Oxidative stress in chronic diseases: causal inference from observational studies

Luo, J.

CHAPTER 3

Low mitochondrial DNA copy number drives atherogenic cardiovascular disease: cohort and genetic studies

Jiao Luo, Raymond Noordam, J Wouter Jukema, Ko Willems van Dijk, Sara Hägg, Felix Grassmann, Saskia le Cessie, Diana van Heemst

In revision

Abstract

Aim: Mitochondrial DNA dysfunction has been implicated in the pathogenesis of cardiovascular disease. We aimed to investigate the associations of mitochondrial DNA copy number (mtDNA-CN), as a proxy of mitochondrial function, and coronary artery disease (CAD) and heart failure (HF) in a cohort study and test the causal nature of any potential associations using Mendelian randomization (MR).

Methods and Results: Multivariable-adjusted regression analyses were conducted in 273,619 unrelated participants of European ancestry from UK Biobank (UKB). In two-sample MR analyses, single nucleotide polymorphisms (SNPs) associated with mtDNA-CN were retrieved from published genome-wide association studies. SNP-disease associations were obtained for CAD from CARDIoGRAMplusC4D, UKB, and FinnGen, comprising 902,538 participants (134,759 cases), and for HF from the HERMES consortium and FinnGen, comprising 1,195,531 participants (70,706 cases). MR analyses were performed per database and results were subsequently meta-analyzed using fixed-effects models. During a median follow-up of 11.8 years, cox regression restricted cubic spline analyses showed associations between lower mtDNA-CN and higher risk of CAD and HF. Hazard ratios for participants in the lowest quintile of mtDNA-CN 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, the pooled odds ratios from two-sample MR of genetically predicted per one-SD decrease in mtDNA-CN were 1.16 (1.05, 1.27) for CAD and 1.00 (0.90, 1.10) for HF, respectively.

Conclusion: Our findings support a possible causal role of lower mtDNA-CN in higher CAD risk, but not in HF.

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 of the mitochondrial genome, known as mtDNA copy numbers (mtDNA-CN). The mtDNA-CN are 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-CN may provide early opportunities in the prevention and treatment of CVD.

Recent epidemiological studies have assessed the associations between peripheral blood 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 cardiovascular disease 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-CN was associated with a cluster of cardiometabolic traits that increase the risk of CVD, including obesity, hypertension, and hyperlipidemia[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 of a causal nature, therefore, remains unclear.

Triangulation of causal inference in etiological epidemiology has been proposed, which integrates results from different methodological approaches to enhance the reliability of a research study[16]. The confidence of 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-CN is associated with an increased risk of incident CVD. Consequently, we first examined the associations between mtDNA-CN and incidence of coronary artery disease (CAD) and heart failure (HF) in participants of European ancestry in the UK Biobank (UKB) using cox-proportional hazard regression models. Second, we exploited publicly available data to perform a two-sample Mendelian randomization (MR) to investigate whether genetically predicted low mtDNA-CN were causally associated with increased risk of diseases.

Methods

Prospective study

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 centers 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 of 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 who were: 1) not used to compute principal components; 2) identified as outliers in heterozygosity and missing rates, which is indicative of poor-quality genotype data for these samples; 3) identified as putatively sex chromosome aneuploidy; 4) sex inferred from genotype data did not match their self-reported sex; 5) have 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 further refined the population definition 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 on the different exclusions is provided in **Figure S1**.

mtDNA-CN computation

Somatic mtDNA-CN were assessed from the intensities of genotyping probes on the mitochondrial chromosome on the Affymetrix Array. The method for computing mtDNA-CN has been described in detail before[18]. We followed the same pipeline to calculate mtDNA-CN 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 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-CN. 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-CN estimate for each individual. To eliminate the plate effect, we subsequently

standardized the CN to a mean of zero and standard deviation (SD) of one within each genotyping plate comprising 96 wells. An additional quality control step was performed by eliminating individuals with high standard deviation (SD) (two SD from the mean) of autosomal probes log2 ratio (L2R). Consequently, 293,245 individuals remained in the cohort.

