 902 (87.5%)	48 061 (87.8%)	
Physical activity (moderate-vigorous MET hours/week)	26.5 (33.8)	27.0 (34.3)	27.1 (34.6)	27.1 (33.8)	27.1 (33.8)	0.054
Alcohol consumption frequency						<0.001
At least three times per week	24 939 (45.6%)	24 836 (45.4%)	25 303 (46.2%)	25 400 (46.4%)	25 679 (46.9%)	
Twice or less per week	29 751 (54.4%)	29 851 (54.5%)	29 381 (53.7%)	29 287 (53.5%)	29 015 (53.0%)	
Data missing	34 (0.1%)	37 (0.1%)	39 (0.1%)	37 (0.1%)	30 (0.1%)	
Smoking status						<0.001
Current	6124 (11.2%)	5643 (10.3%)	5369 (9.8)	5046 (9.2%)	4521 (8.3%)	
Previous	18 646 (34.1%)	18 831 (34.4%)	18 588 (34.0%)	18 586 (34.0%)	18 702 (34.2%)	
Never	29 732 (54.3%)	30 079 (55.0%)	30 594 (55.9%)	30 914 (56.5%)	31 350 (57.3%)	
Data missing	222 (0.4%)	171 (0.3%)	172 (0.3%)	178 (0.3%)	151 (0.3%)	
Sleep duration (hours)	7.1 (1.2)	7.1 (1.2)	7.1 (1.2)	7.1 (1.2)	7.1 (1.2)	0.6
Insomnia						0.02
Sometimes	26 161 (47.8%)	26 273 (48.0%)	26 420 (48.3%)	26 243 (48.0%)	26 321 (48.1%)	
Usually	15 488 (28.3%)	15 145 (27.7%)	15 031 (27.5%)	14 982 (27.4%)	14 991 (27.4%)	
Never/rarely	13 038 (23.8%)	13 271 (24.3%)	13 237 (24.2%)	13 454 (24.6%)	13 384 (24.5%)	
Data missing	37 (0.1%)	35 (0.1%)	35 (0.1%)	45 (0.1%)	28 (0.1%)	
Familial CVD history						<0.001
Yes	21 539 (39.4%)	21 716 (39.7%)	21 690 (39.6%)	21 687 (39.6%)	21 359 (39.0%)	
No	27 791 (50.8%)	27 822 (50.8%)	28 005 (51.2%)	28 026 (51.2%)	28 420 (51.9%)	
Data missing	5394 (9.8%)	5186 (9.5%)	5028 (9.2%)	5011 (9.2%)	4945 (9.0%)	
T2DM history (yes)						<0.001
Yes	1184 (2.2%)	1093 (2.0%)	969 (1.8%)	914 (1.7%)	879 (1.6%)	
No	53 540 (97.8%)	53 631 (98.0%)	53 755 (98.2%)	53 810 (98.3%)	53 845 (98.4%)	

BMI, Body mass index; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; CVD, cardiovascular disease; CAD, coronary artery disease; HF, heart failure. Data are mean (SD) or median (interquartile range, IQR) for continuous variables and frequency (percentage) for categorical variables. Some percentages do not add up to 100 because of rounding. *P*-values were obtained from *Kruskal–Wallis H* test or χ^2 test as appropriate.

mtDNA abundance, accounting for death as a competing risk (*P* for Gray's test < 0.001) (*Figure 1*). In multivariable-adjusted Cox-proportional hazard models, restricted cubic spline analyses showed an approximately linear dose-response relationship between lower mtDNA abundance with the higher risk of CAD (*P* for non-linearity = 0.1) and HF (*P* for non-linearity = 0.7), as shown in *Figure 2*. Continuously (model 1), a one-SD decrease in mtDNA abundance, was associated with a 1.06-fold (95% confidence interval, CI: 1.05, 1.08) and a 1.09-fold (95% CI: 1.06, 1.12) higher hazard of CAD and HF, respectively. Categorically, adjusted HRs for the first (lowest mtDNA abundance) vs. the fifth (reference, highest mtDNA abundance) quintile of mtDNA abundance were 1.18 (95% CI: 1.13, 1.24) for CAD and 1.28 (95% CI: 1.17, 1.39) for HF. Additional adjustment for CVD risk factors only minimally attenuated the estimates of CAD and HF (*Figure 3* and Supplementary material online, *Table S4*).

