UCLA
UCLA Previously Published Works

Title
Association of mitochondrial DNA copy number with cardiometabolic diseases

Permalink
https://escholarship.org/uc/item/5p44c5x6

Journal
Cell Genomics, 1(1)

ISSN
2666-979X

Authors
Liu, Xue
Longchamps, Ryan J
Wiggins, Kerri L
et al.

Publication Date
2021-10-01

DOI
10.1016/j.xgen.2021.100006

Peer reviewed

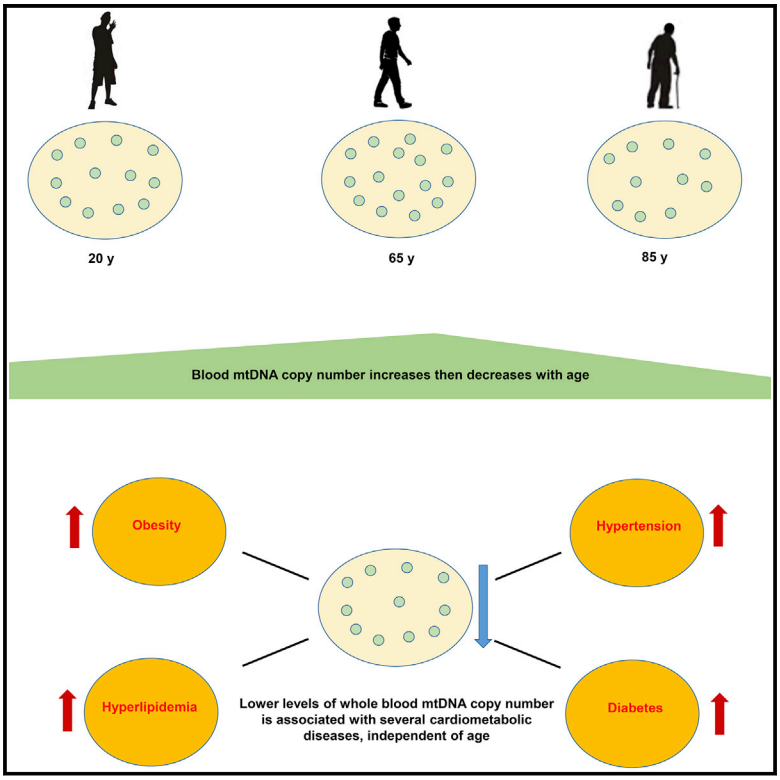
eScholarship.org Powered by the California Digital Library
University of California

Cell Genomics

Article

Association of mitochondrial DNA copy number with cardiometabolic diseases

Graphical abstract



Highlights

- mtDNA copy number in peripheral blood measured in large multi-ancestry cohorts
- mtDNA copy number in blood cells declined with age after age 65
- Lower mtDNA copy number was associated with cardiometabolic disease traits

Authors

Xue Liu, Ryan J. Longchamps, Kerri L. Wiggins, ..., Claudia L. Satizabal, Dan E. Arking, Chunyu Liu

Correspondence

liuc@bu.edu

In brief

Liu et al. examined the association of mitochondrial DNA (mtDNA) copy number (CN) with cardiometabolic traits in 408,361 individuals from TOPMed and UK Biobank, representing the most comprehensive cross-ancestry analyses for these traits. They identify a decline in mtDNA CN in participants older than 65 years and, at lower mtDNA CN levels, age-independent associations with obesity, diabetes, hypertension, and hyperlipidemia.

Liu et al., 2021, Cell Genomics *1*, 100006
October 13, 2021 © 2021 The Author(s).
https://doi.org/10.1016/j.xgen.2021.100006

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Association of mitochondrial DNA copy number with cardiometabolic diseases

Xue Liu,[1] Ryan J. Longchamps,[2] Kerri L. Wiggins,[3] Laura M. Raffield,[4] Lawrence F. Bielak,[5] Wei Zhao,[5] Achilleas Pitsillides,[1] Thomas W. Blackwell,[6] Jie Yao,[7] Xiuqing Guo,[7] Nuzulul Kurniansyah,[8] Bharat Thyagarajan,[9] Nathan Pankratz,[10] Stephen S. Rich,[11] Kent D. Taylor,[7] Patricia A. Peyser,[5] Susan R. Heckbert,[12] Sudha Seshadri,[13,14,15] L. Adrienne Cupples,[1] Eric Boerwinkle,[16,17] Megan L. Grove,[16] Nicholas B. Larson,[18] Jennifer A. Smith,[5]

(Author list continued on next page)

[1]Department of Biostatistics, School of Public Health, Boston University, Boston, MA 02118, USA
[2]McKusick-Nathans Institute, Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[3]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA
[4]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[5]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA
[6]TOPMed Informatics Research Center, University of Michigan, Ann Arbor, MI 48109, USA
[7]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA 90502, USA
[8]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA 02115, USA
[9]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA
[10]Department of Computational Pathology, University of Minnesota, Minneapolis, MN 55455, USA
[11]Department of Public Health Services, Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA
[12]Cardiovascular Health Research Unit and Department of Epidemiology, University of Washington, Seattle, WA 98101, USA
[13]Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA
[14]Framingham Heart Study, NHLBI/NIH, Framingham, MA 01702, USA
[15]Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA
[16]Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, the University of Texas Health Science Center at Houston, Houston, TX 77030, USA
[17]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030, USA
[18]Mayo Clinic College of Medicine and Science, Rochester, MN 55905, USA
[19]Sections of Preventive Medicine and Epidemiology, and Cardiovascular Medicine, Boston University School of Medicine, Boston, MA 02118, USA
[20]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[21]Departments of Family Medicine, Epidemiology, and Global Health, University of Washington, Seattle, WA 98195, USA

(Affiliations continued on next page)

SUMMARY

Mitochondrial DNA (mtDNA) is present in multiple copies in human cells. We evaluated cross-sectional associations of whole-blood mtDNA copy number (CN) with several cardiometabolic disease traits in 408,361 participants of multiple ancestries in TOPMed and UK Biobank. Age showed a threshold association with mtDNA CN: each additional 10 years of age was associated with a 0.03 SD higher level of mtDNA CN ($p = 0.0014$) among younger participants (younger than 65 years) versus a 0.14 SD lower level of mtDNA CN ($p = 1.82 \times 10^{-13}$) among older participants (65 years and older). At lower mtDNA CN levels, we found age-independent associations with increased odds of obesity ($p = 5.6 \times 10^{-238}$), hypertension ($p = 2.8 \times 10^{-50}$), diabetes ($p = 3.6 \times 10^{-7}$), and hyperlipidemia ($p = 6.3 \times 10^{-56}$). The observed decline in mtDNA CN after 65 years of age may be a key to understanding age-related diseases.

INTRODUCTION

Mitochondria convert dietary calories to molecular energy through oxidative phosphorylation (OXPHOS).[1] In addition, mitochondria have essential roles in cellular differentiation, proliferation, reprogramming, and aging.[2–7] Mitochondria contain their own genome (mtDNA), which encodes 37 genes.[1] Multiple copies of mtDNA are present per mitochondrion, and cells contain up to 7,000 mitochondria per cell.[8] The mtDNA copy number (mtDNA CN) correlates with cellular energy generating capacity and metabolic status[9] and, therefore, varies greatly across tissue and cell types.[1,10,11]

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Ramachandran S. Vasan,[14,19] Tamar Sofer,[8,20] Annette L. Fitzpatrick,[21] Myriam Fornage,[22] Jun Ding,[23] Adolfo Correa,[24] Goncalo Abecasis,[6] Bruce M. Psaty,[3,25] James G. Wilson,[26] Daniel Levy,[14,27] Jerome I. Rotter,[7] Joshua C. Bis,[3] TOPMed mtDNA Working Group in NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium,[28] Claudia L. Satizabal,[13,14,15,29] Dan E. Arking,[2,29] and Chunyu Liu[1,14,29,30,*]

[22]Center for Human Genetics, University of Texas Health Science Center at Houston, Houston, TX 77030, USA
[23]Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, NIH, Baltimore, MD 21224, USA
[24]Department of Medicine, University of Mississippi Medical Center, Jackson, MS 39216, USA
[25]Departments of Epidemiology, and Health Services, University of Washington, Seattle, WA 98101, USA
[26]Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS 39216, USA
[27]Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA
[28]See supplemental note for membership list
[29]Senior author
[30]Lead contact
*Correspondence: liuc@bu.edu
https://doi.org/10.1016/j.xgen.2021.100006

Several previous studies have identified a lower level of mtDNA CN in older individuals. This reduced mtDNA CN has been associated with a general decline in health[12–14] and an increased risk of developing cardiovascular disease (CVD) outcomes.[15] Cardiometabolic diseases (CMDs), including obesity, abnormal lipid level and glucose level in plasma, and high blood pressure, are known risk factors for the development of CVD.[16,17] Thus, one mechanism by which a decrease in mtDNA CN could adversely affect not only CVD but also health in general is if the reduction in mtDNA CN was associated with an increase in CMDs. However, the associations of mtDNA CN with the CMD traits have not been consistently reported.[18–20] We investigated associations of mtDNA CN with CMD traits in eight US cohorts from the Trans-Omics for Precision Medicine (TOPMed), representing the most comprehensive cross-ancestry analyses for these traits. These cohort studies included extensive cardiometabolic phenotyping and mtDNA CN estimated from whole-genome sequencing (WGS) data. For validation analyses, we analyzed individuals with whole-exome sequencing (WES) from the UK Biobank (Figure 1).