Outcome definition

Outcomes in the analysis were incident cardiovascular diseases during the time period from recruitment to January 1st, 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 ischemic heart disease (IHD) (I24 and I25); Incident HF cases were defined as I50. In addition, we analyzed acute myocardial infarction (MI) and chronic ischemic heart disease (IHD) as separate outcomes in sensitivity analyses. Follow-up time is computed from the baseline visit to the diagnosis of incident disease, death, or the censoring, whichever occurred first.

Covariates

Covariates were from baseline measurements, which included demographic parameters (age at recruitment, sex, deprivation index); the first ten principal components (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.

Statistical analysis

After further exclusion of participants with any prevalent cardiovascular disease or withdrawn informed consent, the study cohort comprised an analytic sample of 273,619 individuals (**Figure S1**). Baseline characteristics of the study population were described in quintiles of mtDNA-CN 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-CN quintiles, accounting for death as a competing event.

Cox proportional hazards models were used to estimate hazard ratios (HR) and corresponding 95% confidence intervals (CI) for the association between mtDNA-CN and incident CAD and HF separately. Two multivariable-adjusted regression models were fitted: Model 1 was adjusted for age, sex, the first ten 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-CN as a continuous variable and assessed the risk of incident diseases associated with per one-SD decrease in mtDNA-CN using cox restricted cubic spline curves, with knots located at 5th, 50th, and 95th percentiles. Subsequently, mtDNA-CN was categorized into quintiles, and hazard ratios (HR) compared the 1st to 4th quintiles with the 5th quintile (reference). The proportional hazard assumption was graphically assessed by plotting log(-log[survival]) versus 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)[19], 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-CN, 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-CN 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)

Mendelian randomization analyses

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-CN at a genome-wide significance threshold ($p < 5e-08$), as identified in a recent genome-wide association study (GWAS) by Longchamps et al20. 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 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. F statistics [$(\beta/se)^2$] were computed to evaluate instrumental strength, and SNPs with a value of less than 10 were considered weak instruments. Furthermore,

we calculated the proportion of total variance in the exposure explained by each instrument (R^2) separately[21].

Outcome data source

Summary statistics for instrument-CAD associations were extracted from 3 large databases separately, the CARDIoGRAMplusC4D (Coronary Artery Disease Genome-Wide Replication and Meta-analysis plus the Coronary Artery Disease Genetics from Nikpay et al.[22], 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 associations tests of each of the databases are presented in detail in **Supplementary methods** and **Table S1**.

Mendelian randomization analysis

SNP-exposure and SNP-outcome data were harmonized to make alignment on effect alleles. Palindromic SNPs were eliminated[23]. 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[24]. Results were expressed as an odds ratio (OR) on disease risk for a one-SD decrease in genetically predicted mtDNA-CN. 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[25,26], both of which assume 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[27]. 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, with the exception of the MR-Egger sensitivity analysis[28].

