# Oxidative stress in chronic diseases: causal inference from observational studies

Luo, J.

# CHAPTER 4

**Mitochondrial DNA copy number and circulating metabolomic profiling: multivariable-adjusted and Mendelian randomization analyses**

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

Under submission

## Abstract

**Background:** Low leukocyte mitochondrial DNA copy number (mtDNA-CN) has been associated with a high risk of atherosclerotic cardiovascular disease, but the mechanisms remain unclear. Changes in metabolomic profiles have frequently been observed during the transition from health to disease and are considered important intermediate phenotypes. Therefore, we aimed to investigate whether low mtDNA-CN is associated with worse metabolomic profiling using cross-sectional and genetic studies.
**Methods:** Among 61,186 unrelated European participants from the UK Biobank, we performed multivariable-adjusted linear regression analyses to examine the associations between mtDNA-CN and 168 NMR-based circulating metabolic biomarkers and 9 metabolic principal components (PCs) that collectively covered 91.5% of the total variation of individual metabolic biomarkers. Subsequently, we conducted Mendelian randomization (MR) to estimate the causal effects of mtDNA-CN on individual biomarkers and the metabolic PCs. Genetic variants associated with mtDNA-CN, used as instrumental variables, were derived from a publicly available genome-wide association study of 465,809 individuals of White European ancestry, and genetic-biomarker associations were derived from the UK Biobank.
**Results:** After correction for multiple testing, low mtDNA-CN was associated with 130 metabolic biomarkers, predominantly lower concentrations of some amino acids and higher concentrations of lipids, lipoproteins and fatty acids; moreover, mtDNA-CN was associated with 7 out of the 9 metabolic PCs. Using MR, genetically-influenced low mtDNA-CN was associated with lower lactate, higher acetate, and a higher degree of unsaturation, and corresponding standardized betas (95% CIs) from the inverse variance weighted method were -0.17 (-0.26, -0.09), 0.15 (0.07, 0.23), and 0.14 (0.08, 0.20), respectively. In addition, per-SD decrease of mtDNA-CN was associated with lower metabolic PC2 (related to lower concentrations of lipids and fatty acids), and higher metabolic PC9 (related to lower concentrations of glycolysis-related metabolites).
**Conclusion:** Low mtDNA-CN was associated with metabolomic perturbations, particularly reflecting a pro-atherogenic metabolic profile. These findings support the potential of metabolomics to provide new insights into the mitochondrial induced changes preceding chronic diseases.

## Introduction

Mitochondria play a central role in the human body, as they are responsible for most of the cellular energy production through oxidative phosphorylation, biosynthetic processes of macromolecules such as heme and steroid hormone, and maintenance of calcium and redox homeostasis for signaling pathways[1-3]. Mitochondria have their own circular genome, the mitochondrial DNA (mtDNA), consisting of 37 genes, 13 of which encode proteins on the electron transport chain. The individual mitochondrion contains several copies of the mitochondrial genome, known as mtDNA copy number (mtDNA-CN), and therefore large variations of mtDNA-CN may exist between cells, tissues, and individuals. Altered mtDNA levels contribute to defects of mtDNA-coded proteins and RNAs, and disrupted cellular signals for apoptosis and autophagy, resulting in mitochondrial dysfunction[4,5]. Consequently, this will further aggravate oxidative stress and inflammation, both of which are key components in many pathophysiological processes, thus promoting chronic diseases and accelerating ageing[6,7].

Mitochondrial dysfunction has been widely considered as a hallmark of the ageing process and compromised health[6,7]. Indeed, mtDNA-CN decreases with increasing age, especially among individuals older than 65 years[8,9]. In prospective cohort studies, low mtDNA-CN measured in the blood leukocyte has been linked to increased risk of several cardiovascular outcomes among the general population[10-13] and mortality among patients with chronic kidney diseases[14], as well as poor poststroke prognosis[15]. Similarly, in our previous studies, we observed associations between low mtDNA-CN and higher risks of coronary artery disease (CAD), heart failure, and stroke[16,17], and for CAD we provided evidence favoring possible causality. However, the underlying mechanisms are not clear yet.

Importantly, all these diseases experience metabolic changes prior to and during their onset and progression. Metabolites are small molecules that result from biological processes, representing effective intermediate phenotypes to a given disease[18]. Metabolomics has offered the opportunity to comprehensively measure a broad range of metabolites in biological fluids, and thus to provide insight into disease mechanisms and to discover potential biomarkers. Therefore, we investigated the associations between mtDNA-CN in leukocytes and blood-derived metabolomic profiles with 168 metabolic biomarkers of predominantly lipids and lipoproteins (sub)particles measured using the Nightingale NMR-based platform in the UK Biobank (UKB). We triangulated our study by performing cross-sectional multivariable-adjusted analyses, followed by Mendelian randomization (MR) studies that are free of most confounding factors and reverse causation.