RESULTS

Characteristics of study participants

The current study included 26,891 participants of eight cohorts from the TOPMed Consortium, including 13,378 European Americans, 8,020 African Americans, 601 Chinese Americans, and 4,892 Hispanic/Latino Americans, as well as 381,470 individuals of European ancestry from the UK Biobank.[21–33] On average, 55% of the study participants were women, and the participants' mean age was 57 years (range, 20–100 years; Table S1). We observed moderate to high heterogeneity in distributions of age, sex, and cardiometabolic phenotypes across cohorts and ancestries. For example, hypertension (HTN), obesity, diabetes, and hyperlipidemia were more prevalent in African Americans than in participants of other ancestry groups (Table S1).

A threshold effect between age and mtDNA CN

The standardized residuals of mtDNA CN were obtained by regressing mtDNA CN on "blood collection year" (see STAR Methods; Figure S1) to study the relationship of mtDNA CN (as the outcome) with age at blood collection. We observed a threshold effect of age on mtDNA CN (Figures 2A and S2–S4). On average, age was associated with a slightly increased level of mtDNA CN (0.032 SD/10 years [95% confidence interval (CI) = 0.013, 0.052], p = 0.0014) from age 20 to 65 years. However, after 65 years, every additional 10 years of age was associated with a 0.14 SD lower level of mtDNA CN (95% CI = −0.18, −0.10; $p = 1.82 \times 10^{-13}$). The relationship between mtDNA CN and age was similar in men and women, although women had higher mtDNA CNs than men had ($\beta = 0.23$; 95% CI = 0.20, 0.26; $p = 7.4 \times 10^{-60}$), as noted previously (Figure S5).[14,34] The threshold effect between age and mtDNA CN remained similar after adjusting for white blood cell (WBC) compositions and platelet count (Figure S2).

Association analyses in European American participants

We generated cohort- and ancestry-specific mtDNA CN standardized residuals by regressing mtDNA CN on age, age-squared, sex, and "blood collection year" in primary analyses in TOPMed cohorts and UK Biobank (see STAR Methods). We then performed cohort- and ancestry-specific association analyses of the standardized mtDNA residuals (as the main predictor) with CMD traits (as the outcome), adjusting for age, age-squared (for blood pressure phenotypes), sex, body mass index (BMI, not for obesity), and smoking status (see STAR Methods).[35–37] Meta-analysis was performed using the fixed-effects inverse-variance method to summarize results based on an *a priori* assumption that there is only one true treatment effect between studies (n = 13,378) (Table 1; Figures 2B, S6, and S7). Because low mtDNA CN was reported to be associated with an increased CMD risk,[18–20,38] we reported β estimates as the change in a CMD outcome variable in response to 1 SD lower mtDNA CN in all analyses. We found that 1 SD decrease in mtDNA CN was significantly associated with 1.10-fold odds of obesity (95% CI = 1.05,1.15; $p = 1.0 \times 10^{-4}$), 1.08-fold odds of HTN (95% CI = 1.03, 1.12; $p = 1.2 \times 10^{-3}$), and 1.22-fold odds of diabetes (95% CI = 1.13, 1.30; $p = 2.8 \times 10^{-8}$), whereas it was not associated with hyperlipidemia (p = 0.13). For continuous traits, 1 SD decrease in mtDNA CN was significantly associated with a 0.030-unit (95% CI = 0.021, 0.039; $p = 2.5 \times 10^{-11}$) increase in log-transformed triglyceride (TRIG) value and a

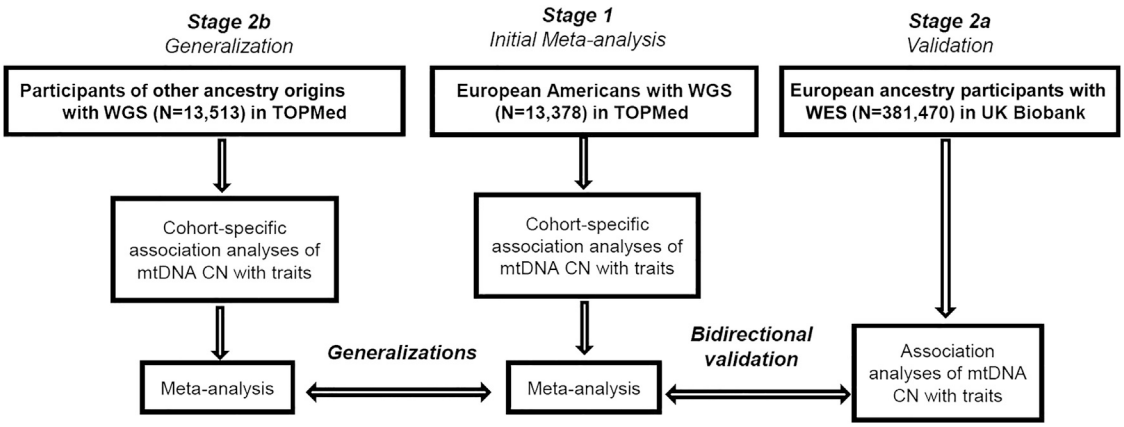


Figure 1. Study design

Association analysis of mtDNA copy number (CN) with cardiometabolic disease traits was performed in cohorts from European Americans (n = 13,378), African Americans (n = 8,020), Chinese Americans (n = 601), and Hispanic and Latino Americans (n = 4,892) in TOPMed and from the UK Biobank (n = 381,470) of European ancestry participants. Meta-analysis was performed using the fixed-effects inverse-variance method to summarize the results among European Americans and among African Americans in TOPMed. WGS, whole-genome sequencing; WES, whole-exome sequencing. See also Tables S1–S4 and Figures S9–S11.

0.20 kg/m^2 (95% CI = 0.11, 0.29; p = 2.0 × 10^{-5}) increase in BMI and was nominally associated with 0.43 mm Hg (95% CI = 0.077, 0.78; p = 0.019) increase in systolic blood pressure (SBP). In contrast, a 1 SD decrease in mtDNA CN was significantly associated with a 0.012-unit (95% CI = −0.017, −0.0071; p = 2.9 × 10^{-6}) decrease in log-transformed high-density lipoprotein (HDL) value and a 0.0075 unit decrease in low-density lipoprotein (LDL) (95% CI= −0.013, −0.0022; p = 0.0058). mtDNA CN was not significantly associated with either diastolic blood pressure (DBP) or fasting plasma glucose (FBG; p > 0.05) in the initial meta-analysis (Table 1; Figures S6 and S7).

Meta-analysis with participants of European ancestry in UK Biobank

We tested seven associations with p < 0.01 from the initial meta-analysis in TOPMed for association in participants of European ancestry in the UK Biobank. Five of those seven associations were validated in the direction of association in the UK Biobank (Table 1; Figures 2B and S6–S8). Compared with those in the initial meta-analysis, analyses in the UK Biobank data yielded larger effect sizes for associations of mtDNA CN with five traits (Table 1). For example, a 1 SD decrease in mtDNA CN was associated with a 1.14-fold (95% CI = 1.13, 1.15) odds of obesity in the UK Biobank versus a 1.10-fold (95% CI = 1.05–1.15) odds of obesity in the initial meta-analysis in TOPMed. mtDNA CN was not significantly associated with hyperlipidemia in TOPMed, whereas it was significantly associated with 1.08-fold odds of hyperlipidemia (95% CI = 1.07, 1.09; p = 2.2 × 10^{-56}) in the UK Biobank. Similarly, mtDNA CN displayed significant association with DBP (mm Hg) (β = 0.24; 95% CI = 0.20, 0.28; p = 5.9 × 10^{-39}) and FBG (mg/dL) (β = 0.23; 95% CI = 0.19, 0.26; p = 1.1 × 10^{-32}) in the UK Biobank, whereas it was not associated with those traits in TOPMed. A low level of mtDNA CN was associated with low levels of log-transformed LDL and HDL in TOPMed, but a 1 SD unit decrease in mtDNA CN was significantly associated with a 0.0034-unit (95% CI = 0.0026, 0.0042; p = 3.0 × 10^{-17}) increase in log-transformed HDL and a 0.0066-unit increase in log-transformed LDL (95% CI = 0.0057, 0.0075; p = 1.7 × 10^{-43}) in the UK Biobank. The association of mtDNA CN with TRIG, however, displayed consistent directionality in both TOPMed and the UK Biobank data (Table 1; Figures S4 and S5). Because of a much larger sample size in the UK Biobank, a meta-analysis combining all participants of European ancestry (total n = 394,848) in the TOPMed and UK Biobank yielded results similar to those of the UK-Biobank-only analysis for all 11 traits (Table 1; Figures S4 and S5).