Meta-analysis of estimates from different databases

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

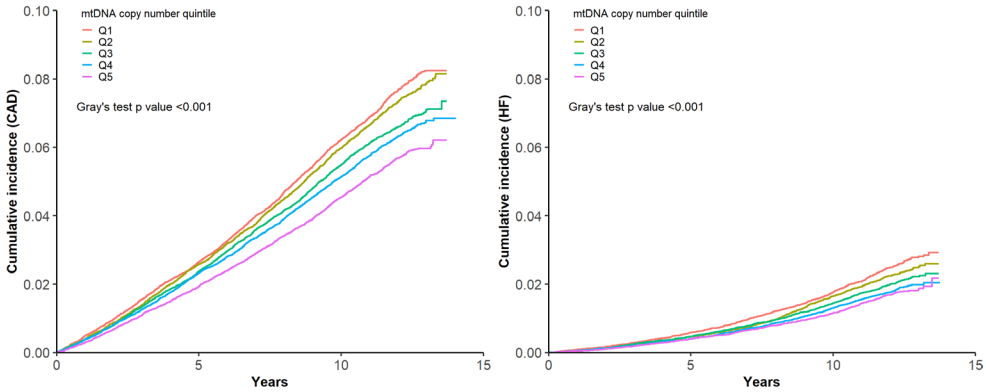

Figure 1 Cumulative Incidence of CAD and HF by quintiles of mtDNA copy number

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

Sensitivity analysis

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-CN 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 in the UKB only. Therefore, 66 independent (linkage disequilibrium < 0.1) SNPs were used that were associated with mtDNA-CN at a genome-wide significance threshold ($p < 5e\text{-}08$) 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. All MR analyses were repeated with the substitution of the 69 genetic instruments for mtDNA-CN.

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 nonlinear dose-response associations, "TwoSampleMR" for MR analyses, and "meta" for meta-analyses. All results were reported as odds ratio with accompanied 95% confidence interval.

Results from the multivariable-adjusted prospective analyses

Prospective results

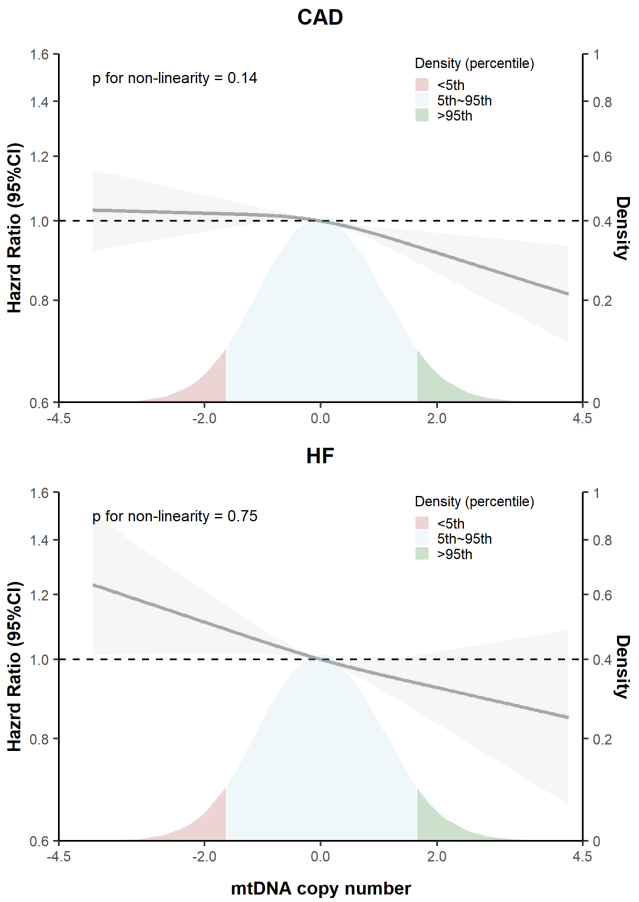

Figure 2 Hazard ratios for incident CAD and HF by levels of mtDNA copy number.

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 (gray shadowed area) using restricted cubic splines for mtDNA copy number with knots at distribution of 5th, 50th, and 95th percentiles. The density on the right y-axis shows distribution of baseline mtDNA copy number. Since mtDNA copy number has been standardized during coputation, the distribution is approximately normal. Since mtDNA copy number was scaled before analyses, the distribution is normal. CAD: coronary artery disease; HF: heart failure.