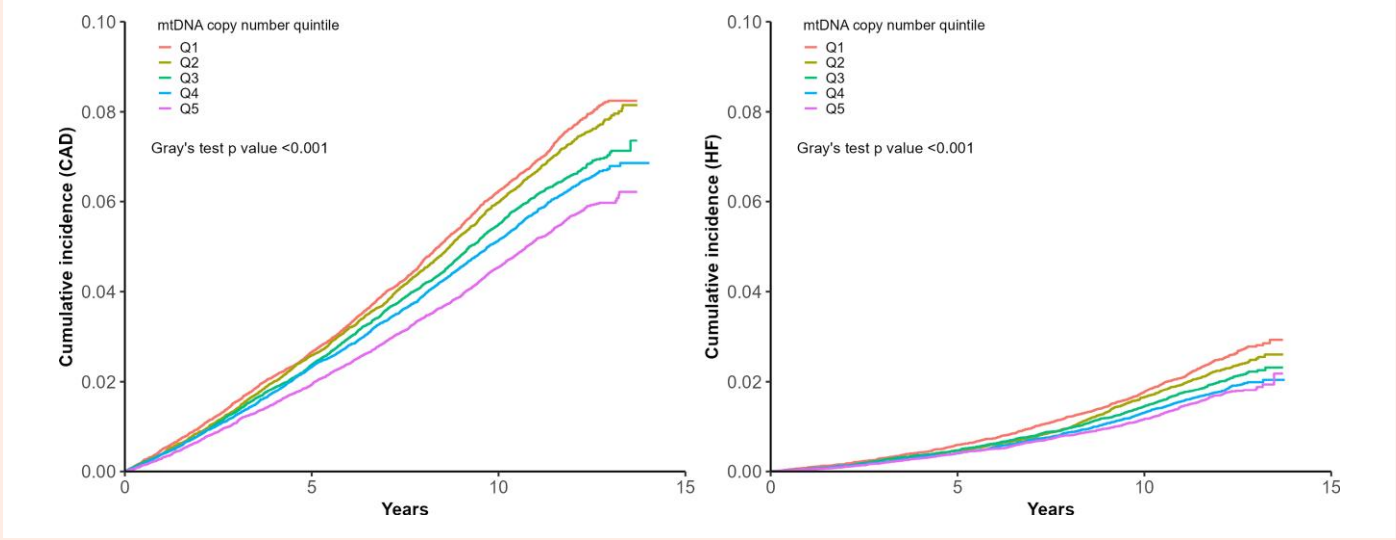


Figure 1 Cumulative incidence of CAD and HF by quintiles of mtDNA abundance. We calculated cumulative incidence for CAD and HF, accounting for death as a competing event. Differences in cumulative incidence between mtDNA abundance quintiles were assessed using Gray's test. CAD, coronary artery disease; HF, heart failure.

3.1.2 Sensitivity analysis

We observed no evidence favouring interactions between mtDNA abundance and sex (*P*-values for interaction = 0.2 for CAD, 0.7 for HF); in line, in sex-stratified analyses, the estimates between men and women were similar (see Supplementary material online, *Table S5*). Interaction was observed between mtDNA abundance and the baseline age for CAD (*P* for interaction < 0.001). After stratification by age groups, HRs obtained from model 2 for CAD slightly attenuated from the younger group (<50 years) to older groups (50~60 years and >60 years) (HR: 1.06, 1.04, and 1.02, respectively) (see Supplementary material online, *Table S6*). However, no interaction was detected between mtDNA abundance and age at baseline for HF (*P* for interaction = 0.2); though HR in the younger group was also higher for HF, this may be due to the very limited number of cases in this group. When analyses were conducted for MI and IHD separately, cumulative incidences were higher in lower quintiles compared with the highest quintile (see Supplementary material online, *Figure S2*) for MI and IHD; estimates from Cox proportional hazard regression models did not differ considerably from when all CAD were considered (see Supplementary material online, *Tables S4–S6*).

In addition, missing data in covariates were present (see Supplementary material online, *Table S7*), and 162 002 (59%) of 273 619 individuals included in the current study provided complete data for all variables. The absolute difference in the baseline characteristics between these participants with and without complete data was very limited (see Supplementary material online, *Table S8*). Furthermore, the main results from sensitivity analyses restricting to complete cases did not materially differ from the results obtained after imputation (see Supplementary material online, *Table S9*).

3.2 Mendelian randomization

3.2.1 Individual-level data

In the UKB, lower weighted GRS was continuously associated with a stepwise lower mtDNA abundance (see Supplementary material online, *Figure S3*), and was also associated with a higher risk of CAD [OR (95% CI): 1.18 (1.09, 1.28)] but not with the risk of HF [OR (95%CI): 1.07 (0.92, 1.24)]; categorically, compared with the highest GRS quintile (highest mtDNA abundance), the ORs increased according to quintiles, with significant estimates for CAD in the 1st and 2nd quintiles (lower mtDNA abundance) [OR (95%CI): 1.05 (1.02, 1.09) and 1.07 (1.03, 1.11)] (see Supplementary material online, *Table S10*). MR estimates [OR (95%CI)] using the ratio method were 1.25 (1.12, 1.38) and 1.09 (0.89, 1.34) for CAD and HF, respectively. We did not observe any non-linear association across different quintiles or from the piecewise method for both CAD and HF by the non-linear test (see Supplementary material online, *Figure S4*).

3.2.2 Summary-level data analyses

In total, 101 distinct SNPs were present in at least one of the outcome databases. *F*-statistics for each SNP were higher than 10 and ranged from 16 to 634, and a total of 1.8% variation were explained by the instruments (see Supplementary material online, *Table S11*).