Comparison of directionality of associations between European and other ancestries

The directionality of associations of mtDNA CN with CMD traits was consistent in African Americans (n = 8,020) (Table S2; Figure S9), Hispanic/Latino Americans (n = 4,892), and Asian Americans (n = 601) compared with that of the participants of European ancestry for most of the CMD traits (Table S3; Figures S10 and S11). In the meta-analysis of African American participants, 1 SD decrease in mtDNA CN was significantly associated with 1.14-fold odds of diabetes (95% CI = 1.07, 1.23; p = 2.0 × 10^{-4}), a 0.75 mm Hg increase in SBP (95% CI = 0.27, 1.22; p = 2.0 × 10^{-3}), and a 0.0077 unit increase in TRIG (95% CI = 0.0024, 0.013; p = 0.0039). In Asian-only participants (n = 601), 1 SD decrease in mtDNA CN was significantly associated with 1.43-fold odds of hyperlipidemia (95% CI = 1.19, 1.72; p = 0.00014). In Hispanic-only TOPMed participants (n = 4,892), 1 SD decrease in mtDNA CN was significantly associated with an increase of odds of diabetes (odds ratio [OR] = 1.14; 95% CI = 1.05, 1.24; p = 0.002), and significantly associated with a 0.016-unit decrease in LDL (95% CI = −0.027, −0.0050; p = 4.3 × 10^{-3}) (Table S3). A pan-ancestry meta-analysis of all participants (n = 408,361), combining the TOPMed and the UK Biobank data, gave rise to similar results for all CMD traits to

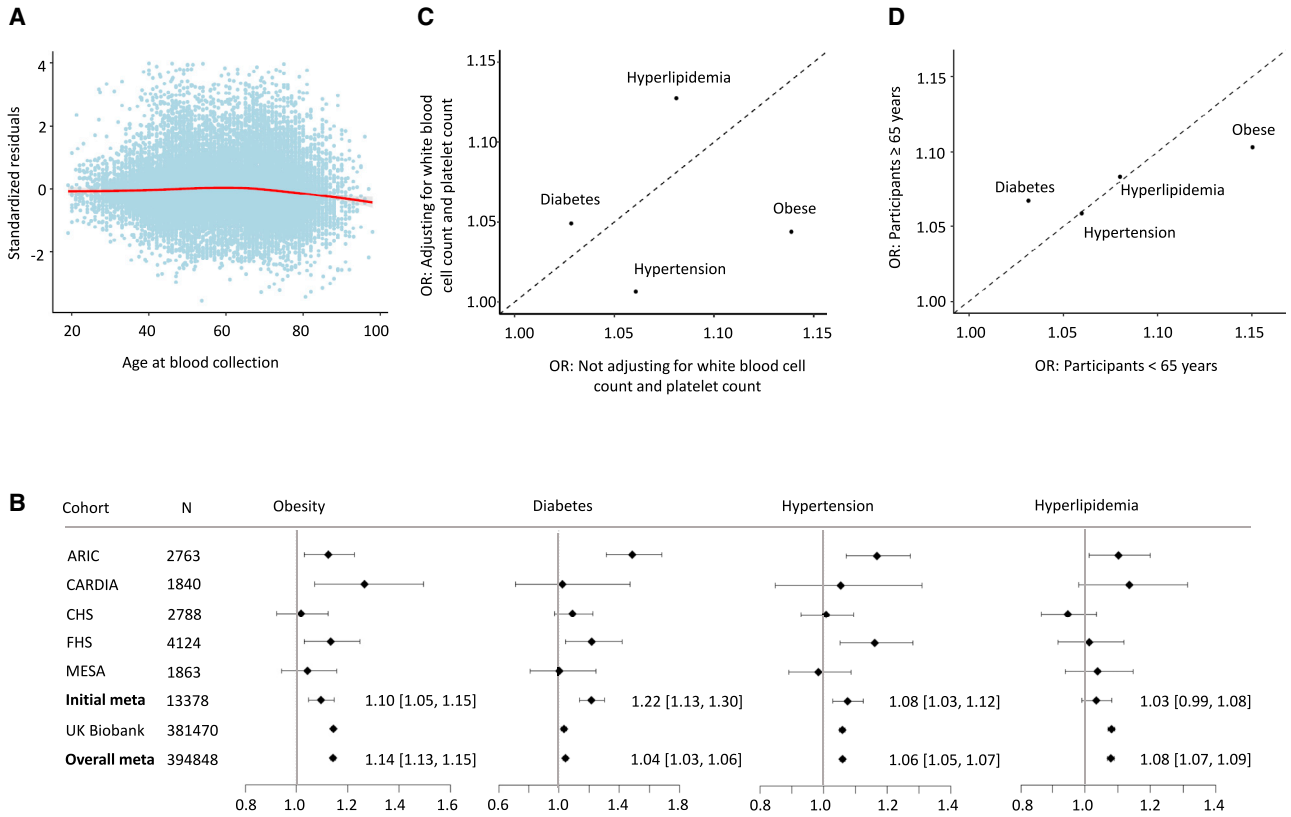


Figure 2. Association of mtDNA CN and cardiometabolic disease traits

(A) The relationship of mtDNA CN with age in TOPMed European American and African American participants (n = 21,398).

(B) Association and meta-analyses of mtDNA CN (n = 394,748) with obesity, hypertension (HTN), diabetes, and hyperlipidemia in TOPMed European Americans (n = 13,378) and in UK Biobank European ancestry participants (n = 381,470). ARIC, Atherosclerosis Risk in Communities study; CARDIA, Coronary Artery Risk Development in Young Adults Study; CHS, Cardiovascular Health Study; FHS, Framingham Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis. A horizontal line represents the 95% confidence intervals of an odds ratio (represented by the box in the middle of the line) for a cohort or meta-analysis; N, sample size; X axis, odds ratio; a number with a square bracket represents an odds ratio with 95% confidence intervals for a meta-analysis.

(C) Comparison of odds ratio of cardiometabolic disease traits in participants with and without adjusting for white blood cell and platelet counts. Meta-analysis (n = 386,526) using inverse-variance weighting, combining the TOPMed European American participants and UK Biobank participants, with cell counts. The odds ratio (OR) corresponds to a 1-SD decrease in the mtDNA CN level.

(D) Age-specific meta-analysis (age younger than 65 years, n = 315,708; age 65 years and older, n = 79,782) combining European ancestry participants in TOPMed and UK Biobank. The effect size estimates are in units of cardiometabolic traits corresponding to a 1-SD decrease in mtDNA CN. See Tables S6–S8 and Figures S2–S8 and S12–S16.

those using the UK-Biobank-only data because of a much larger sample size from the UK Biobank (Table S4).

Accounting for WBC compositions and platelets as covariates

WBC compositions and platelets were available in a subset of participants in TOPMed (n = 12,402) and in all participants of the UK Biobank (n = 381,470). We investigated the relationship of mtDNA CN with WBC compositions (e.g., neutrophil and lymphocyte) and platelet count (see STAR Methods).[39–42] mtDNA CN was inversely associated with the total WBC count and neutrophil count and positively associated with platelet count (Table S5). Moreover, we found that WBC compositions and platelets together explained about 10%–14% of the variation in mtDNA CN, and these blood cell components explained about 0.5%–6% of the variations in CMD traits across a few cohorts (Table S5). Therefore, WBC compositions and platelets are strong confounders for associations between mtDNA CN and several CMD traits. Thus, we compared results between models with and without WBC compositions and platelet count as additional covariates in the same participants. Directionality remained the same for all associations after adjusting for WBC and platelet count in the meta-analysis of participants of European ancestry in the TOPMed and UK Biobank data (Figure 2C; Table S6) and in the ancestry-specific participants and the UK-Biobank-only analyses (Figures S12–S15). Most non-lipid traits, e.g., HTN/SBP and obesity/BMI, displayed a great attenuation in their associations with mtDNA CN after adjusting for WBC compositions and platelet count. In contrast, the associations of mtDNA CN with hyperlipidemia, HDL, and LDL became moderately strengthened after adjusting for WBC and platelet counts (Figure 2C; Table S6).

Table 1. Cross-sectional association and meta-analysis of mtDNA CN with metabolic disease phenotypes

Traits	Initial (TopMed, European ancestry) (N = 13,378)				Validation (UK Biobank, European ancestry) (N = 381,470)				Overall meta-analysis (N = 394,848)			
	OR	95% CI	B	SE	p value	OR	95% CI	β	SE	p value	OR	95% CI	β	SE	p value
Binary outcomes															
Obesity	1.10	1.05–1.15	–	–	1.0×10^{-4}	1.14	1.13–1.15	–	–	9.1×10^{-237}	1.14	1.13–1.15	–	–	5.6×10^{-238}
HTN	1.08	1.03–1.12	–	–	1.2×10^{-3}	1.06	1.05–1.07	–	–	3.8×10^{-48}	1.06	1.05–1.07	–	–	2.8×10^{-50}
Diabetes	1.22	1.13–1.30	–	–	2.8×10^{-8}	1.03	1.02–1.05	–	–	1.0×10^{-4}	1.04	1.03–1.06	–	–	3.6×10^{-7}
Hyperlipidemia	1.03	0.99–1.08	–	–	0.13	1.08	1.07–1.09	–	–	2.2×10^{-56}	1.08	1.07–1.09	–	–	6.3×10^{-56}
Continuous outcomes															
BMI	–	–	0.2	0.046	2.0×10^{-5}	–	–	0.3	0.0077	2.1×10^{-332}	–	–	0.3	0.0076	1.0×10^{-335}
DBP	–	–	0.042	0.1	0.69	–	–	0.24	0.019	5.9×10^{-39}	–	–	0.24	0.018	4.6×10^{-38}
SBP	–	–	0.43	0.18	0.019	–	–	0.64	0.033	7.5×10^{-86}	–	–	0.64	0.032	9.5×10^{-87}
FBG	–	–	0.054	0.093	0.56	–	–	0.23	0.019	1.1×10^{-32}	–	–	0.22	0.019	6.5×10^{-32}
HDL	–	–	−0.012	0.0025	2.9×10^{-6}	–	–	0.0034	0.0004	3.0×10^{-17}	–	–	0.003	0.0004	2.2×10^{-14}
LDL	–	–	−0.0075	0.0027	0.0058	–	–	0.0066	0.00047	1.7×10^{-43}	–	–	0.0061	0.0005	1.7×10^{-39}
TRIG	–	–	0.03	0.0045	2.5×10^{-11}	–	–	0.021	0.00082	4.5×10^{-149}	–	–	0.022	0.0008	2.9×10^{-158}

The β estimates are in units of metabolic traits corresponding to 1 SD lower mtDNA CN. Association analysis of mtDNA CN with metabolic traits was performed in cohorts of European Americans in TOPMed and also in European-ancestry participants in UK Biobank. Meta-analysis with the fixed or random-effects inverse-variance method was applied to summarize the results. See also Tables S2–S4, S6, and S8 and Figures S4–S13.