Main analyses

A total of 273,619 participants were eligible for analyses after exclusion (**Figure S1**). Compared with the highest quintile of mtDNA-CN (**Table 1**), participants in the lower quintiles were more likely to have more unfavorable CVD risk factors, including older age, male sex, higher BMI, higher blood pressure and more

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%)	18702 (34.2%)	
Never	29,732 (54.3%)	30,079 (55.0%)	30,594 (55.9%)	30,914 (56.5%)	31350 (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	

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 chi-square test as appropriate. BMI: Body mass index; HDL: high-density lipoprotein cholesterol; LDL: low-density lipoprotein cholesterol; CVD: cardiovascular disease; CAD: coronary artery disease; HF: heart failure.

Table 1 Baseline charactertistics of the study participants by quintiles of mtDNA copy number

Variable	Q1	Q2	Q3	Q4	Q5	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	47128 (86.1%)	47330 (86.5%)	47575 (86.9%)	47902 (87.5%)	48061 (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%)	

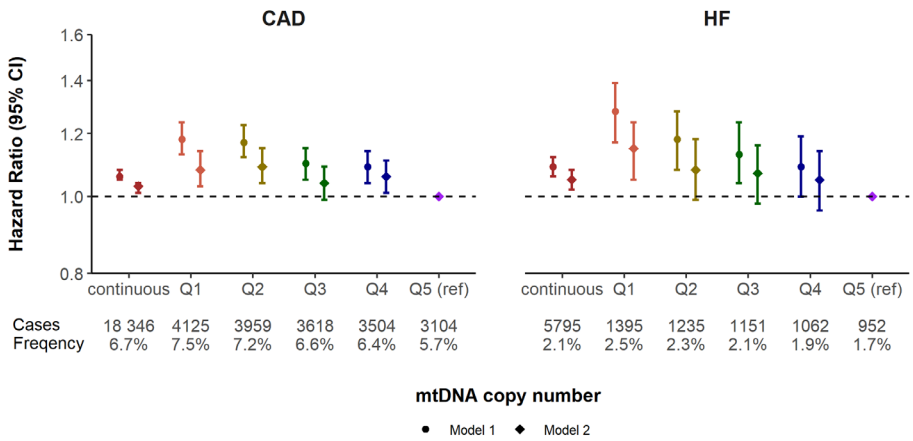

Figure 3 Hazard ratios for incident CAD and HF by quintiles of mtDNA copy number

Estimated hazard ratios for the effect of per-SD decrease in mtDNA copy number (continuous), or for the 1st to the 4th quintile compared to the 5th (reference) 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.

blood pressure-lowering medication, higher lipids (total cholesterol and LDL) 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 (interquartile range: 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 mtDNA-CN, 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-CN with the higher risk of CAD (p for non-linearity = 0.14) and HF (p for non-linearity = 0.73), as shown in **Figure 2**. Categorically (model 1), a one-SD decrease in mtDNA-CN was associated with 1.06-fold (95% confidence interval, CI: 1.05, 1.08) and 1.09-fold (95%CI: 1.06, 1.12) higher hazard of CAD and HF, respectively; adjusted HRs for the first versus the fifth (reference) quintile of mtDNA-CN were 1.18 (95%CI: 1.23, 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 **Table S2**).