# Methods

## Cross-sectional study

### *Study population*

The UKB 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. The UK biobank study was approved by the North-West Multi-center Research Ethics Committee (MREC). Access for information to invite participants was approved by the Patient Information Advisory Group (PIAG) for England and Wales. All participants in the UK Biobank provided written informed consent. Invitation letters were sent to eligible adults registered to the National Health Services (NHS) and living within a 25 miles distance from one of the assessment centers. Participants provided information on their lifestyle and medical history through touch-screen questionnaires and physical measurements. Blood samples were collected for biochemical analyses and genotyping. All participants provided electronically written informed consent for the study. Detailed information about the study design, investigation methods, as well as limitations has been reported previously[19]. Further details of the array design, genotyping, and imputation procedures have been described elsewhere[20]. All analyses were conducted under UKB application 56340.

### *mtDNA-CN*

Full genotyped data were available from 488,377 individuals. MtDNA-CN was computed for unrelated European ancestry participants. A detailed step-wise exclusion of the study population was described elsewhere[16]. Somatic blood-derived 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 (*https://github.com/GrassmannLab/MT_UKB*)[8]. Briefly, 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, L2R values of each probe were 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. This resulted 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 with eligible mtDNA-CN remained.

***Metabolomic profiling***

A subset of ~120,000 UKB participants were randomly selected for plasma metabolomic profiling using a high throughput NMR-metabolomics platform (Nightingale Health, Helsinki, Finland). Details of the experimentation and applications of the platform have been described previously[21], as well as representative coefficients of variations for the metabolic biomarkers.[22] This platform provides simultaneous quantification of 168 metabolomic measures and 81 ratios from the plasma samples. We included only the direct measures in the current analyses and thus excluded the ratios, including lipoprotein subclasses (size-specific, n = 98), lipoprotein particle sizes (n = 7), cholesterol (n = 15), glycerides and phospholipids (n = 12), total lipids (n = 4), fatty acids (n= 9), amino acids (n= 10), glycolysis-related metabolites (n= 4), fluid balance (n= 2), inflammation (n= 1), ketone bodies (n= 4), apolipoproteins (n= 2). A full list of the measured metabolic biomarkers and their concentrations are presented in **Table S1**.



***Statistical analyses***

Individuals with any prevalent cardiovascular disease at baseline assessment were excluded. All metabolic biomarkers were log-transformed and subsequently scaled to standard deviation (SD) units for comparison before analyses.
Confounding factors were selected based on previous studies. Confounders included demographic parameters (age at recruitment, sex, deprivation index); the first ten genetic principal components to correct for possible remaining population stratification; genotyping batch; cell numbers (white blood cell counts and platelet counts); an 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]); lipid-lowering medication, blood pressure (mmHg, average of the two measurements taken a few moments apart when applicable).

Multivariable-adjusted linear regression models were performed to estimate the coefficient, representing changes in SD unit of individual metabolic biomarker corresponding to per one-SD lower mtDNA-CN, adjusted for confounding factors. Due to the high correlation of most metabolomic traits, particularly lipid subclasses, conventional correction for multiple testing is too stringent. Therefore, we estimated the number of independent tests from the correlation matrix of the metabolomic measures using the “effective number” (Meff) procedure as described previously to correct for the significance level[23], with an adjusted p-value of 0.05/37 (0.00134), where 37 represents the number of independent metabolomic measures.

Furthermore, given the intercorrelation of metabolic biomarkers, especially among lipid and lipoprotein (sub)particles, we subsequently performed principal component analysis (PCA), which aims at explaining the maximum variation in a multivariate dataset without a priori defined group, on all biomarkers to reduce the large number of correlated biomarkers to a much smaller number of uncorrelated metabolic principal components (PCs). PCs explained at least 1% variations and had an eigenvalue above 1 were selected for further analyses.

Consequently, 9 PCs met the criteria and were selected. The same analyses were performed for each PC as those for individual biomarkers.

We additionally conducted sensitivity analyses for individual biomarkers by excluding lipid-lowering medication users (N = 7770). Missing data in the covariates (all less than 20%) were imputed using multiple imputations by chain equations (MICE)[24], setting the number of imputed datasets to 10. The imputation model included all covariates. A detailed description and comparisons between individuals with and without missing data in the UKB were available in our previous study[16].