Interaction analyses

Because of a large sample size, age showed significant interaction with mtDNA CN for multiple traits, including obesity/BMI, FBG, HDL, LDL, and TRIG, and sex showed a significant interaction with mtDNA CN for HTN and BMI in meta-analyses of all participants (Table S7). Because of the threshold effect of age on mtDNA CN, we further performed stratified analyses in younger (younger than 65 years) and older (65 years and older) participants. Hyperlipidemia/lipid traits and HTN displayed consistent effect sizes in associations of mtDNA CN between younger and older age groups. The effect sizes of obesity/BMI and DBP were larger in younger individuals, whereas the effect sizes for type 2 diabetes (T2D)/FBG and SBP were larger in older individuals, although the directionality remained the same between the two age groups (Figures 2D and S16; Table S8).

DISCUSSION

We demonstrate associations of low levels of mtDNA CN in peripheral blood with an increased risk of CMDsin 408,361 individuals of multiple ancestries in TOPMed and UK Biobank, with adjustments made for traditional clinical covariates as well as for blood cell compositions. Cardiometabolic factors are known risks for the development of CVD. Therefore, our association findings further suggest that altered levels of mitochondrial energy production may be involved in the development of a cluster of conditions that increase the risk of CVD. More specifically, the CMD traits that were significantly associated with low levels mtDNA CN—increased odds for obesity, HTN, diabetes, and hyperglycemia—are all components of the metabolic syndrome.[43] Because the metabolic syndrome is the clinical surrogate for insulin resistance, these data suggest that decreasing mtDNA CN may contribute to the insulin resistance accompanying aging.[44]

We identified a threshold effect of age on mtDNA CN, with a large decline in mtDNA CN observed from 65 years of age. Reduced mitochondrial function is considered one of the hallmarks of aging.[45] Therefore, the observed age-related decline in mtDNA CN is potentially important in studying age-related diseases. Our stratified analysis in younger (younger than 65 years) and older (65 years and older) participants found that the effect sizes of associations varied by age groups for six traits in participants of European ancestry.

WBC compositions and platelets are blood biomarkers of systemic inflammation.[8,46] It has been increasingly recognized that a chronic low-grade inflammatory state accompanies CMD risk[47] and is associated with an increasing risk of obesity,[48,49] diabetes,[47,50,51] and HTN,[52–54] whereas it is heterogeneously related to lipid levels.[55] Complementary to those previous findings, this study found complex relationships between WBC compositions/platelets and CMD traits with respect to their heterogeneous directionalities and strengths in the associations of blood compositions with CMD traits. Similarly, WBC compositions (e.g., neutrophils and lymphocytes) and platelets displayed different directionalities in their associations with mtDNA CN, although we observed that a high WBC count was associated with a low level of mtDNA CN, which is consistent with previous findings.[46,56–60] These results indicate that mtDNA CN, WBC

compositions, and platelets may represent an interplay for CVD risk by contributing to those CMD risk factors that are components of the metabolic insulin-resistance syndrome. Further studies are needed to investigate the underlying molecular mechanisms and potential causal pathway among mtDNA CN, inflammation, and CMD risk.

Strengths of the study

This study included a large sample of men and women of multiple ancestries across a wide age range. In TOPMed, mtDNA CN was jointly estimated from WGS. A comprehensive examination showed that the mtDNA CN derived from WGS produced comparable or better results with known correlates (e.g., age and sex) compared with qPCR or other methods (e.g., mtDNA CN estimated from genotyping arrays and whole-exome sequencing).[61] We also performed careful phenotype harmonization and examined several potential confounding variables of mtDNA CN in an association analysis with CMD traits. The UK Biobank, a large prospective cohort study, applied a range of approaches to its sample collection, processing, and assay data monitoring to minimize measurement error in traits and biomarker data.[62] This may partially explain why we observed larger effect sizes in association of mtDNA CN with five of the CMD traits in the UK Biobank data than we did in the TOPMed data. Our combined analyses using TOPMed and UK Biobank provided a large dataset with the comprehensive data collection and quality control needed to enable testing of associations between mtDNA CN and CMD traits through the adult life.

Limitations of the study

Several limitations of the study should be noted. In this study, we used mtDNA CN estimated from whole blood, because peripheral blood is easily accessible, and changes in mtDNA in whole blood are likely to reflect metabolic health across multiple systems. However, this may not be the most-relevant tissue for cardiometabolic targets (e.g., cardiac muscle, skeletal muscle, or adipose tissue) and aging-related (e.g., brain) disease phenotypes. A previous study compared mtDNA CN in whole blood and plasma in the same participants with T2D and found a significant correlation between mtDNA CN of whole blood and plasma in those patients.[63] Another study investigated mtDNA CN in skeletal muscle and cardiac muscle samples through autopsy and heart bypass surgery.[8] However, none of those studies directly compared the mtDNA CN measured from both whole blood and skeletal muscles in the same human samples. A more recent study found that blood-derived mtDNA CN was associated with gene expression across multiple tissues and is predictive for incident neurodegenerative disease, which provides evidence supporting the hypothesis that changes in mtDNA in whole blood may reflect metabolic health across multiple systems.[64]

Second, although we accounted for confounders and known batch effects in mtDNA CN and harmonized metabolic traits, we still observed a moderate to high heterogeneity in the association coefficients in meta-analysis of most of the phenotypes in both ancestry-specific analyses and in TOPMed cohorts. Different distributions of age, sex, and phenotypes across study cohorts may partially explain the heterogeneity in those associations. Unobserved confounding factors, such as experimental conditions for blood draws, DNA extraction, and storage, may also have contributed to the heterogeneity. Finally, we were unable to determine causal relationships between mtDNA CN and CMD traits because of the cross-sectional nature of the study. A reverse causation from a CMD endpoint to mtDNA CN is also possible. A recent study found that mtDNA CN was associated with prevalent diabetes but not with incident diabetes, indicating that diabetes is likely to result in lower levels of mtDNA CN, rather than a lower level mtDNA CN resulting in diabetes.[65] In further studies, it would be of interest to include analyses of mtDNA CN and CMD traits at two time points to provide further insight into associations of aging-related mtDNA CN change with CMD traits.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - mtDNA copy number estimation
  - Cardiometabolic disease phenotypes
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2021.100006.

ACKNOWLEDGMENTS

Detailed acknowledgment for each cohort is included in the supplemental information. In brief, the authors thank the staff and participants of the Atherosclerosis Risk in Communities study, the Cardiovascular Health Study, the Coronary Artery Risk Development in Young Adults Study, the Framingham Heart Study, the Jackson Heart Study, the Genetic Epidemiology Network of Arteriopathy Study, the Hispanic Community Health Study/Study of Latinos, the Multi-Ethnic Study of Atherosclerosis Study, and the UK Biobank for their important contributions. We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. Whole-genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung, and Blood Institute (NHLBI). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (R01HL-117626-02S1; contract HHSN268201800002I). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393-02S1; contract HHSN268201800001I). Additional phenotype harmonization was performed by the current study (R01AG059727). The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the US Department of Health and Human Services.

AUTHOR CONTRIBUTIONS

Data preparation, X.L., R.J.L., D.E.A., C.L., K.L.W., J.C.B., L.M.R., L.F.B., W.Z., J.A.S., A.P., J.Y., X.G., N.K., B.T., M.L.G., N.B.L., A.L.F., M.F., N.P., S.R.H., and T.S.; mtDNA CN estimation, T.W.B., J.D., G.A., and D.E.A.; statistical analyses, X.L., R.J.L., D.E.A., N.K., T.S., and C.L.; manuscript preparation and revision, X.L., R.J.L., J.C.B., D.E.A., C.L., A.L.F., S.S., S.R.H., D.L., J.I.R., and B.M.P.; funding support, C.L., C.L.S., S.S.R., K.D.T., P.A.P., L.A.C., E.B., R.S.V., J.G.W., M.F., J.I.R., A.C., and B.M.P.