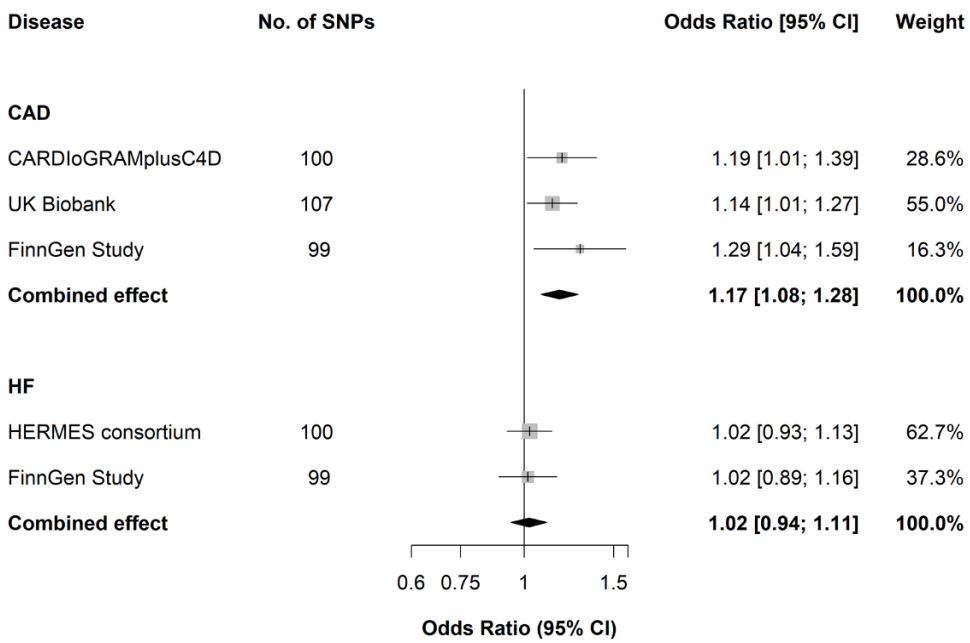

Figure 4 Mendelian randomization study of mtDNA copy number on the risk of CAD and HF

Estimated ORs for the effect of per-SD decrease in mtDNA copy number 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.

Sensitivity analysis

We observed no evidence favoring an interaction between mtDNA-CN and sex (p for interaction = 0.2 for CAD, 0.7 for HF); in line, in sex-stratified analyses, the estimates between men and women were similar (**Table S3**). The interaction was observed between mtDNA-CN and age at baseline 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) (**Table S4**). However, no interaction was detected between mtDNA-CN 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 (**Figure S2**) for MI and IHD; estimates from cox proportional hazard regression models did not differ considerably from when all CADs were considered (**Table S2-S4**).

In addition, missing data in covariates were present (**Table S5**), 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 (**Table S6**). Furthermore, the main results from sensitivity analyses restricting to complete cases did not differ materially from the results obtained after imputation (**Table S7**).

Mendelian randomization

Main analyses

For the included 129 genetic instrumental variables, 4 of which with an F-statistics below 10 were discarded to avoid weak instrumental bias. In total, 108 distinct SNPs were present in at least one of the outcome databases. F-statistics for each SNP were higher than 10 and ranged from 15.6 to 634.4, and a total of 2.0% variation were explained by the instruments (**Table S8**).

For CAD, the pooled OR of the primary IVW estimates from CARDIoGRAMplusC4D, UKB, and FinnGen of a one-SD decrease in mtDNA-CN was 1.17 (95%CI: 1.08, 1.28) (**Figure 4**). Estimates from WME and MR-Egger generally did not differ substantially with the exception of 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 (**Table S9**).

For HF, the combined OR from IVW obtained in the HERMES consortium and FinnGen of per one-SD decrease in mtDNA-CN was 1.02 (95%CI: 0.94, 1.11) (**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 with those generated from IVW (**Table S10**).

Sensitivity analyses

When we used genetic instrumental variables from Hägg et al.[18], 64 distinct SNPs were included. F statistics for each SNPs were higher than 10 and ranged from 29.8 to 277.4, and a total of 1.2% variation were explained by the. Detailed full information of the used genetic instruments is presented in **Table S11**. A one-SD decrease in mtDNA-CN was associated with 1.16-fold (95%CI:1.05, 1.27), 1.00-fold (95%CI:0.90, 1.10) higher risk of CAD and HF in meta-analysis, respectively (**Figure S3**). MR sensitivity analyses including WME, MR-Egger, and MR PRESSO are presented in **Table S12-S13**.