For CAD, the pooled OR of the primary IVW estimates from CARDIoGRAMplusC4D, UKB, and FinnGen of a one-SD decrease in mtDNA abundance was 1.09 (95% CI: 1.03, 1.16) (*Figure 4*). Estimates from WME and MR-Egger generally did not differ substantially except for UKB where the point estimates attenuated to some extent. No pleiotropy was detected by the intercept of MR-Egger ($P > 0.05$). Though outliers were identified by MR-PRESSO in each database, estimates after outlier removal remained similar to those obtained from IVW (see Supplementary material online, *Table S12*).

For HF, the combined OR from IVW obtained in the HERMES consortium and FinnGen per one-SD decrease in mtDNA abundance was 0.99 (95% CI: 0.92, 1.08) (*Figure 4*). Results from WME were similar, and we observed no evidence for horizontal pleiotropy from MR-Egger intercept ($P > 0.05$); outliers were spotted in the HERMES consortium assessed by MR-PRESSO, but outlier-corrected estimates did not materially differ from those generated from IVW (see Supplementary material online, *Table S13*).

3.2.3 Sensitivity analyses

Steiger filtering detected several SNPs across different outcome datasets with possible reverse causation, and these were likely to be primarily associated with the outcomes rather than mtDNA abundance (see Supplementary material online, *Table S11*). After removing these SNPs, IVW analyses were repeated and study-specific estimates were meta-analysed; results from these efforts and estimates did not substantially differ with the overall analysis. Although the individual estimate became insignificant due to fewer instrumental variables, the meta-analysis of three estimates remained significant [OR (95%CI): 1.06 (1.00, 1.12)] for CAD and remain null [OR (95%CI): 0.99 (0.92, 1.08)] for HF (see Supplementary material online, *Figure S5*).