DECLARATION OF INTERESTS

The authors declare no competing interests, except for the disclosure from the following authors: Dr. Rocha Abecasis reports grants from National Heart Lung and Blood Institute (NIH) during the conduct of the study and personal fees and other from Regeneron Pharmaceuticals outside of the submitted work; Dr. Cupples reports personal fees from Dyslipidemia Foundation during the conduct of the study and personal fees from Veterans Administration outside of the submitted work; Dr. Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson outside of the submitted work.

Received: September 17, 2020
Revised: February 10, 2021
Accepted: March 30, 2021
Published: October 13, 2021

REFERENCES

1. Voet, D.V.J., and Pratt, C.W. (2005). Fundamentals of Biochemistry, Second Edition (John Wiley and Sons), p. 547.
2. Osellame, L.D., Blacker, T.S., and Duchen, M.R. (2012). Cellular and molecular mechanisms of mitochondrial function. Best Pract. Res. Clin. Endocrinol. Metab. *26*, 711–723.
3. Antico Arciuch, V.G., Elguero, M.E., Poderoso, J.J., and Carreras, M.C. (2012). Mitochondrial regulation of cell cycle and proliferation. Antioxid. Redox Signal. *16*, 1150–1180.
4. Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.
5. Misko, A.L., Sasaki, Y., Tuck, E., Milbrandt, J., and Baloh, R.H. (2012). Mitofusin2 mutations disrupt axonal mitochondrial positioning and promote axon degeneration. J. Neurosci. *32*, 4145–4155.
6. Clayton, D.A., Doda, J.N., and Friedberg, E.C. (1974). The absence of a pyrimidine dimer repair mechanism in mammalian mitochondria. Proc. Natl. Acad. Sci. USA *71*, 2777–2781.
7. Seo, A.Y., Joseph, A.M., Dutta, D., Hwang, J.C., Aris, J.P., and Leeuwenburgh, C. (2010). New insights into the role of mitochondria in aging: mitochondrial dynamics and more. J. Cell Sci. *123*, 2533–2542.
8. Miller, F.J., Rosenfeldt, F.L., Zhang, C., Linnane, A.W., and Nagley, P. (2003). Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. Nucleic Acids Res. *31*, e61.
9. St John, J.C. (2016). Mitochondrial DNA copy number and replication in reprogramming and differentiation. Semin. Cell Dev. Biol. *52*, 93–101.
10. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (1994). Molecular Biology of the Cell (Garland Publishing).
11. Moyes, C.D., Battersby, B.J., and Leary, S.C. (1998). Regulation of muscle mitochondrial design. J. Exp. Biol. *201*, 299–307.
12. Mengel-From, J., Thinggaard, M., Dalgård, C., Kyvik, K.O., Christensen, K., and Christiansen, L. (2014). Mitochondrial DNA copy number in peripheral blood cells declines with age and is associated with general health among elderly. Hum. Genet. *133*, 1149–1159.
13. Zhang, R., Wang, Y., Ye, K., Picard, M., and Gu, Z. (2017). Independent impacts of aging on mitochondrial DNA quantity and quality in humans. BMC Genomics *18*, 890.
14. Ashar, F.N., Moes, A., Moore, A.Z., Grove, M.L., Chaves, P.H.M., Coresh, J., Newman, A.B., Matteini, A.M., Bandeen-Roche, K., Boerwinkle, E., et al. (2015). Association of mitochondrial DNA levels with frailty and all-cause mortality. J. Mol. Med. (Berl.) *93*, 177–186.
15. Ashar, F.N., Zhang, Y., Longchamps, R.J., Lane, J., Moes, A., Grove, M.L., Mychaleckyj, J.C., Taylor, K.D., Coresh, J., Rotter, J.I., et al. (2017). Association of mitochondrial DNA copy number with cardiovascular disease. JAMA Cardiol. *2*, 1247–1255.
16. Leon, B.M., and Maddox, T.M. (2015). Diabetes and cardiovascular disease: Epidemiology, biological mechanisms, treatment recommendations and future research. World J. Diabetes *6*, 1246–1258.
17. Tune, J.D., Goodwill, A.G., Sassoon, D.J., and Mather, K.J. (2017). Cardiovascular consequences of metabolic syndrome. Transl. Res. *183*, 57–70.
18. Xu, F.X., Zhou, X., Shen, F., Pang, R., and Liu, S.M. (2012). Decreased peripheral blood mitochondrial DNA content is related to HbA1c, fasting plasma glucose level and age of onset in type 2 diabetes mellitus. Diabet. Med. *29*, e47–e54.
19. Lee, J.Y., Lee, D.C., Im, J.A., and Lee, J.W. (2014). Mitochondrial DNA copy number in peripheral blood is independently associated with visceral fat accumulation in healthy young adults. Int. J. Endocrinol. *2014*, 586017.
20. Guyatt, A.L., Burrows, K., Guthrie, P.A.I., Ring, S., McArdle, W., Day, I.N.M., Ascione, R., Lawlor, D.A., Gaunt, T.R., and Rodriguez, S. (2018). Cardiometabolic phenotypes and mitochondrial DNA copy number in two cohorts of UK women. Mitochondrion *39*, 9–19.
21. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2019). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. bioRxiv, 563866.
22. Friedman, G.D., Tekawa, I., Grimm, R.H., Manolio, T., Shannon, S.G., and Sidney, S. (1990). The leucocyte count: correlates and relationship to coronary risk factors: the CARDIA study. Int. J. Epidemiol. *19*, 889–893.
23. Fried, L.P., Borhani, N.O., Enright, P., Furberg, C.D., Gardin, J.M., Kronmal, R.A., Kuller, L.H., Manolio, T.A., Mittelmark, M.B., Newman, A., et al. (1991). The Cardiovascular Health Study: Design and rationale. Ann. Epidemiol. *1*, 263–276.
24. Dawber, T.R., Meadors, G.F., and Moore, F.E., Jr. (1951). Epidemiological approaches to heart disease: the Framingham Study. Am. J. Public Health Nations Health *41*, 279–281.
25. Feinleib, M., Kannel, W.B., Garrison, R.J., McNamara, P.M., and Castelli, W.P. (1975). The Framingham Offspring Study: Design and preliminary data. Prev. Med. *4*, 518–525.
26. Splansky, G.L., Corey, D., Yang, Q., Atwood, L.D., Cupples, L.A., Benjamin, E.J., D'Agostino, R.B., Sr., Fox, C.S., Larson, M.G., Murabito, J.M., et al. (2007). The third generation cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: Design, recruitment, and initial examination. Am. J. Epidemiol. *165*, 1328–1335.
27. Daniels, P.R., Kardia, S.L., Hanis, C.L., Brown, C.A., Hutchinson, R., Boerwinkle, E., and Turner, S.T.; Genetic Epidemiology Network of Arteriopathy study (2004). Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. Am. J. Med. *116*, 676–681.
28. Lavange, L.M., Kalsbeek, W.D., Sorlie, P.D., Avilés-Santa, L.M., Kaplan, R.C., Barnhart, J., Liu, K., Giachello, A., Lee, D.J., Ryan, J., et al. (2010). Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. Ann. Epidemiol. *20*, 642–649.
29. Sempos, C.T., Bild, D.E., and Manolio, T.A. (1999). Overview of the Jackson Heart Study: a study of cardiovascular diseases in African American men and women. Am. J. Med. Sci. *317*, 142–146.
30. Wilson, J.G., Rotimi, C.N., Ekunwe, L., Royal, C.D., Crump, M.E., Wyatt, S.B., Steffes, M.W., Adeyemo, A., Zhou, J., Taylor, H.A., Jr., et al.

(2005). Study design for genetic analysis in the Jackson Heart Study. Ethn. Dis. *15* (*suppl 6*), S6–S37.