Discussion

In the present study, we implemented a prospective cohort study design and MR study to assess the relationship of mtDNA-CN with the risk of incident CAD and HF. Results from the multivariable-adjusted prospective analyses suggested an association between lower mtDNA-CN with a higher risk of CAD and HF, whereas

findings from MR analyses only confirmed an association between genetically predicted lower mtDNA-CN with 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 peripheral blood 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 Atherosclerosis Risk in Communities (ARIC) Study[12,13]. In the ARIC study, composed of 20,163 participants (2460 incident CHD) with a mean follow-up of 13.5 years, a lower mtDNA-CN was associated with an increased risk of incident CHD. Similarly, with 10,802 participants (2227 incident HF cases) followed-up for a mean of 23.1 years, lower mtDNA-CN 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 association between mtDNA-CN and risk of CVD. MR analyses with the genetic instruments for mtDNA-CN confirmed the detrimental effect of lower mtDNA content on the risk of CAD observed in the cohort studies. Mitochondrial dysfunction, indicating by low mtDNA-CN, 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 low-density lipoprotein and dysfunction of endothelial cells that are critical factors to promote atherosclerosis, and further CAD events[29]. 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[30]. Moreover, a large degree of variation has been described even within patients with preserved EF[31,32]. It has also been shown previously that the associations between mtDNA-CN and HF with preserved and reduced EF were different and possibly would make the association into the direction of 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. In addition, cause-specific GWAS summary-level data of HF are currently not available. For these reasons, the lack of a clear association between mtDNA-CN and HF should be interpreted with caution, and more follow-up analyses are required to investigate the cause-specific HF in more detail.

Study strengths and limitations

The main strength of our study is that we adopted the triangulation of causal inference in etiological epidemiology[16]. The consistency between biochemically measured and genetically determined mtDNA-CN in relation to CAD increased the credibility of the results. Given the absence of randomized clinical trials with respect to mtDNA-CN 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 size and the considerable number of incident cases from UKB, comprehensive assessment of confounding factors, and subtype analyses of MI and IHD within CAD. In MR studies, we meta-analyzed three large databases where SNP-outcome associations were derived, comprising a substantial size of overall participants and cases. The results are consistent 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 peripheral blood cells, which may be different from cells in the vasculature or in the heart. Nevertheless, it has been shown that the blood cell and cardiomyocyte mtDNA-CN were significantly correlated within individuals, with a Pearson correlation coefficient of 0.72[33]. In addition, the initial calculation of mtDNA-CN from chip arrays might have introduced noise due to the small number of variants. To this end, a weighted mtDNA-CN was implemented, which approximates what would be estimated from exome sequencing and has been validated[18]. Second, despite the large number of instrumental variables in the MR analyses, the variation of mtDNA-CN explained by these SNPs was small. Notwithstanding, we had more than sufficient power to detect the true causal effect in MR analyses (**Figure S4**). Moreover, while we acknowledge the possibility of pleiotropic effects of included genetic instruments, this is likely to be vertical (**Supplementary discussion**). When we additionally stringently excluded the 22 SNPs on chromosome 19, where APOE and LDLR gene locate, the estimates were slightly attenuated in each database but remained statistically significant in the meta-analysis (**Figure S5**). Third, since the population of non-Europeans was highly heterogeneous in UKB, we restricted the prospective analyses to White European populations; furthermore, MR analyses were also performed predominantly in European-descent individuals, except for 23% of individuals with a non-European background in CARDIoGRAMplusC4D. 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[34]. Consequently, there is a need for accurate deep sequencing to simultaneously analyze the entire mitochondrial genome in order to better understand the relationships between mtDNA-CN function, germline and acquired mutations, and CVD.