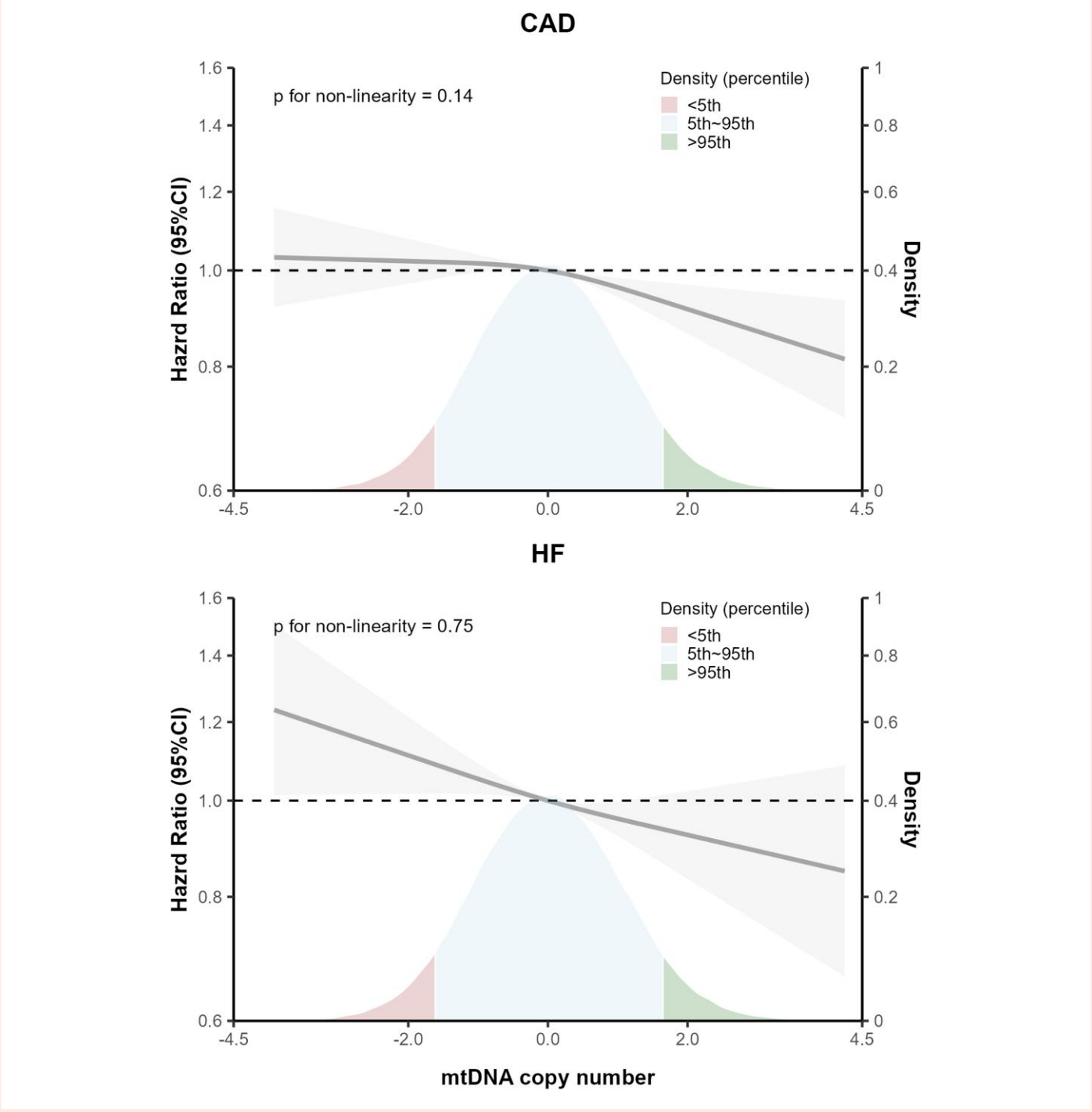


Figure 2 Hazard ratios for incident CAD and HF by levels of mtDNA abundance. Solid lines represented hazard ratios (derived from model 2 adjusted for age, sex, genotyping batch, the first two principal components, white blood cell count and platelet count, body mass index, physical activity, smoking status, alcohol consumption frequency, blood pressure and blood pressure-lowering medication, cholesterol, triglycerides and lipid-lowering medication, sleep duration and insomnia, type 2 diabetes status, and familial history of cardiovascular disease) and corresponding 95% confidence intervals (grey shadowed area) using restricted cubic splines for mtDNA abundance with knots at distribution of 5th, 50th, and 95th percentiles. The density on the right *y*-axis shows the distribution of baseline participants. Since mtDNA abundance has been standardized during computation, the distribution is approximately normal. CAD, coronary artery disease; HF, heart failure.

When we used genetic instrumental variables from Hägg *et al.*,[18] 59 distinct SNPs were included. *F*-statistics for each SNP were higher than 10 and ranged from 30 to 277, and a total of 1.1% variation were explained by the instruments. Detailed full information on the used genetic instruments is presented in Supplementary material online, *Table S14*. A one-SD decrease in mtDNA abundance was associated with a 1.16-fold (95% CI: 1.07, 1.26), a 1.04-fold (95% CI: 0.94, 1.15) higher risk of CAD and HF in the meta-analyses, respectively (see Supplementary material

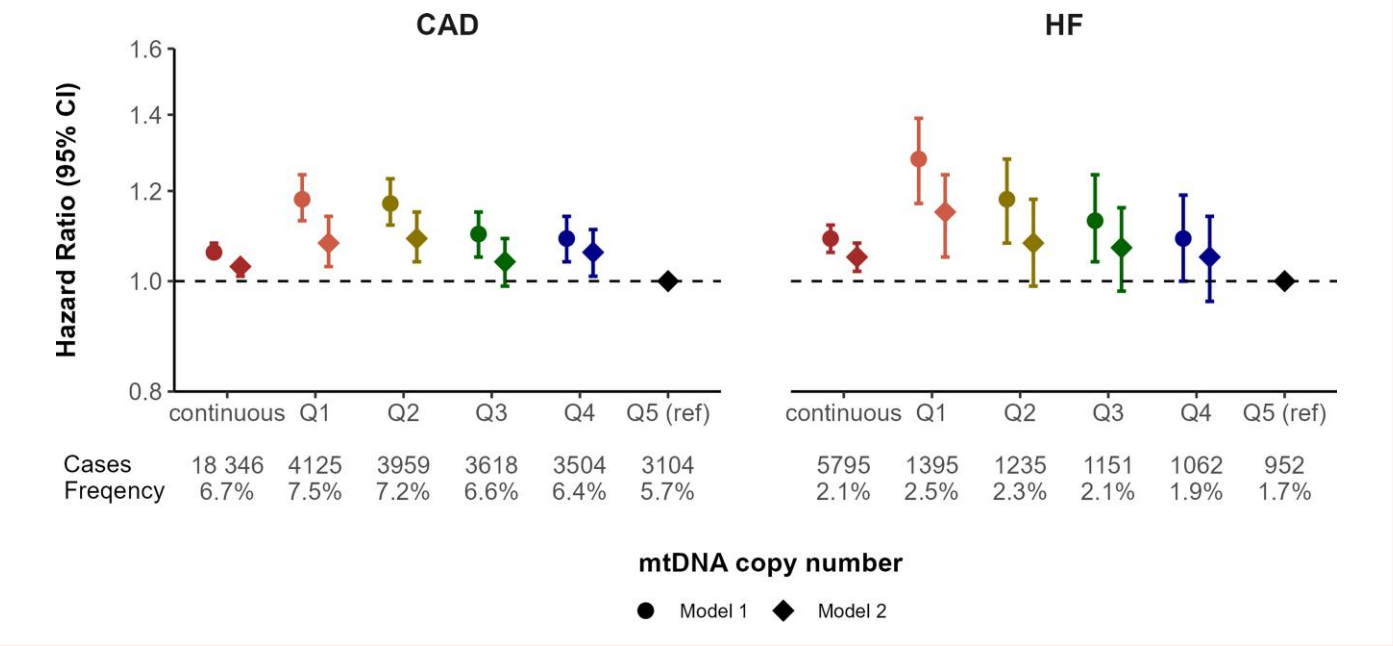


Figure 3 Hazard ratios for incident CAD and HF by quintiles of mtDNA abundance. Estimated hazard ratios for the effect of per-SD decrease in mtDNA abundance (continuous), or for the 1st (lowest mtDNA abundance) to the 4th quintile compared with the 5th (reference, highest mtDNA abundance) quintile (categorical) on CAD and HF. Model 1 was adjusted for age, sex, genotyping batch, the first two principal components, white blood cell count, and platelet count. Model 2 was model 1 additionally adjusted for body mass index, physical activity, smoking status, alcohol consumption frequency, blood pressure and blood pressure-lowering medication, cholesterol, triglycerides and lipid-lowering medication, sleep duration and insomnia, type 2 diabetes status, and familial history of cardiovascular disease. CAD, coronary artery disease; HF, heart failure.

online, *Figure S6*). MR sensitivity analyses including WME, MR-Egger, and MR-PRESSO are presented in Supplementary material online, *Tables S15* and *S16*.