## Mendelian randomization

### *Data source*

133 independent (linkage disequilibrium < 0.05) nuclear single-nucleotide polymorphisms (SNPs) on autosomal chromosomes were used as genetic instruments that were previously identified in relation to mtDNA-CN at a genome-wide significance threshold ($p < 5e\text{-}08$)[25]. This genome-wide association study (GWAS) 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 genetic PCs, blood collection sites, family structure, and cell composition. F statistics [$(\beta/se)^2$] were computed to evaluate instrumental strength. Furthermore, we calculated the proportion of total variance in the exposure explained by each instrument ($R^2$) separately[26].

For gene-outcome associations, summary statistics (standardized regression coefficients for the SNP-metabolic biomarker association) for individual metabolic biomarkers were obtained from the MR-Base database by the MRC Integrative Epidemiology Unit at the University of Bristol (*https://gwas.mrcieu.ac.uk/*). In addition, we performed new GWAS on 9 identified metabolic PCs. We used the BOLT-LMM software with adjustment for age, sex, genetic principal components. Standardized regression coefficients for the SNP-metabolic PCs were used in the MR analyses.

### *Mendelian randomization analysis*

The primary MR analysis was performed using the inverse-variance weighted (IVW) method to combine the SNP-specific Wald ratios for both individual metabolic biomarkers and metabolic PCs, assuming all instrumental variables are valid[27]. Results were expressed as changes in SD of each biomarker or metabolic PC per one-SD decrease in genetically-predicted mtDNA-CN. When the MR assumptions were met, this estimate approximated the causal effect. Sensitivity analyses accounting for pleiotropy were conducted, including Weighted-Median Estimator and MR-Egger regression[28,29], both of which assume the instrumental variable assumptions are valis for at least half of the SNPs. 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 plei-

**Table 1 Baseline charactertistics of the study participants**

| Variables | N = 61,186 |
|---|---|
| Age (years) | 56.5 (8.0) |
| Sex (female) | 33,417 (54.6%) |
| White blood cell count ($10^9$/L) | 6.9 (1.9) |
| Platelet count ($10^9$/L) | 253.9 (59.1) |
| Deprivation index | -1.7 (2.9) |
| Body mass index (kg/m$^2$) | 27.2 (4.6) |
| Diastolic blood pressure (mmHg) | 82.4 (10.1) |
| Systolic blood pressure (mmHg) | 138.1 (18.6) |
| Blood pressure-lowering medication | |
| Yes | 10,664 (17.4) |
| No | 50,522 (82.6) |
| Blood chemistry | |
| Total cholesterol (mmol/L) | 5.8 (1.1) |
| HDL (mmol/L) | 1.5 (0.4) |
| LDL (mmol/L) | 3.6 (0.8) |
| Triglycerides (mmol/L) | 1.7 (1.0) |
| Cholesterol lowering medication | |
| Yes | 7,770 (12.7) |
| No | 53,416 (87.3) |
| Physical activity (moderate-vigorous MET hours/week) | 27.0 (34.2) |
| Alcohol consumption frequency | |
| At least three times per week | 28,246 (46.2) |
| Twice or less per week | 32,899 (53.8) |
| Data missing | 41 (0.1) |
| Smoking status | |
| Current | 5,928 (9.7) |
| Previous | 20,810 (34.0) |
| Never | 34,242 (56.0) |
| Data missing | 206 (0.3) |
| Sleep duration (hours) | 7.1 (1.2) |
| Insomnia | |
| Usually | 16,716 (27.3) |
| Sometimes | 29,575 (48.3) |
| Never/rarely | 14,855 (24.3) |
| Data missing | 40 (0.1) |

Data are mean (SD) for continuous variables and frequency (percentage) for categorical variables. Some percentages do not add up to 100 because of rounding. BMI: Body mass index; HDL: high-density lipoprotein cholesterol; LDL: low-density lipoprotein cholesterol; CVD: cardiovascular disease.

**Table 2 Cross-sectional associations of mtDNA-CN with the first 9 metabolomic principal components (PCs)**

| PCs | Main loadings | Main model1 |
|---|---|---|
| 1 | Positive associations with most lipids and lipoproteins (except for extra-large and large HDL particles) and fatty acids. | 0.262 (0.194, 0.330) |
| 2 | Negative associations with most lipids and lipoproteins (except for LDL subclass chylomicrons and triglycerides within each subparticle), and fatty acids. | -0.194 (-0.235, -0.153) |
| 3 | Negatively associated with IDL, LDL, and HDL (but not extra-large HDL); | -0.027 (-0.060, 0.006) |
| 4 | Negative associations with HDL | 0.022 (-0.001, 0.045) |
| 5 | Positive associations with amino acids | -0.049 (-0.066, -0.032) |
| 6 | Negative associations with ketone bodies | -0.02 (-0.035, -0.006) |
| 7 | Negative associations with ketone bodies (associations with lipids and lipoproteins are opposite to the associations of PC6 with lipids and lipoproteins, despite very weak) | -0.031 (-0.043, -0.018) |
| 8 | Positive associations with fatty acids | -0.096 (-0.107, -0.085) |
| 9 | Negative associations with glycolysis-related metabolites | -0.038 (-0.049, -0.027) |