Conclusion

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

References

1. J. Schaper, E. Meiser, G. Stämmler. Ultrastructural morphometric analysis of myocardium from dogs, rats, hamsters, mice, and from human hearts. Circ Res 1985; 56(3): 377-91.
2. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer. The hallmarks of aging. Cell 2013; 153(6): 1194-217.
3. C. A. Castellani, R. J. Longchamps, J. Sun, E. Guallar, D. E. Arking. Thinking outside the nucleus: Mitochondrial DNA copy number in health and disease. Mitochondrion 2020; 53: 214-23.
4. D. F. Dai, P. S. Rabinovitch, Z. Ungvari. Mitochondria and cardiovascular aging. Circ Res 2012; 110(8): 1109-24.
5. D. C. Wallace. A mitochondrial bioenergetic etiology of disease. J Clin Invest 2013; 123(4): 1405-12.
6. A. N. Malik, A. Czajka. Is mitochondrial DNA content a potential biomarker of mitochondrial dysfunction? Mitochondrion 2013; 13(5): 481-92.
7. R. Wei, Y. Ni, P. Bazeley, et al. Mitochondrial DNA Content Is Linked to Cardiovascular Disease Patient Phenotypes. J Am Heart Assoc 2021; 10(4): e018776.
8. S. Chen, X. Xie, Y. Wang, et al. Association between leukocyte mitochondrial DNA content and risk of coronary heart disease: a case-control study. Atherosclerosis 2014; 237(1): 220-6.
9. X. B. Wang, N. H. Cui, S. Zhang, Z. J. Liu, J. F. Ma, L. Ming. 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: 165-72.
10. A. Koller, F. Fazzini, C. Lamina, et al. Mitochondrial DNA copy number is associated with all-cause mortality and cardiovascular events in patients with peripheral arterial disease. J Intern Med 2020; 287(5): 569-79.
11. Y. Zhang, E. Guallar, F. N. Ashar, et al. Association between mitochondrial DNA copy number and sudden cardiac death: findings from the Atherosclerosis Risk in Communities study (ARIC). Eur Heart J 2017; 38(46): 3443-8.
12. F. N. Ashar, Y. Zhang, R. J. Longchamps, et al. Association of Mitochondrial DNA Copy Number With Cardiovascular Disease. JAMA Cardiol 2017; 2(11): 1247-55.
13. Y. S. Hong, R. J. Longchamps, D. Zhao, et al. Mitochondrial DNA Copy Number and Incident Heart Failure: The Atherosclerosis Risk in Communities (ARIC) Study. Circulation 2020; 141(22): 1823-5.
14. D. Zhao, T. M. Bartz, N. Sotoodehnia, et al. Mitochondrial DNA copy number and incident atrial fibrillation. BMC Med 2020; 18(1): 246.
15. Xue Liu, Ryan J. Longchamps, Kerri L. Wiggins, et al. Association of mitochondrial DNA copy number with cardiometabolic diseases. Cell Genomics 2021; 1(1): 100006.
16. D. A. Lawlor, K. Tilling, G. Davey Smith. Triangulation in aetiological epidemiology. Int J Epidemiol 2016; 45(6): 1866-86.
17. C. Sudlow, J. Gallacher, N. Allen, et al. 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(3): e1001779.