4. Discussion

In the present study, we implemented a prospective cohort study design and MR study to assess the relationship of leukocyte mtDNA abundance with the risk of incident CAD and HF. Results from the multivariable-adjusted prospective analyses suggested associations between lower mtDNA abundance and higher risks of CAD and HF, whereas findings from MR analyses using either individual-level and summary-level data only confirmed an association between genetically influenced lower mtDNA abundance and a higher risk of CAD, possibly reflecting evidence of causality for CAD.

Consistent with our observational findings, previous studies showed that lower mtDNA-CN measured from leukocyte was related to an increased risk of CVD and its risk factors.[7–15] The only prospective study that assessed the relationship between mtDNA content and either CAD or HF used the ARIC study.[12,13] In the ARIC study, composed of 20 163 participants (2460 incident CAD) with a mean follow-up of 13.5 years, a lower mtDNA abundance was associated with an increased risk of incident CAD. Similarly, with 10 802 participants (2227 incident HF cases) followed-up for a mean of 23.1 years, lower mtDNA abundance was linked to an increased risk of HF. Residual confounding, in particular factors relevant to both mitochondrial function and CVD such as physical activity and insomnia, was not taken into account. However, in our multivariable-adjusted analysis, additional adjustment for these covariates did not further attenuate the estimates substantially.

To the best of our knowledge, the current study is the first to evaluate the causal nature of the associations between mtDNA abundance and the risk of CVD. MR analyses with the genetic instruments for mtDNA abundance confirmed the detrimental effect of lower mtDNA content on the risk of CAD observed in the cohort studies. Mitochondrial dysfunction, indicated by low mtDNA abundance, would lead to increased production of reactive oxygen species (ROS) in mitochondria.[5] Those maladaptive overproduced mitochondrial ROS mediate irreversible damage to macromolecules, such as increased oxidation of LDL and dysfunction of endothelial cells that are critical factors to promote atherosclerosis, and further CAD events.[33] Nevertheless, several factors merit thoughtful consideration in terms of the interpretation of the null effect on HF in MR analyses. HF has substantial phenotypic heterogeneity, which can be defined by ejection fraction (EF) and diastolic function; more than half of patients have preserved EF while over 40% of cases have isolated diastolic dysfunction.[34] Moreover, a large degree of variation has been described even within patients with preserved EF.[35,36] It has also been shown previously that the associations between mtDNA abundance and HF with preserved and reduced EF were different and possibly would make the association in the direction to zero when we combined the two subgroups in a single analysis.[13] However, stratification by cause of HF in the UKB ended up with a low number of cases and insufficient statistical power and cause-specific GWAS summary-level data of HF are currently not available. For these reasons, the lack of a clear association between mtDNA abundance and HF should be interpreted with caution, and more follow-up analyses are required to investigate the cause-specific HF in more detail.

4.1 Study strengths and limitations

The main strength of our study is that we adopted the triangulation of causal inference in aetiological epidemiology.[16] The consistency between biochemically measured and genetically determined mtDNA abundance with CAD increased the validity of the results. Given the absence of randomized clinical trials concerning mtDNA abundance and CAD to date, the analyses that have been performed in the present study provide the foremost evidence on the association between mtDNA content and CAD. Other important strengths of our prospective cohort study include the large sample

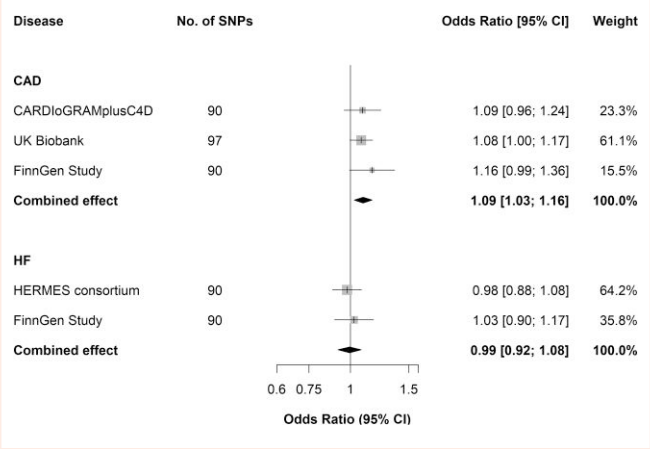


Figure 4 Mendelian randomization study of mtDNA abundance on the risk of CAD and HF. Estimated ORs for the effect of per-SD decrease in mtDNA abundance on CAD and HF, obtained from an MR inverse-variance weighted method, per outcome database separately and combined over the different databases using fixed-effect meta-analyses. CARDIoGRAMplusC4D, Coronary Artery Disease Genome-Wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics; HERMES consortium, Heart Failure Molecular Epidemiology for Therapeutic Targets consortium. UK Biobank data of heart failure was already integrated into HERMES consortium.

size and the considerable number of incident cases from the UKB, comprehensive assessment of confounding factors, and subtype analyses of MI and IHD within CAD. In MR studies, we meta-analysed three large databases where SNP-outcome associations were derived, comprising a substantial size of overall participants and cases. The results wereconsistent across different databases, and the precision of the pooled MR estimates obtained from different databases increased significantly.