[1]Betas (95%CI) were adjusted for age at recruitment, sex, deprivation index, the first ten genetic principal components, genotyping batch, white blood cell counts and platelet counts, body mass index, smoking status, alcohol consumption, physical activity, sleep duration and insomnia, lipid-lowering medication, blood pressure and hypertensive medication.

otropy through removing outliers[30]. Moreover, we examined the heterogeneity using Cochran's Q statistic among all SNPs within each outcome database.

All the analyses were performed using R (v4.0.2) statistical software (The R Foundation for Statistical Computing, Vienna, Austria). Packages used in the analyses included "mice" for multiple imputations, "TwoSampleMR" for MR analyses. All results were reported as betas with accompanied 95% confidence intervals.

# Results

Characteristics of the study population are presented in **Table 1**. In total, 61,186 unrelated individuals with no history of cardiovascular diseases were included in the analyses with available mtDNA-CN and metabolomic measures. The mean (SD) age was 56.5 (8.0) years, and 54.6% (N= 33,417) were women.

### Individual metabolic biomarker

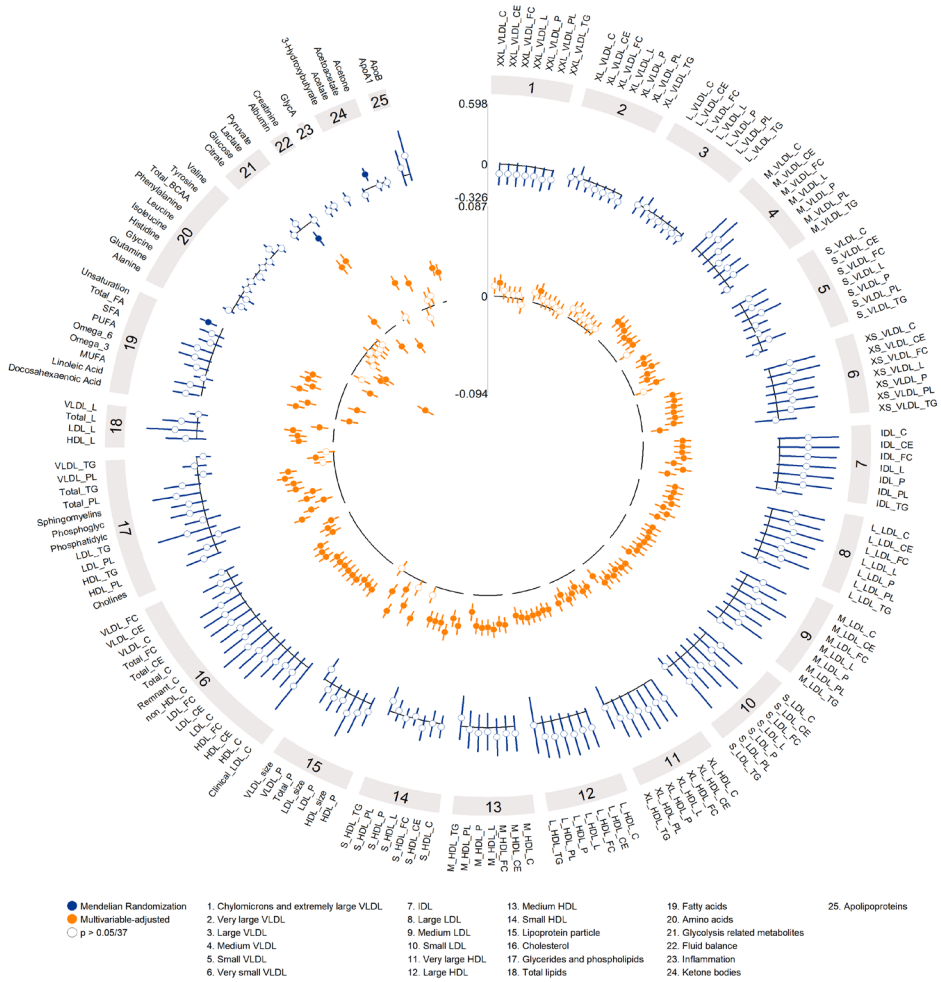


**Figure 1 Associations between low leukocytes mtDNA-CN and 168 NMR-based circulating metabolic biomarkers**

The inner circle represents the associations derived from multivariable linear regression models among 61,168 unrelated individuals from the UK Biobank, adjusted for age, sex, deprivation index, the first ten genetic principal components, genotyping batch, white blood cell counts and platelet counts, body mass index, smoking status, alcohol consumption, physical activity, sleep duration in hours and insomnia, lipid-lowering medication, blood pressure and hypertensive medication. The outer circle represents the estimates obtained from Mendelian