31. Bild, D.E., Bluemke, D.A., Burke, G.L., Detrano, R., Diez Roux, A.V., Folsom, A.R., Greenland, P., Jacob, D.R., Jr., Kronmal, R., Liu, K., et al. (2002). Multi-ethnic study of atherosclerosis: Objectives and design. Am. J. Epidemiol. *156*, 871–881.
32. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. *12*, e1001779.
33. ARIC Investigators (1989). The Atherosclerosis Risk in Communities (ARIC) Study: Design and objectives. Am. J. Epidemiol. *129*, 687–702.
34. Ding, J., Sidore, C., Butler, T.J., Wing, M.K., Qian, Y., Meirelles, O., Busonero, F., Tsoi, L.C., Maschio, A., Angius, A., et al. (2015). Assessing mitochondrial DNA variation and copy number in lymphocytes of ~2,000 Sardinians using tailored sequencing analysis tools. PLoS Genet. *11*, e1005306.
35. Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Mägi, R., Strawbridge, R.J., Rehnberg, E., Gustafsson, S., et al.; DIAbetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat. Genet. *44*, 991–1005.
36. Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., et al. (2009). Genome-wide association study of blood pressure and hypertension. Nat. Genet. *41*, 677–687.
37. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., et al.; Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.
38. Lee, H.K., Song, J.H., Shin, C.S., Park, D.J., Park, K.S., Lee, K.U., and Koh, C.S. (1998). Decreased mitochondrial DNA content in peripheral blood precedes the development of non-insulin-dependent diabetes mellitus. Diabetes Res. Clin. Pract. *42*, 161–167.
39. Tin, A., Grams, M.E., Ashar, F.N., Lane, J.A., Rosenberg, A.Z., Grove, M.L., Boerwinkle, E., Selvin, E., Coresh, J., Pankratz, N., and Arking, D.E. (2016). Association between mitochondrial DNA copy number in peripheral blood and incident CKD in the Atherosclerosis Risk in Communities Study. J. Am. Soc. Nephrol. *27*, 2467–2473.
40. Houseman, E.A., Accomando, W.P., Koestler, D.C., Christensen, B.C., Marsit, C.J., Nelson, H.H., Wiencke, J.K., and Kelsey, K.T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics *13*, 86.
41. Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). Wiley Interdiscip. Rev. Comput. Stat. *2*, 97–106.
42. Liu, C., Dupuis, J., Larson, M.G., and Levy, D. (2013). Association testing of the mitochondrial genome using pedigree data. Genet. Epidemiol. *37*, 239–247.
43. Huang, P.L. (2009). A comprehensive definition for metabolic syndrome. Dis. Model. Mech. *2*, 231–237.
44. Barzilai, N., and Ferrucci, L. (2012). Insulin resistance and aging: A cause or a protective response? J. Gerontol. A Biol. Sci. Med. Sci. *67*, 1329–1331.
45. López-Otín, C., Blasco, M.A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. Cell *153*, 1194–1217.
46. Esposito, K., and Giugliano, D. (2004). The metabolic syndrome and inflammation: association or causation? Nutr. Metab. Cardiovasc. Dis. *14*, 228–232.
47. Vozarova, B., Weyer, C., Lindsay, R.S., Pratley, R.E., Bogardus, C., and Tataranni, P.A. (2002). High white blood cell count is associated with a worsening of insulin sensitivity and predicts the development of type 2 diabetes. Diabetes *51*, 455–461.
48. Dixon, J.B., and O'Brien, P.E. (2006). Obesity and the white blood cell count: Changes with sustained weight loss. Obes. Surg. *16*, 251–257.
49. Kullo, I.J., Hensrud, D.D., and Allison, T.G. (2002). Comparison of numbers of circulating blood monocytes in men grouped by body mass index (<25, 25 to <30, > or =30). Am. J. Cardiol. *89*, 1441–1443.
50. Ohshita, K., Yamane, K., Hanafusa, M., Mori, H., Mito, K., Okubo, M., Hara, H., and Kohno, N. (2004). Elevated white blood cell count in subjects with impaired glucose tolerance. Diabetes Care *27*, 491–496.
51. Twig, G., Afek, A., Shamiss, A., Derazne, E., Tzur, D., Gordon, B., and Tirosh, A. (2013). White blood cells count and incidence of type 2 diabetes in young men. Diabetes Care *36*, 276–282.
52. Orakzai, R.H., Orakzai, S.H., Nasir, K., Santos, R.D., Rana, J.S., Pimentel, I., Carvalho, J.A.M., Meneghello, R., and Blumenthal, R.S. (2006). Association of white blood cell count with systolic blood pressure within the normotensive range. J. Hum. Hypertens. *20*, 341–347.
53. Karthikeyan, V.J., and Lip, G.Y.H. (2006). White blood cell count and hypertension. J. Hum. Hypertens. *20*, 310–312.
54. Gillum, R.F., and Mussolino, M.E. (1994). White blood cell count and hypertension incidence. The NHANES I Epidemiologic Follow-up Study. J. Clin. Epidemiol. *47*, 911–919.
55. Lai, Y.C., Woollard, K.J., McClelland, R.L., Allison, M.A., Rye, K.A., Ong, K.L., and Cochran, B.J. (2019). The association of plasma lipids with white blood cell counts: Results from the Multi-Ethnic Study of Atherosclerosis. J. Clin. Lipidol. *13*, 812–820.
56. Wu, I.C., Lin, C.C., Liu, C.S., Hsu, C.C., Chen, C.Y., and Hsiung, C.A. (2017). Interrelations between mitochondrial DNA copy number and inflammation in older adults. J. Gerontol. A Biol. Sci. Med. Sci. *72*, 937–944.
57. Knez, J., Marrachelli, V.G., Cauwenberghs, N., Winckelmans, E., Zhang, Z., Thijs, L., Brguljan-Hitij, J., Plusquin, M., Delles, C., Monleon, D., et al. (2017). Peripheral blood mitochondrial DNA content in relation to circulating metabolites and inflammatory markers: A population study. PLoS ONE *12*, e0181036.
58. Sharma, P. (2011). Inflammation and the metabolic syndrome. Indian J. Clin. Biochem. *26*, 317–318.
59. Saltiel, A.R., and Olefsky, J.M. (2017). Inflammatory mechanisms linking obesity and metabolic disease. J. Clin. Invest. *127*, 1–4.
60. Paoletti, R., Bolego, C., Poli, A., and Cignarella, A. (2006). Metabolic syndrome, inflammation and atherosclerosis. Vasc. Health Risk Manag. *2*, 145–152.
61. Longchamps, R.J., Castellani, C.A., Yang, S.Y., Newcomb, C.E., Sumpter, J.A., Lane, J., Grove, M.L., Guallar, E., Pankratz, N., Taylor, K.D., et al. (2020). Evaluation of mitochondrial DNA copy number estimation techniques. PLoS ONE *15*, e0228166.
62. UK Biobank (2019). Biomarker assay quality procedures: Approaches used to minimise systematic and random errors (and the wider epidemiological implications). Version 1.2. https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/biomarker_issues.pdf.
63. Rosa, H.S., Ajaz, S., Gnudi, L., and Malik, A.N. (2020). A case for measuring both cellular and cell-free mitochondrial DNA as a disease biomarker in human blood. FASEB J. *34*, 12278–12288.
64. Yang, S.Y., Castellani, C.A., Longchamps, R.J., Pillalamarri, V.K., O'Rourke, B., Guallar, E., and Arking, D.E. (2021). Blood-derived mitochondrial DNA copy number is associated with gene expression across multiple tissues and is predictive for incident neurodegenerative disease. Genome Res. *31*, 349–358.
65. DeBarmore, B., Longchamps, R.J., Zhang, Y., Kalyani, R.R., Guallar, E., Arking, D.E., Selvin, E., and Young, J.H. (2020). Mitochondrial DNA copy number and diabetes: the Atherosclerosis Risk in Communities (ARIC) study. BMJ Open Diabetes Res. Care *8*, e001204.

STAR★METHODS

KEY RESOURCES TABLE

RESOURCES	SOURCE	IDENTIFIER
Deposited data		
Atherosclerosis Risk in Communities study	TOPMed	phs001211.v3.p2.c1
Coronary Artery Risk Development in Young Adults Study	TOPMed	phs001612
The Cardiovascular Health Study	TOPMed	phs001368.v1.p1
The Framingham Heart Study	TOPMed	phs000974
The Genetic Epidemiology Network of Arteriopathy	TOPMed	phs001345.v1.p1
Hispanic Community Health Study/Study of Latinos	TOPMed	phs001395
The Jackson Heart Study	TOPMed	phs000964.v1.p1
Multi-Ethnic Study of Atherosclerosis Study	TOPMed	phs001416.v1.p1
The UK Biobank	UK Biobank	Application Number 17731
Other		
mtDNA copy number estimation by mitoAnalyzer software (https://lgsun.irp.nia.nih.gov/hsgu/software/mitoAnalyzer/)	Ding et al.[34]	https://www.nia.nih.gov/
Summary data	This paper	This paper
Analysis codes	This paper	https://github.com/chunyuliu1/mtDNA-copy-number-and-cardiometabolic-traits

RESOURCE AVAILABILITY

Lead contact

Further information and requests for data, resources, and reagents should be directed to and will be fulfilled by the lead contact, Chunyu Liu (liuc@bu.edu).

Materials availability

This study did not generate new unique reagents. mtDNA CN and phenotype data are available in the Database of Genotypes and Phenotypes (dbGaP) upon request (below).

Data and code availability

This study did not generate any unique datasets. For the US cohorts, whole-genome sequencing data were generated by the Trans-Omics for Precision Medicine (TOPMed) program supported by National Heart, Lung, and Blood Institute. For TOPMed cohorts, mtDNA CN and phenotype data are available in the dbGaP upon request. The steps to request dbGAP access includes 1) obtain eRA Commons account, 2) obtain dbGaP access, 3) obtain access to Research Project through dbGaP, 4) grant access to individuals to your lab, and 5) log into GDC data portal. The detailed instructions can be found at the following link: (https://gdc.cancer.gov/access-data/obtaining-access-controlled-data). The access numbers are as follows: The Atherosclerosis Risk in Communities study, dbGaP: phs001211; The Coronary Artery Risk Development in Young Adults Study, dbGaP: phs001612; The Cardiovascular Health Study, dbGaP: phs001368; The Framingham Heart Study, dbGaP: phs000974; The Genetic Epidemiology Network of Arteriopathy, dbGaP: phs001345; The Hispanic Community Health Study/Study of Latinos, dbGaP: phs001395; The Jackson Heart Study, dbGaP: phs000964; The Multi-Ethnic Study of Atherosclerosis Study, dbGaP: phs001416; The UK Biobank data (whole-exome sequencing and phenotype data) were downloaded at https://www.ukbiobank.ac.uk/, UKBB: 17731. Codes for association analyses in TOPMed cohorts and estimation of mtDNA CN in the UK Biobank can be accessed at https://github.com/chunyuliu1/mtDNA-copy-number-and-cardiometabolic-traits.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