Several limitations should be acknowledged. First, mtDNA content was measured in leukocytes. Without experimental confirmation in specific target tissues (e.g. heart, endothelium) that are relevant in the pathogenesis of CVD, our results therefore cannot be applied to indicate mtDNA abundance in cardiac or vascular tissues, which are highly physiologically distinct from cells circulating in the blood. More studies are needed to address the correlation between blood- and tissue-derived mtDNA abundance with larger sample sizes. In addition, the initial calculation of mtDNA abundance from chip arrays might have introduced noise due to the small number of variants. To this end, a weighted mtDNA abundance was implemented, which approximates what would be estimated from exome sequencing and has been validated.[18] Second, in MR analyses in the UKB, the GRS for mtDNA was generated based on the genetic associations identified from a combination of the CHARGE consortium and the UKB, and thus, the GRS is not completely externally weighted. Nevertheless, with the available data, this has been the utmost effort; in the two-sample MR, despite a large number of instrumental variables the variation of mtDNA abundance explained by these SNPs was small. Notwithstanding, we had more than sufficient power to detect the true causal effect in MR analyses (see Supplementary material online, *Figure S7*). Moreover, while we acknowledge the possibility of pleiotropic effects of included genetic instruments, this is likely to be vertical (Supplementary discussion). When we excluded genetic variants mapped in lipid-metabolism-related genes, and additionally stringently remove a large proportion of genetic variants with possible reverse causation detected by Steiger filtering test, the results remained similar. Third, since the population of non-Europeans was highly heterogeneous in UKB, we restricted the individual observational and genetic analyses to White European populations; furthermore, two-sample MR analyses were also performed predominantly in European-descent individuals. It is therefore inappropriate to extrapolate our findings to other populations with different ethnic backgrounds. Lastly, we were not able to dissect the potential impact of other mtDNA alterations, such as mtDNA mutations or deletions which have been proposed to contribute to the initiation and progression of atherosclerosis.[37] Consequently, there is a need for accurate deep sequencing to simultaneously analyse the entire mitochondrial genome to better understand the relationships between mtDNA abundance function, germline and acquired mutations, and CVD.

5. Conclusions

This study provides the first evidence of a possible causal association between leukocyte mtDNA abundance and the risk of CAD. Further studies are required to fully understand how mtDNA affects atherogenic risk development.

Supplementary material

Supplementary material is available at *Cardiovascular Research* online.

Acknowledgements

The present study has been conducted using the UK Biobank Resource (Application Number 56340) that is available to researchers. The authors acknowledge the participants and investigators of all consortia that contributed summary statistics data, including the UK Biobank, CARDIoGRAMplusC4D, HERMES consortium, and the FinnGen study.

Conflict of interest: None declared.

Funding

This work was supported by the VELUX Stiftung (grant number 1156) (to D.H. and R.N.). J.L. was supported by the China Scholarship Council (No. 201808500155). R.N. was supported by an innovation grant from the Dutch Heart Foundation (grant number 2019T103). S.H. is supported by the Karolinska Institutet Strategic Research Area in Epidemiology and the Swedish Research Council (2019-01272).

Data availability

Data used in the multivariable-adjusted analyses will be made available upon request in adherence with transparency conventions in medical research and through reasonable requests to the corresponding author. Data used in the Mendelian randomization analyses are all publicly available provided in the article or via a corresponding consortium.

References

1. Schaper J, Meiser E, Stämmler G. Ultrastructural morphometric analysis of myocardium from dogs, rats, hamsters, mice, and from human hearts. *Circ Res* 1985;**56**:377–391.
2. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell* 2013;**153**:1194–1217.
3. Castellani CA, Longchamps RJ, Sun J, Guallar E, Arking DE. Thinking outside the nucleus: mitochondrial DNA copy number in health and disease. *Mitochondrion* 2020;**53**:214–223.
4. Dai DF, Rabinovitch PS, Ungvari Z. Mitochondria and cardiovascular aging. *Circ Res* 2012;**110**:1109–1124.
5. Wallace DC. A mitochondrial bioenergetic etiology of disease. *J Clin Invest* 2013;**123**:1405–1412.
6. Malik AN, Czajka A. Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction? *Mitochondrion* 2013;**13**:481–492.
7. Wei R, Ni Y, Bazeley P, Grandhi S, Wang J, Li ST, Hazen SL, Wilson Tang WH, LaFramboise T. Mitochondrial DNA content is linked to cardiovascular disease patient phenotypes. *J Am Heart Assoc* 2021;**10**:e018776.
8. Chen S, Xie X, Wang Y, Gao Y, Xie X, Yang J, Ye J. Association between leukocyte mitochondrial DNA content and risk of coronary heart disease: a case-control study. *Atherosclerosis* 2014;**237**:220–226.
9. Wang XB, Cui NH, Zhang S, Liu ZJ, Ma JF, Ming L. Leukocyte telomere length, mitochondrial DNA copy number, and coronary artery disease risk and severity: a two-stage case-control study of 3064 Chinese subjects. *Atherosclerosis* 2019;**284**:165–172.