randomization inverse-variance weighted method. All point estimates represent the standardized change in metabolic biomarker concentration per standard deviation change in mtDNA-CN. Hollow points indicate the associations were not statistically significant after correction for multiple testing ($p > 0.05/37$, where 37 represents independent metabolomic measures).

After adjustment for potential confounding factors, mtDNA-CN was significantly associated with 130 out of 168 metabolic biomarkers after correcting for mul-

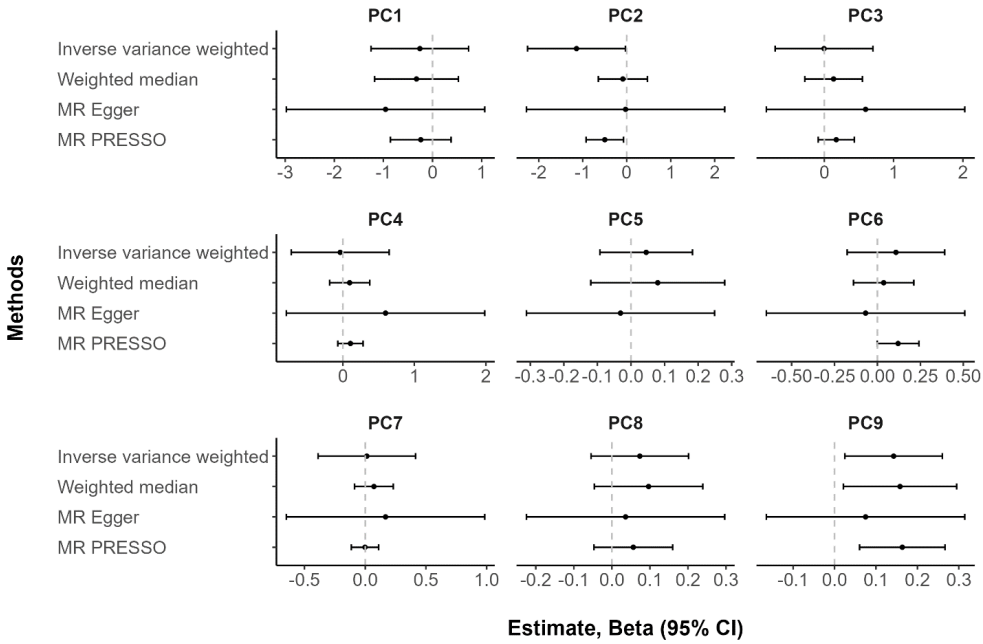


**Figure 2 Associations between leukocyte mtDNA-CN with 9 metabolic principal components (PCs) derived from 168 NMR-based metabolic biomarkers in Mendelian randomization (MR)**

X-axis represents the estimates derived from Mendelian randomization (MR) analyses and are expressed as changes in SD of each metabolic PC per one-SD decrease in genetically-predicted mtDNA-CN. Y-axis represent different MR methods. No outlier was detected for PC5, and therefore no estimate was available in MR-PRESSO for outlier correction.

tiple testing ($p < 0.05/37$), as shown in **Figure 1** (inner circle). A one-SD lower mtDNA-CN was associated with higher concentrations of most lipoprotein (sub) particles, except for chylomicrons, and the associations with triglycerides within each subclass were generally smaller than other (sub)particles. Similarly, low mtDNA-CN was related to higher concentrations of several biomarkers, including most fatty acids except for omega-3 fatty acids and docosahexaenoic acid, amino acids of alanine, glutamine, glycine, and valine, glycolysis related metabolites of citrate, lactate, and pyruvate, and two apolipoproteins. Results from sensitivity analyses after excluding individuals with lipid-lowering medication remained similar to those obtained from the main analyses (**Table S2**).