This study only included human participants from prospective cohort studies. We included eight cohorts from the NHLBI's TOPMed program.[21] These eight cohorts included Atherosclerosis Risk in Communities study (ARIC) (n = 2,964), Coronary Artery Risk Development in Young Adults Study (CARDIA) (n = 3,452), The Cardiovascular Health Study (CHS) (n = 3,493), The Framingham Heart

Study (FHS) (n = 4,124), The Genetic Epidemiology Network of Arteriopathy (GENOA) (n = 1,234), Hispanic Community Health Study/Study of Latinos (HCHS/SOL) (n = 3,868), The Jackson Heart Study (JHS) (n = 3,160), and Multi-Ethnic Study of Atherosclerosis Study (MESA) (n = 4,596). ARIC is a prospective epidemiologic study conducted in four communities which are Forsyth County, NC; Jackson, MS; the northwest suburbs of Minneapolis, MN; and Washington County, MD.[33] Focusing on cardiovascular disease outcomes, event adjudication through 2017 consisted of expert committee review of death certificates, hospital records and telephone interviews. Buffy coat was purified using the Gentra Puregene Blood Kit (QIAGEN) using blood samples collected from several health exam visits. mtDNA-CN was available for 2,964 participants of European Americans and African Americans with WGS from TOPMed. CARDIA is a prospective cohort study which was initiated in 1984 to investigate life-style and other factors that influence cardiovascular disease and their risk factors during young adulthood. The study recruited and examined 5,116 African American and European ancestry women and men aged 18-30 years in four urban areas: Birmingham, Alabama; Chicago, Illinois; Minneapolis, Minnesota, and Oakland, California.[22] The initial examination included carefully standardized measurements of major risk factors as well as assessments of psychosocial, dietary, and exercise-related characteristics that might influence them, or that might be independent risk factors. mtDNA CN was available for 3,452 participants with WGS sequencing in TOPMed. CHS is a population based, longitudinal, multicenter study of coronary heart disease and stroke in 5,888 elderly adults aged 65 years and older. The CHS originated in 1988 to recruit participants from four U.S. communities.[23] The original cohort recruited 5,201 participants and 687 predominately African-American participants were recruited at three of the four field centers in 1992. The first exam began in June 1989. A second comprehensive exam began 3 years after the first exam. A total of n = 3,493 CHS participants (mean age 74 and 58% women) with WGS were included in this study. The FHS is a single-site, community-based, prospective study that was initiated in 1948 to investigate the risk factors for CVD.[24] The second generation[25] was recruited in 1971 and the third generation[26] was recruited between 2002 and 2005. The first generation has been examined every two years. The second generation has been examined every 4-8 years. The third generation has had three examinations. A small number of spouse individuals of the second generation was examined at the same time when the third generation had their first examination. A total of 4,196 FHS participants were whole genome sequenced by TOPMed; of those, 376 were the first generation, 2218 were the second generation and 95 were spouses of the second generation participants; and 1507 were the third generation participants. This study included 4,124 FHS participants. GENOA study enrolled sibships in which at least 2 siblings had essential hypertension diagnosed prior to age 60 years. From 1995 to 2000, the first exam enrolled 1583 non-Hispanic white Americans from Rochester, Minnesota, and 1854 African Americans from Jackson, Mississippi.[27] All siblings within the sibship were invited to participate, including both normotensive and hypertensive siblings. The second exam re-recruited 80% of participants from 2000 to 2005. The GENOA data consists of biological samples (DNA, serum, urine) as well as demographic, anthropometric, environmental, clinical, biochemical, physiological, and genetic data for understanding the genetic predictors of diseases of the heart, brain, kidney, and peripheral arteries. This study included 1,234 participants of African Americans. HCHS/SOL is a longitudinal cohort study established in 2008 following Hispanics/Latinos from four US cities: Bronx, NY; Chicago, IL; Miami, FL; San Diego, CA. This study was approved by the IRB in all field centers.[28] This study included 3,868 participants with available whole-genome sequencing data from blood drawn in their first field center visit. Detailed information on HCHS/SOL was provided previously.[28] Statistical methods used to analyze the present data account for the complex study design, including stratified sampling, clustering, and sampling probabilities. We also adjusted for 11 principal components, estimated from the TOPMed DCC. This study included 3,868 individuals with WGS and matched metabolic phenotypes. The JHS cohort is one of the largest prospective, epidemiologic investigation of CVD among African Americans residing in the three counties (Hinds, Madison, and Rankin) that make up the Jackson, Mississippi metropolitan area.[29,30] Data and biologic materials have been collected from 5,306 participants, including a nested family cohort of 1,498 members of 264 families. The age at enrollment for the unrelated cohort was 35-84 years; the family cohort included related individuals > 21 years old. Participants provided extensive medical and social history and had an array of physical and biochemical measurements and diagnostic procedures during a baseline examination (2000-2004), two follow-up examinations (2005-2008 and 2009-2012), and ancillary studies. Samples for genomic DNA were collected during the first two examinations. Consent for genetic studies and broad sharing of genetic data was provided by 3,482 participants. After all quality control procedures, whole genome sequence data are available for 3,406 participants. Follow-up information on vital status, major illnesses or injuries, and hospitalizations to identify intervening clinical events is done annually by phone. Medical records of cardiovascular disease related hospitalizations and death certificates are abstracted and used for adjudication of cardiovascular events and related deaths. MESA (n = 4,596) is a study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease. MESA researchers study a diverse, population-based sample of 6,814 men and women 45-84 years of age and free of prevalent clinical CVD when recruited from six field centers across the United States in 2000-2002.[31] Event adjudication through 2015 consists of expert committee review of death certificates, hospital records and telephone interviews. DNA for mtDNA-CN analyses was isolated from exam 1 peripheral leukocytes using the Gentra Puregene Blood Kit. mtDNA-CN was available for 4,596 individuals (24.1% Black, 22.3% Hispanic, 13.1% Chinese, 40.5% White) derived from TOPMed WGS sequencing. Several of the TOPMed cohorts contained a small number of duplicated participants. After removing the duplicates, this study included 26,890 individuals with WGS from the TOPMed program (67.4% women; age range of 20-100 years; 45.4% European Americans, 32.6% African Americans, 19.6% Hispanic/Latino Americans and 2.4% Chinese Americans) (Table S1). Additionally, we included 381,470 participants of European ancestry from the UK Biobank with WES (54% women; 40-75 years) for validation (Table S1). The UK Biobank is a prospective cohort study, with the aim of improving the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses – including cancer, heart diseases,

stroke, diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dementia. UK Biobank recruited 500,000 individuals from across the United Kingdom aged between 40-69 years in 2006-2010, who have undergone measures, provided blood, urine and saliva samples for future analysis.[32] 381,470 participants with whole exome sequencing and matched metabolic phenotypes were included in this study. This research has been conducted using the UK Biobank Resource under application number UKBB: 17731.

All study participants provided written informed consent for genetic studies. The protocols for WGS and WES were approved by the institutional review boards (IRB) of the participating institutions (supplemental information).

METHOD DETAILS

mtDNA copy number estimation

mtDNA CN estimation in WGS

Whole blood derived DNA was used for WGS from TOPMed sequencing centers. In analyzing sequencing data, the coverage was defined as the number of reads that were mapped to a given nucleotide in the reconstructed sequence. The average coverage was ~39x across samples in TOPMed. The program *fastMitoCalc* of the software package *mitoAnalyzer*[34] was used to estimate mtDNA copy number across TOPMed participants. Because nuclear DNA (nDNA) is diploid, ordinarily inheriting the DNA from two parents, while mitochondrial DNA is haploid, coming only from the mother, the average mtDNA CN per cell was estimated as twice the ratio of the average coverage of mtDNA to the average coverage of the nuclear DNA (nDNA).[34]

