Genome-wide analysis of mitochondrial DNA copy number reveals multiple loci implicated in nucleotide metabolism, platelet activation, and megakaryocyte proliferation

Longchamps RJ[1*], Yang SY[1*], Castellani CA[1,2], Shi W[1], Lane J[3], Grove ML[4], Bartz TM[5], Sarnowski C[6], Burrows K[7,8], Guyatt AL[9], Gaunt TR[7,8], Kacprowski T[10], Yang J[11], De Jager PL[12,13], Yu L[11], CHARGE Aging and Longevity Group, Bergman A[14], Xia R[15], Fornage M[15,16], Feitosa MF[17], Wojczynski MK[17], Kraja AT[17], Province MA[17], Amin N[18], Rivadeneira F[19], Tiemeier H[18,20], Uitterlinden AG[18,19], Broer L[19], Van Meurs JBJ[18,19], Van Duijn CM[18], Raffield LM[21], Lange L[22], Rich SS[23], Lemaitre RN[24], Goodarzi MO[25], Sitlani CM[24], Mak ACY[26], Bennett DA[11], Rodriguez S[7,8], Murabito JM[27], Lunetta KL[6], Sotoodehnia N[28], Atzmon G[29], Kenny Y[30], Barzilai N[31], Brody JA[32], Psaty BM[33], Taylor KD[34], Rotter JI[34], Boerwinkle E[4,35], Pankratz N[3], Arking DE[1]

*** Indicates authors contributed equally to this work.**

1 McKusick-Nathans Institute, Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD

2 Department of Pathology and Laboratory Medicine, Western University, London, ON, Canada

3 Department of Laboratory Medicine and Pathology, University of Minnesota Medical School, Minneapolis, MN

4 Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX

5 Cardiovascular Health Research Unit, Departments of Medicine and Biostatistics, University of Washington, Seattle, WA

6 Department of Biostatistics, Boston University School of Public Health, Boston, MA

7 MRC Integrative Epidemiology Unit at the University of Bristol, University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK

8 Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, UK

9 Department of Health Sciences, University of Leicester, University Road, Leicester, UK

10 Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Greifswald, Germany; Division Data Science in Biomedicine, Peter L. Reichertz Institute for Medical Informatics, TU Braunschweig and Hannover Medical School, Brunswick, Germany

11 Rush Alzheimer's Disease Center & Department of Neurological Sciences, Rush University Medical Center, Chicago, IL

12 Center for Translational and Systems Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY

13 Program in Medical and Population Genetics, Broad Institute, Cambridge, MA

14 Department of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, New York, USA

15 Institute of Molecular Medicine, The University of Texas health Science Center at Houston, Houston TX

16 Human Genetics Center, The University of Texas Health Science Center at Houston, Houston

17 Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine

18 Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands

19 Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

20 Department of Social and Behavioral Science, Harvard T.H. School of Public Health, Boston, USA

21 Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC

22 Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO

23 Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA

24 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA

25 Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA

26 Cardiovascular Research Institute and Institute for Human Genetics, University of California, San Francisco, California

27 Boston University School of Medicine, Boston University, Boston, MA

28 Cardiovascular Health Research Unit, Division of Cardiology, University of Washington, Seattle, WA

29 Department of Natural science, University of Haifa, Haifa, Israel; Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY, 10461, USA.

30 Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, 10461, USA.

31 Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY, 10461, USA.

32 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA

33 Cardiovascular Health Research Unit, Departments of Epidemiology, Medicine and Health Services, University of Washington, Seattle, WA