We included 112 genetic variants available in individual metabolic biomarker GWAS in the MR analyses. F-statistics for each SNP ranged from 15.6 to 634.4, and all the included genetic variants explained approximately 2.0% variation in mtDNA-CN. Detailed information on genetic variants is available in **Table S3**. Genetically-influenced mtDNA-CN were not associated with any lipid (sub)particles or amino acids upon correction for multiple testing (0.05/37) in the IVW analyses **(Figure 1**, outer circle). Notwithstanding, the associations of mtDNA-CN with lipids and lipoproteins in MR analyses generally showed similar trends

To the best of our knowledge, we provide the first evidence for potential causal associations of mtDNA-CN on metabolomic profiling in a large cohort. Previously, the association between mtDNA-CN with lipid dysregulation defined according to biochemically measured lipids levels is inconclusive. In a recent cross-sectional study including 408,361 participants of multiple ancestries (including UKB), lower mtDNA-CN was associated with higher odds of hyperlipidemia as high total cholesterol or triglycerides or use of lipid-lowering medication[9]. Conversely, mtDNA-CN was not associated with any cardiometabolic phenotypes including lipids among two cohorts of UK young and old women[31]. Nevertheless, in case-control settings, mtDNA-CN was lower among individuals with either hyperlipidemia[32] or metabolic syndrome[33-35]. In only one study with merely 310 participants and 44 measured metabolites, it was observed that mtDNA-CN was associated with circulating metabolites indicative of perturbed lipid metabolism[36]. In the present study, we specifically identify the relationship between low mtDNA-CN with several lipids and lipoproteins reflecting a pro-atherogenic metabolic profile through cross-sectional analyses and MR analyses on metabolic PC2. Although MR results from individual metabolic biomarker did not yield any significant associations on lipids and lipoproteins, this might be due to insufficient statistical power and the trend of these associations were similar to those from multivariable-adjusted linear models. Collectively, altered mitochondrial respiration, reflected by low mtDNA-CN may contribute to lipid dysregulation that leads to further cardiovascular diseases.

Glycolysis is the most important pathway for glucose metabolism in cellular metabolism converting glucose to pyruvate for supplying energy via a series of intermediate metabolites. Pyruvate lies at the crossroads of glycolysis and mitochondrial oxidation, and lactate elevates when the flux through glycolysis overwhelms the utilization of pyruvate in the mitochondria. Often, patients with mitochondrial disease suffer from lactic acidosis, a condition in which lactic acids are increased in the body. In experimental studies, pyruvate has been shown to protect against stroke possibly by mitigating inflammation[37,38]. Nevertheless, most of the currently available epidemiological studies have shown positive associations between high levels of either lactate or pyruvate with a higher risk of several cardiovascular endpoints and poor prognosis[39-42], whereas other studies failed to observe similar relationships[42-45]. In our cross-sectional study, we observed an association between low mtDNA-CN and high concentrations of both lactate and pyruvate. However, in MR analyses, genetically-determined low mtDNA-CN was associated with low lactate and high metabolic PC9 (inversely related to glycolysis-related metabolites). These discrepancies are likely owing to that either lactate or pyruvate solely is not a prominent biomarker of mitochondrial function. Interestingly, the lactate-to-pyruvate ratio in conjunction with high lactate has been considered as the most reliable biomarker to differentiate inherited mitochondrial disorders or disturbances in the tricarboxylic acid cycle[46], whether it is also a stable biomarker for compromised mitochondrial function needs to be validated in future research.

There is no evidence favoring a causal link between mtDNA-CN with other metabolomic measures, including amino acids and fatty acids. For amino acids, results from both cross-sectional and MR consistently showed null associations. Although amino acids have been associated to several metabolic-related dis-

eases from both observational and MR studies[46,47], this may indicate that the effect of low mtDNA-CN on chronic diseases is independent of dysregulation of amino acids. For fatty acids, in spite of the solid associations in cross-sectional analyses, MR analyses that are free of reverse causation and most of the confounding factors from either individual biomarker or metabolic PC that are closely related to fatty acids (PC6) failed to validate these associations. Taken together, fatty acids are unlikely the causal intermediate phenotypes from mitochondrial function to chronic diseases.

The main strength of our study is that we used two different epidemiological methods that have different and largely unrelated sources of potential bias, and two different dimensions, i.e. individual metabolic biomarkers and metabolic PCs, to infer causality. Similar effects converged from different approaches largely strengthens the associations and favors potential causality between mtDNA-CN and metabolomic profiling. Our study had several limitations. First, mtDNA-CN were derived from the blood, which might not be representative for those in different tissues and organs. Tissues with higher energy demands harbor more mitochondria and consequently have a higher mtDNA-CN, such as the heart, brain, and skeletal muscles. However, mtDNA-CN measured in the blood and different tissues are not necessarily intercorrelated[48], and therefore the interpretation of the findings should cautiously restrict to the quantitative mtDNA information specifically about the blood. In addition, the calculation of mtDNA-CN from chip arrays might have introduced noise due to the small number of variants. Consequently, we exploited a validated weighted mtDNA-CN approximating what would be estimated from exome sequencing[8]. Furthermore, mtDNA mutations and deletions have been implicated in the pathogenesis of several diseases, particularly atherosclerosis. Nevertheless, due to the lack of data, we could not perform additional analyses and therefore further deep sequencing of the entire mitochondrial genome will facilitate the understanding of mitochondrial function in metabolomic changes.