mtDNA CN estimation in UK BioBank

Whole blood derived DNA was used for WES from the UK BioBank. In UK Biobank, we started with 49,997 Exome SPB CRAM files (version Jul 2018) downloaded from the UKB data repository, and used Samtools (ver1.9) to extract read summary statistics ('idxstats' command). A custom perl script was used to aggregate the summary statistics from each individual file into the following categories (see perl script and example stats file): 1) Total Reads (sum of columns 3 and 4, across all rows), 2) Mapped Reads (sum of column 3, across all rows), 3) Unmapped Reads (some of column 4 across all rows), 4) Autosomal Reads (sum of column 3, rows 1-22), 5) Chr X, 6) Chr Y, 7) Chr MT, 8) 'Random' Reads (sum of column 3, across rows 26-67), 9) 'Unknown' Reads (sum of column 3 across rows 68-194), 10) EBV Reads, 11) 'Decoy1' Reads (sum of column 3 across rows 196-582), 12) 'Decoy2′ Reads (sum of column 3 across rows 583-2580). Linear regression models were used to adjust for total DNA and potential technical artifacts. Specifically, we used 10-fold cross validation for variable selection, using the 'leaps' R package (version 3.0), with an initial model with chrMT read count as the dependent variable, and 'Total', 'Mapped', 'unknown', 'random', 'decoy1' and 'decoy2′ read counts as the independent variables. For each of the independent variables, we included a natural spline with df = 4 to allow for non-linear effects. The independent variables 'Total', 'unknown', 'decoy1' and 'decoy2′ read counts were selected. We then increased the natural spline df to 15, and then used backward selection to reduction model complexity, requiring $p < 0.005$ to keep a term in the model. The final regression model residuals were generated with the following R (version 3.6.0) code: WES.mtDNA = residuals(lm(chrMT ~ns(Total,df = 3) + ns(unknown,df = 4) + ns(decoy1,df = 7) + decoy2)). Mitochondrial SNP probe intensities were obtained from the "ukb_chrMT_l2r.txt" file downloaded from the UKBiobank, and samples were stratified by array type (UK BiLEVE, Axiom). To correct for potential artifacts and/or batch effects, we generated 250 principal components (PCs) using the 'rpca' command from the 'rsvd' package (version 1.0.3) from autosomal nuclear probes by randomly sampling 5% of probes from either even or odd chromosomes that were required to be present on both array types (n~19,500 probes). Note that we generated the two independent sets of PCs so that we could ensure that probe selection for PCA did not bias results. Prior to PCA, all probe intensities were rank transformed to reduce the impact of any outliers. For each array type, all mitochondrial SNP probes (UKBelieve, n = 181; Axiom, n = 244) along with the 250 PCs were regressed on the 'WES.mtDNA' metric derived as described above. Beta estimates from these analyses were then used to generate fitted values in the full UKBiobank dataset using the 'predict' function ('array.mtDNA'). Given the known impact of age, sex, and cell counts on mtDNA-CN, we first used visual inspection to identify outliers for cell counts: Log(WBC) ≤ 1.25 or ≥ 3; Log(RBC) ≤ 1.4 or ≥ 2; Platelet ≤ 10 or ≥ 500; Log(Lymphocyte) ≤ 0.10 or ≥ 2; Log(Mono) ≥ 0.9; Log(Neutrophil) ≤ 0.75 or ≥ 2.75; Log(Eos) ≥ 0.75; Log(Baso) ≥ 0.45. We then excluded non-Whites, related individuals (used.in.pca.calculation = 0), and cell count outliers and then adjusted for age, sex, and cell counts using a backward regression, starting with a natural spline (df = 4) for each covariate. The final model obtained was ("log_" indicates log-transformed variable):

$$\mathrm{mtDNA-CN} = \mathrm{residuals}\left[\mathrm{lm}\left(\begin{array}{l}\mathrm{array.mtDNA} \sim \mathrm{ns(age, df = 4)} + \\ \mathrm{sex} + \mathrm{ns(log_WBC, df = 4)} + \\ \mathrm{ns(log_RBC, df = 4)} + \mathrm{ns(platelet, df = 4)} + \\ \mathrm{ns(log_lymph, df = 4)} + \mathrm{ns(log_neutrophil, df = 4)} \\ + \mathrm{log_Eos} + \mathrm{log_Baso} + \mathrm{log_NucRBC}\end{array}\right)\right]$$

Beta estimates from these analyses were then used to generate fitted values in the full UK Biobank dataset (n = 381,470) using the 'predict' function. For all analyses, mtDNA-CN was standardized by subtracting the mean and dividing by the standard deviation.

In ARIC, mtDNA CN has also been estimated from low-pass WGS and Affymetrix Genome-Wide Human SNP Array 6.0.[61] We provided association results of mtDNA CN estimated from these two platforms to provide additional information on whether mtDNA CN estimated from different technologies gave rise to consistent results compared to that estimated from WGS (Table S9; supplemental information). These results were not included in any of the meta-analyses and comparisons in the main text.

Cardiometabolic disease phenotypes

Metabolic disease phenotypes were mapped to the health exams when blood was drawn for DNA extraction for mtDNA CN estimates. Our primary analysis focused on four CMD phenotypes – obesity, hypertension (HTN), diabetes, and hyperlipidemia. We analyzed binary traits in the primary analyses for reducing the multiple testing burden. Obesity was defined as body mass index (BMI) $\geq$ 30 (kg/m^2). For the majority, but not all, of the TOPMed cohorts, a therapeutic indication was provided for a medication treatment, but this was not clear for the UK BioBank. T2D was defined as having a fasting blood glucose level of $\geq$ 126 mg/dL or currently receiving medications to lower blood glucose levels to treat diabetes. Hypertension (HTN) was defined as SBP $\geq$ 140 mmHg, or DBP $\geq$ 90 mmHg, or use of antihypertensive medication(s). Hyperlipidemia was defined as fasting total cholesterol (TC) $\geq$ 200 mg/dL or TRIG $\geq$ 150 mg/dL, or use of any lipid-lowering medication.

We also analyzed the association of mtDNA CN with continuous cardiometabolic traits that defined the binary traits: BMI, SBP, DBP, FBG, HDL cholesterol, LDL cholesterol, and TRIG levels. In the analysis of FBG, we excluded individuals with diabetes, defined as glucose value $\geq$ 126 mg/dL and/or taking glucose-lowering or diabetes medications.[35] SBP and DBP values (mmHg) were derived from the averages of two measurements. We added 15 mmHg and 10 mmHg to SBP and DBP, respectively, for individuals taking any BP lowering medications.[36] The TC measurements were divided by 0.8 for individuals using lipid treatment medications.[37] LDL (mg/dL) was calculated as (TC - HDL - TRIG/5) in individuals with TRIG < 400 mg/dL using imputed TC values.[37] In analyses of FBG and lipid levels, we excluded individuals whose fasting status was not established. TRIG, LDL and HDL values were log-transformed to approximate normality. Other continuous outcome variables were not transformed.

QUANTIFICATION AND STATISTICAL ANALYSIS

We used mtDNA CN as the primary independent variable in all association analyses with CMD traits. To identify confounders and covariates, we first examined whether mtDNA CN levels were associated with the 'blood collection year' (i.e., the year when blood was drawn, as a surrogate of batch effects for blood-derived DNA samples) in all participating cohorts. We discovered that 'blood collection year' explained a 0.9% to 16% variation in mtDNA CN (Figure S1). White blood cell (WBC) count, blood differential count and platelets were previously reported to be associated with mtDNA CN.[14,39] To further understand possible confounding effects of these blood components on association analyses, we investigated whether mtDNA CN and CMD traits were associated with total WBC count, blood differential count, and platelet count that were measured or imputed using the Houseman method or a partial least-squares method[40,41] (Table S5). We further examined the effect of age (Figure S2) and sex (Figure S3) on mtDNA CN after adjusting for 'blood collection year'.

Based on observing significant associations of mtDNA CN in relation to 'blood collection year', age, and sex, we generated mtDNA CN residuals by regressing mtDNA CN on age, age squared, sex and blood collection year (as a factored variable) in each cohort for primary analyses. The residuals were standardized to a mean of zero and standard deviation (s.d.) of one, and used as the main predictor in all regression models. In the primary analysis, we used logistic regression (for unrelated individuals) and mixed effects logistic regression model (related individuals) to analyze binary outcomes (i.e., obesity, HTN) in relation to mtDNA CN residuals. Because age, sex, and BMI are important confounders or covariates for cardiometabolic traits, we further adjusted for sex and age as covariates in the analysis of obesity, and adjusted for sex, age, age-squared (only for HTN) and BMI as covariates in the analysis of T2D, hyperlipidemia, and HTN. Smoking was a traditional covariate in association analysis of mtDNA with disease phenotypes. Although we included smoking as an additional covariate, we found that the impact of smoking on associations of mtDNA CN with CMD traits was minimal (supplemental information; Figure S17). We excluded any value in mtDNA CN or a trait measurement if it was beyond 4 standard deviation of the mean of mtDNA CN residuals or a trait. We used linear effects models to analyze continuous outcome variables, adjusting for the same set of covariates as for the respective binary outcomes. For cohorts with family structure, we accounted for maternal lineage as random effects in linear or logistic mixed models. A maternal lineage was defined to include a founder woman with all of her children, and all grandchildren from daughters of the founder woman.[42]

We performed an initial meta-analysis in European American participants in TOPMed with fixed effects inverse variance method based on an *a priori* assumption that there is only one true treatment effect between studies.[14,15] We performed validation analyses using data for European ancestry participants in the UK Biobank (Figure 1). We further compared meta-analysis results in participants of European ancestry to those from other ancestry origins in TOPMed cohorts. Finally, we performed fixed inverse-variance meta-analysis to combine results from the TOPMed and UK Biobank. We used p = 0.0125 (0.05/4) for significance to account for multiple testing for the primary results from the four binary traits, and used p = 0.05/9$\sim$0.006 for significance in analysis of continuous outcomes.

We compared associations between mtDNA CN and individual outcomes in the same participants with and without WBC count, differential count, and platelet count as additional covariates. We further investigated whether sex or age modified the association between mtDNA CN and outcome variables, adjusting for the same set of covariates described in the primary analyses. In these analyses, we generated mtDNA CN residuals by regressing the mtDNA CN on the blood collection year (as a factored variable) in each cohort for the primary analyses. We included an interaction term between mtDNA CN and sex/age in the association analyses. We also performed age-group stratified analyses between mtDNA CN and CMD traits in younger (< 65 years) and older (≥ 65 years) participants (supplemental information). The statistical software R (version 3.6.0) was used for all statistical analyses.