18. S. Hägg, J. Jylhävä, Y. Wang, K. Czene, F. Grassmann. Deciphering the genetic and epidemiological landscape of mitochondrial DNA abundance. Hum Genet 2020.
19. I. R. White, P. Royston, A. M. Wood. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011; 30(4): 377-99.
20. R. J. Longchamps, S. Y. Yang, C. A. Castellani, et al. Genome-wide analysis of mitochondrial DNA copy number reveals loci implicated in nucleotide metabolism, platelet activation, and megakaryocyte proliferation. Hum Genet 2022; 141(1): 127-46.
21. T. M. Teslovich, K. Musunuru, A. V. Smith, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010; 466(7307): 707-13.
22. M. Nikpay, A. Goel, H. H. Won, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 2015; 47(10): 1121-30.
23. F. P. Hartwig, N. M. Davies, G. Hemani, G. Davey Smith. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. Int J Epidemiol 2016; 45(6): 1717-26.
24. S. Burgess, A. Butterworth, S. G. Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013; 37(7): 658-65.
25. J. Bowden, G. Davey Smith, P. C. Haycock, S. Burgess. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol 2016; 40(4): 304-14.
26. J. Bowden, G. Davey Smith, S. Burgess. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015; 44(2): 512-25.
27. M. Verbanck, C. Y. Chen, B. Neale, R. Do. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet 2018; 50(5): 693-8.
28. C. Minelli, M. F. Del Greco, D. A. van der Plaat, J. Bowden, N. A. Sheehan, J. Thompson. The use of two-sample methods for Mendelian randomization analyses on single large datasets. Int J Epidemiol 2021.
29. N. R. Madamanchi, M. S. Runge. Mitochondrial dysfunction in atherosclerosis. Circ Res 2007; 100(4): 460-73.
30. F. Bursi, S. A. Weston, M. M. Redfield, et al. Systolic and diastolic heart failure in the community. Jama 2006; 296(18): 2209-16.
31. T. Ahmad, M. J. Pencina, P. J. Schulte, et al. Clinical Implications of Chronic Heart Failure Phenotypes Defined by Cluster Analysis. Journal of the American College of Cardiology 2014; 64(17): 1765-74.
32. G. S. Francis, R. Cogswell, T. Thenappan. The heterogeneity of heart failure: will enhanced phenotyping be necessary for future clinical trial success? J Am Coll Cardiol 2014; 64(17): 1775-6.
33. J. Huang, L. Tan, R. Shen, L. Zhang, H. Zuo, D. W. Wang. Decreased Peripheral Mitochondrial DNA Copy Number is Associated with the Risk of Heart Failure and Long-term Outcomes. Medicine (Baltimore) 2016; 95(15): e3323.
34. A. Volobueva, A. Grechko, S. F. Yet, I. Sobenin, A. Orekhov. Changes in Mitochondrial Genome Associated with Predisposition to Atherosclerosis and Related Disease. Biomolecules 2019; 9(8).

Supplementary materials

Table S1 GWAS data sources for Mendelian Randomization

Table S2 Hazard ratio of incident CAD and HF by quintiles of mtDNA-CN

Table S3 Hazard ratio of incident CAD and HF by levels of mtDN-CN, stratified by sex

Table S4 Hazard ratio of incident CAD and HF by levels of mtDN-CN, stratified by age

Table S5 Number and frequencies of missingness in each variable (N = 273,619)

Table S6 Differences between participants with complete data and participants with missing data

Table S7 Cox proportional hazard regression in complete cases

Table S8 Genetic instruments at genome-wide significant level for mtDNA-CN in the main MR analyses (Instruments retrieved from Longchamps et al.)

Table S9 Mendelian Randomization results of mtDNA-CN on the risk of CAD (Instruments retrieved from Longchamps et al.)

Table S10 Mendelian Randomization results mtDNA-CN on the risk of HF (Instruments retrieved retrieved from Longchamps et al.)

Table S11 Genetic instruments at genome-wide significant level for mtDNA-CN in the sensitivity MR analysis (Instruments retrieved from Hägg et al.)

Table S12 Mendelian Randomization results of mtDNA-CN on the risk of CAD (Instruments retrieved from Hägg et al.)

Table S13 Mendelian Randomization results mtDNA-CN on the risk of HF (Instruments retrieved from Hägg et al.)

Figure S1 Flowchart of participants inclusion in UK Biobank

Figure S2 Cumulative Incidence of MI and IHD by quintiles of mtDNA-CN

Figure S3 Mendelian Randomization study of mtDNA-CN on the risk of CAD and HF (Instruments retrieved from Hägg et al.)

Figure S4 Statistical power of Mendelian Randomization analyses

Figure S5 Mendelian Randomization study of mtDNA-CN on the risk of CAD and HF (Instruments retrieved from Longchamp et al. excluding 22 SNPs from chromosome 19)

Supplementary methods

Supplementary Discussion

The Supplementary materials for this article can be found online at: *https://drive.google.com/drive/folders/1d46G5jf6fIZUUIp6aHpjnz74SnL-jzQPd?usp=sharing*