10. Koller A, Fazzini F, Lamina C, Rantner B, Kollerits B, Stadler M, Klein-Weigel P, Fraedrich G, Kronenberg F. Mitochondrial DNA copy number is associated with all-cause mortality and cardiovascular events in patients with peripheral arterial disease. *J Intern Med* 2020;**287**:569–579.
11. Zhang Y, Guallar E, Ashar FN, Longchamps RJ, Castellani CA, Lane J, Grove ML, Coresh J, Sotoodehnia N, Ilkhanoff L, Boerwinkle E, Pankratz N, Arking DE. Association between mitochondrial DNA copy number and sudden cardiac death: findings from the Atherosclerosis Risk in Communities study (ARIC). *Eur Heart J* 2017;**38**:3443–3448.
12. Ashar FN, Zhang Y, Longchamps RJ, Lane J, Moes A, Grove ML, Mychaleckyj JC, Taylor KD, Coresh J, Rotter JI, Boerwinkle E, Pankratz N, Guallar E, Arking DE. Association of mitochondrial DNA copy number with cardiovascular disease. *JAMA Cardiol* 2017;**2**:1247–1255.
13. Hong YS, Longchamps RJ, Zhao D, Castellani CA, Loehr LR, Chang PP, Matsushita K, Grove ML, Boerwinkle E, Arking DE, Guallar E. Mitochondrial DNA copy number and incident heart failure: the Atherosclerosis Risk in Communities (ARIC) study. *Circulation* 2020;**141**:1823–1825.
14. Zhao D, Bartz TM, Sotoodehnia N, Post WS, Heckbert SR, Alonso A, Longchamps RJ, Castellani CA, Hong YS, Rotter JI, Lin HJ, O'Rourke B, Pankratz N, Lane JA, Yang SY, Guallar E, Arking DE. Mitochondrial DNA copy number and incident atrial fibrillation. *BMC Med* 2020;**18**:246.
15. Liu X, Longchamps RJ, Wiggins KL, Raffield LM, Bielak LF, Zhao W, Pitsillides A, Blackwell TW, Yao J, Guo X, Kurniansyah N, Thyagarajan B, Pankratz N, Rich SS, Taylor KD, Peyser PA, Heckbert SR, Seshadri S, Cupples LA, Boerwinkle E, Grove ML, Larson NB, Smith JA, Vasan RS, Sofer T, Fitzpatrick AL, Fornage M, Ding J, Correa A, Abecasis G, Psaty BM, Wilson JG, Levy D, Rotter JI, Bis JC, Satizabal CL, Arking DE, Liu C. Association of mitochondrial DNA copy number with cardiometabolic diseases. *Cell Genom* 2021;**1**:100006.
16. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *Int J Epidemiol* 2016;**45**:1866–1886.
17. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;**12**:e1001779.
18. Hägg S, Jylhävä J, Wang Y, Czene K, Grassmann F. Deciphering the genetic and epidemiological landscape of mitochondrial DNA abundance. *Hum Genet* 2020;**140**:849–861.
19. Lee HC, Wei YH. Mitochondrial biogenesis and mitochondrial DNA maintenance of mammalian cells under oxidative stress. *Int J Biochem Cell Biol* 2005;**37**:822–834.
20. Longchamps RJ, Castellani CA, Yang SY, Newcomb CE, Sumpter JA, Lane J, Grove ML, Guallar E, Pankratz N, Taylor KD, Rotter JI, Boerwinkle E, Arking DE. Evaluation of mitochondrial DNA copy number estimation techniques. *PloS ONE* 2020;**15**:e0228166.
21. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;**30**:377–399.
22. Longchamps RJ, Yang SY, Castellani CA, Shi W, Lane J, Grove ML, Bartz TM, Sarnowski C, Liu C, Burrows K, Guyatt AL, Gaunt TR, Kacprowski T, Yang J, De Jager PL, Yu L, Bergman A, Xia R, Fornage M, Feitosa MF, Wojczynski MK, Kraja AT, Province MA, Amin N, Rivadeneira F, Tiemeier H, Uitterlinden AG, Broer L, Van Meurs JBJ, Van Duijn CM, Raffield LM, Lange L, Rich SS, Lemaitre RN, Goodarzi MO, Sitlani CM, Mak ACY, Bennett DA, Rodriguez S, Murabito JM, Lunetta KL, Sotoodehnia N, Atzmon G, Ye K, Barzilai N, Brody JA, Psaty BM, Taylor KD, Rotter JI, Boerwinkle E, Pankratz N, Arking DE. Genome-wide analysis of mitochondrial DNA copy number reveals loci implicated in nucleotide metabolism, platelet activation, and megakaryocyte proliferation. *Hum Genet* 2022;**141**:127–146.
23. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan J, Lamina C, Ziegler A, Zhang W, Zee RY, Wright AF, Witteman JC, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJ, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BW, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PK, Lucas G, Luben R, Loos RJ, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, König IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens AC, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Döring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJ, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA, Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJ, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010;**466**:707–713.