## Conclusion

Low mtDNA-CN was associated with metabolomic perturbations, particularly a more pro-atherogenic metabolic profile. These findings support the potential of metabolomics to provide new insights into the mitochondrial-induced changes preceding chronic diseases.

# References


1. H. Vakifahmetoglu-Norberg, A. T. Ouchida, E. Norberg. The role of mitochondria in metabolism and cell death. Biochem Biophys Res Commun 2017; 482(3): 426-31.
2. D. E. Handy, J. Loscalzo. Redox regulation of mitochondrial function. Antioxid Redox Signal 2012; 16(11): 1323-67.
3. L. Contreras, I. Drago, E. Zampese, T. Pozzan. Mitochondria: the calcium connection. Biochim Biophys Acta 2010; 1797(6-7): 607-18.
4. D. C. Wallace. A mitochondrial bioenergetic etiology of disease. J Clin Invest 2013; 123(4): 1405-12.
5. R. W. Taylor, D. M. Turnbull. Mitochondrial DNA mutations in human disease. Nat Rev Genet 2005; 6(5): 389-402.
6. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer. The hallmarks of aging. Cell 2013; 153(6): 1194-217.
7. C. López-Otín, G. Kroemer. Hallmarks of Health. Cell 2021; 184(1): 33-63.
8. S. Hägg, J. Jylhävä, Y. Wang, K. Czene, F. Grassmann. Deciphering the genetic and epidemiological landscape of mitochondrial DNA abundance. Hum Genet 2020.
9. X. Liu, R. J. Longchamps, K. L. Wiggins, et al. Association of mitochondrial DNA copy number with cardiometabolic diseases. Cell Genom 2021; 1(1).
10. 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.
11. 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.
12. 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.
13. D. Zhao, T. M. Bartz, N. Sotoodehnia, et al. Mitochondrial DNA copy number and incident atrial fibrillation. BMC Med 2020; 18(1): 246.
14. F. Fazzini, C. Lamina, L. Fendt, et al. Mitochondrial DNA copy number is associated with mortality and infections in a large cohort of patients with chronic kidney disease. Kidney Int 2019; 96(2): 480-8.
15. M. R. Chong, S. Narula, R. Morton, et al. Mitochondrial DNA Copy Number as a Marker and Mediator of Stroke Prognosis: Observational and Mendelian Randomization Analyses. Neurology 2022; 98(5): e470-e82.
16. J. Luo, R. Noordam, J. W. Jukema, et al. Low mitochondrial copy number drives atherogenic cardiovascular disease: evidence from prospective cohort analyses in the UK Biobank combined with Mendelian Randomization. medRxiv; 2021: 2021.07.01.21259854.
17. L. G. Martens, J. Luo, M. J. H. Wermer, et al. The association between lymphocyte mitochondrial DNA abundance and Stroke: a combination of multivariable-adjusted survival and mendelian randomization analyses. medRxiv; 2021: 2021.10.27.21265463.
18. H. Dharuri, A. Demirkan, J. B. van Klinken, et al. Genetics of the human metabolome, what is next? Biochim Biophys Acta 2014; 1842(10): 1923-31.

19. 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.
20. C. Bycroft, C. Freeman, D. Petkova, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018; 562(7726): 203-9.
21. P. Soininen, A. J. Kangas, P. Würtz, et al. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Cardiovascular Epidemiology and Genetics. Circ Cardiovasc Genet. 2015; 8(1): 192-206.
22. J. Kettunen, A. Demirkan, P. Wurtz, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun 2016; 7: 11122.
23. J. Li, L. Ji. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) 2005; 95(3): 221-7.
24. 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.
25. 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.
26. 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.
27. S. Burgess, A. Butterworth, S. G. Thompson. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol 2013; 37(7): 658-65.
28. 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.
29. 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.
30. 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.
31. A. L. Guyatt, K. Burrows, P. A. I. Guthrie, et al. Cardiometabolic phenotypes and mitochondrial DNA copy number in two cohorts of UK women. Mitochondrion 2018; 39: 9-19.
32. C. S. Liu, C. L. Kuo, W. L. Cheng, C. S. Huang, C. F. Lee, Y. H. Wei. Alteration of the copy number of mitochondrial DNA in leukocytes of patients with hyperlipidemia. Ann N Y Acad Sci 2005; 1042: 70-5.
33. D. Révész, J. E. Verhoeven, M. Picard, et al. Associations Between Cellular Aging Markers and Metabolic Syndrome: Findings From the CARDIA Study. J Clin Endocrinol Metab 2018; 103(1): 148-57.
34. J. H. Kim, J. A. Im, D. C. Lee. The relationship between leukocyte mitochondrial DNA contents and metabolic syndrome in postmenopausal women. Menopause 2012; 19(5): 582-7.
35. C. H. Huang, S. L. Su, M. C. Hsieh, et al. Depleted leukocyte mitochondrial DNA copy number in metabolic syndrome. J Atheroscler Thromb 2011; 18(10): 867-73.

36. J. Knez, V. G. Marrachelli, N. Cauwenberghs, et al. Peripheral blood mitochondrial DNA content in relation to circulating metabolites and inflammatory markers: A population study. PLoS One 2017; 12(7): e0181036.
37. Q. Wang, M. van Hoecke, X. N. Tang, et al. Pyruvate protects against experimental stroke via an anti-inflammatory mechanism. Neurobiol Dis 2009; 36(1): 223-31.
38. M. G. Ryou, R. Liu, M. Ren, J. Sun, R. T. Mallet, S. H. Yang. Pyruvate protects the brain against ischemia-reperfusion injury by activating the erythropoietin signaling pathway. Stroke 2012; 43(4): 1101-7.
39. R. Zymliński, J. Biegus, M. Sokolski, et al. Increased blood lactate is prevalent and identifies poor prognosis in patients with acute heart failure without overt peripheral hypoperfusion. Eur J Heart Fail 2018; 20(6): 1011-8.
40. T. Kawase, M. Toyofuku, T. Higashihara, et al. Validation of lactate level as a predictor of early mortality in acute decompensated heart failure patients who entered intensive care unit. J Cardiol 2015; 65(2): 164-70.
41. D. Vojinovic, M. Kalaoja, S. Trompet, et al. Association of circulating metabolites in plasma or serum and risk of stroke: Meta-analysis from seven prospective cohorts. Neurology 2020; 96(8): e1110-23.
42. K. Matsushita, E. K. Williams, M. L. Mongraw-Chaffin, et al. The association of plasma lactate with incident cardiovascular outcomes: the ARIC Study. Am J Epidemiol 2013; 178(3): 401-9.
43. P. Würtz, A. J. Kangas, P. Soininen, D. A. Lawlor, G. Davey Smith, M. Ala-Korpela. Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. Am J Epidemiol 2017; 186(9): 1084-96.
44. D. Ko, E. M. Riles, E. G. Marcos, et al. Metabolomic Profiling in Relation to New-Onset Atrial Fibrillation (from the Framingham Heart Study). Am J Cardiol 2016; 118(10): 1493-6.
45. N. Becerra-Tomás, M. Ruiz-Canela, P. Hernández-Alonso, et al. Glycolysis Metabolites and Risk of Atrial Fibrillation and Heart Failure in the PREDIMED Trial. Metabolites 2021; 11(5).
46. F. G. Debray, G. A. Mitchell, P. Allard, B. H. Robinson, J. A. Hanley, M. Lambert. Diagnostic accuracy of blood lactate-to-pyruvate molar ratio in the differential diagnosis of congenital lactic acidosis. Clin Chem 2007; 53(5): 916-21.
47. L. A. Lotta, R. A. Scott, S. J. Sharp, et al. Genetic Predisposition to an Impaired Metabolism of the Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation Analysis. PLoS Med 2016; 13(11): e1002179.
48. M. Picard. Blood mitochondrial DNA copy number: What are we counting? Mitochondrion 2021; 60: 1-11.

## Supplementary materials

**Table S1** Information of 168 NMR-based metabolomic biomarkers in the UK Biobank (N= 61,186)

**Table S2** Multivariable regression analysis of mtDNA-CN on different metabolomic biomarker excluding individuals with lipid-lowering medication from the UK Biobank

Table S3 Information of genetic variants for mtDNA-CN included in Mendelian randomization analyses

**Table S4** Mendelian randomization of mtDNA-CN on different metabolomic biomarker

**Table S5** Mendelian randomization of mtDNA-CN on metabolomic PCs from the UK Biobank

**Figure S1** The percentage of variance explained by each metabolic principal components in the UK Biobank

**Figure S2** Individual metabolic biomarker loading in metabolic principal components

**The Supplementary materials for this article can be found online at:**
*https://drive.google.com/drive/folders/1d46G5jf6fIZUUlp6aHpjnz74SnLjzQPd?usp=sharing*