24. Staley JR, Burgess S. Semiparametric methods for estimation of a nonlinear exposure-outcome relationship using instrumental variables with application to Mendelian randomization. *Genet Epidemiol* 2017;**41**:341–352.
25. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, Saleheen D, Kyriakou T, Nelson CP, Hopewell JC, Webb TR, Zeng L, Dehghan A, Alver M, Armasu SM, Auro K, Bjonnes A, Chasman DI, Chen S, Ford I, Franceschini N, Gieger C, Grace C, Gustafsson S, Huang J, Hwang SJ, Kim YK, Kleber ME, Lau KW, Lu X, Lu Y, Lyytikäinen LP, Mihailov E, Morrison AC, Pervjakova N, Qu L, Rose LM, Salfati E, Saxena R, Scholz M, Smith AV, Tikkanen E, Uitterlinden A, Yang X, Zhang W, Zhao W, de Andrade M, de Vries PS, van Zuydam NR, Anand SS, Bertram L, Beutner F, Dedoussis G, Frossard P, Gauguier D, Goodall AH, Gottesman O, Haber M, Han BG, Huang J, Jalilzadeh S, Kessler T, König IR, Lannfelt L, Lieb W, Lind L, Lindgren CM, Lokki ML, Magnusson PK, Mallick NH, Mehra N, Meitinger T, Memon FU, Morris AP, Nieminen MS, Pedersen NL, Peters A, Rallidis LS, Rasheed A, Samuel M, Shah SH, Sinisalo J, Stirrups KE, Trompet S, Wang L, Zaman KS, Ardissino D, Boerwinkle E, Borecki IB, Bottinger EP, Buring JE, Chambers JC, Collins R, Cupples LA, Danesh J, Demuth I, Elosua R, Epstein SE, Esko T, Feitosa MF, Franco OH, Franzosi MG, Granger CB, Gu D, Gudnason V, Hall AS, Hamsten A, Harris TB, Hazen SL, Hengstenberg C, Hofman A, Ingelsson E, Iribarren C, Jukema JW, Karhunen PJ, Kim BJ, Kooner JS, Kullo IJ, Lehtimäki T, Loos RJF, Melander O, Metspalu A, März W, Palmer CN, Perola M, Quertermous T, Rader DJ, Ridker PM, Ripatti S, Roberts R, Salomaa V, Sanghera DK, Schwartz SM, Seedorf U, Stewart AF, Stott DJ, Thiery J, Zalloua PA, O'Donnell CJ, Reilly MP, Assimes TL, Thompson JR, Erdmann J, Clarke R, Watkins H, Kathiresan S, McPherson R, Deloukas P, Schunkert H, Samani NJ, Farrall M. A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 2015;**47**:1121–1130.
26. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol* 2016;**45**:1717–1726.
27. Burgess S, Butterworth A, Thompson SG. Mendelian Randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;**37**:658–665.
28. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a Weighted Median Estimator. *Genet Epidemiol* 2016;**40**:304–314.
29. Bowden J, Davey Smith G, Burgess S. Mendelian Randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;**44**:512–525.
30. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* 2018;**50**:693–698.
31. Minelli C, Del Greco MF, van der Plaat DA, Bowden J, Sheehan NA, Thompson J. The use of two-sample methods for Mendelian randomization analyses on single large datasets. *Int J Epidemiol* 2021.
32. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017;**13**:e1007081.
33. Madamanchi NR, Runge MS. Mitochondrial dysfunction in atherosclerosis. *Circ Research* 2007;**100**:460–473.
34. Bursi F, Weston SA, Redfield MM, Jacobsen SJ, Pakhomov S, Nkomo VT, Meverden RA, Roger VL. Systolic and diastolic heart failure in the community. *JAMA* 2006;**296**:2209–2216.
35. Ahmad T, Pencina MJ, Schulte PJ, O'Brien E, Whellan DJ, Piña IL, Kitzman DW, Lee KL, O'Connor CM, Felker GM. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. *J Am Coll Cardiol* 2014;**64**:1765–1774.
36. Francis GS, Cogswell R, Thenappan T. The heterogeneity of heart failure: will enhanced phenotyping be necessary for future clinical trial success? *J Am Coll Cardiol* 2014;**64**:1775–1776.
37. Volobueva A, Grechko A, Yet SF, Sobenin I, Orekhov A. Changes in mitochondrial genome associated with predisposition to atherosclerosis and related disease. *Biomolecules* 2019;**9**.

Translational perspective

We provide evidence that low leukocyte mtDNA abundance is a causal risk factor, independent of other traditional cardiovascular risk factors, for atherosclerotic cardiovascular diseases among individuals without known cardiovascular diseases. Future studies should address the value of incorporation of mtDNA abundance into risk prediction on cardiovascular risk management in clinical practice.