34 The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA

35 Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX

Abstract

Blood-derived mitochondrial DNA copy number (mtDNA-CN) is a minimally invasive proxy measure of mitochondrial function that exhibits both inter-individual and intercellular variation. While mtDNA-CN has been previously associated with various aging-related diseases, little is known about the genetic factors that may modulate this phenotype. We performed a genome-wide association study (GWAS) in 465,809 individuals of White (European) ancestry from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium and the UK Biobank (UKB). We identified 129 SNPs with statistically significant, independent effects associated with mtDNA-CN across 96 loci. A combination of fine-mapping, variant annotation, co-localization, and gene set enrichment analyses were used to prioritize genes within each of the 129 independent sites. Putative causal genes were enriched for known mitochondrial DNA depletion syndromes ($p = 3.09 \times 10^{-15}$) and the gene ontology (GO) terms for mtDNA metabolism ($p = 1.43 \times 10^{-8}$) and mtDNA replication ($p = 1.2 \times 10^{-7}$). A clustering approach leveraged pleiotropy between mtDNA-CN associated SNPs and 42 mtDNA-CN associated phenotypes to identify functional domains, revealing five distinct groups, including platelet activation, megakaryocyte proliferation, and mtDNA metabolism. In conclusion, in a GWAS of mtDNA-CN conducted in >450,000 individuals, we identified SNPs within loci that implicate novel pathways that provide a framework for defining the underlying mechanisms involved in genetic control of mtDNA-CN.

Introduction

Mitochondria are the cellular organelles primarily responsible for producing the chemical energy required for metabolism, as well as signaling the apoptotic process, maintaining homeostasis, and synthesizing several macromolecules such as lipids, heme and iron-sulfur clusters[1,2].

Mitochondria possess their own genome (mtDNA); a circular, intron-free, double-stranded, haploid, ~ 16.6 kb maternally inherited molecule encoding 37 genes vital for proper mitochondrial function. Due to the integral role of mitochondria in cellular metabolism, mitochondrial dysfunction is known to play a critical role in the underlying etiology of several aging-related diseases[3–5].

Unlike the nuclear genome, a large amount of variation exists in the number of copies of mtDNA present within cells, tissues, and individuals. The relative copy number of mtDNA (mtDNA-CN) has been shown to be positively correlated with oxidative stress[6], energy reserves, and mitochondrial membrane potential[7]. As a minimally invasive proxy measure of mitochondrial dysfunction[8], decreased blood-derived mtDNA-CN has been previously associated with aging-related disease states including frailty[9], cardiovascular disease[10–12], chronic kidney disease[13], neurodegeneration[14,15], and cancer[16].

Although mtDNA-CN measured from whole blood presents itself as an easily accessible and minimally invasive biomarker, cell type composition has been shown to be an important confounder, complicating analyses[17,18]. For example, while platelets generally have fewer mtDNA molecules than leukocytes, the lack of a platelet nuclear genome drastically skews mtDNA-CN estimates. As a result, not only is controlling for cell composition extremely vital for accurate mtDNA-CN estimation, but interpreting the results in relation to the impact of cell composition becomes a necessity[18–20].

Although the comprehensive mechanism through which mtDNA-CN is modulated is largely unknown[21,22], twin studies have estimated broad-sense heritability ~0.65, consistent with moderate genetic control[23]. Several nuclear genes have been shown to directly modulate mtDNA-CN, specifically those within the mtDNA replication machinery such as the mitochondrial

polymerase, *POLG* and *POLG2*[24,25], as well as the mitochondrial DNA helicase, *TWNK*, and the mitochondrial single-stranded binding protein, *mtSSB*[26]. Furthermore, nuclear genes which maintain proper mitochondrial nucleotide supply including *DGUOK* and *TK2* have also been shown to regulate mtDNA-CN[27–29]. To further elucidate the genetic control over mtDNA-CN, several genome-wide association studies (GWAS) of mtDNA-CN have been published[30–33], including a study that was published while the current manuscript was in preparation, analyzing ~300,000 participants from the UK Biobank (UKB), and identifying 50 independent loci[33].

In the present study, we report mtDNA-CN GWAS results from 465,809 individuals across the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium[34] and the UK Biobank (UKB)[35]. Using multiple gene prioritization and functional annotation methods, we assign genes to loci that reach genome-wide significance. Finally, we perform gene expression analyses and gene-set enrichment through PHEWAS-based SNP-clustering to identify functional domains related to mtDNA-CN.

Subjects and Methods

Study Populations

470,579 individuals participated in this GWAS, 465,809 of whom self-identified as White. Participants were derived from 7 population-based cohorts representing the Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) consortium (Avon Longitudinal Study of Parents and Children [ALSPAC], Atherosclerosis Risk in Communities [ARIC], Cardiovascular Health Study [CHS], Multi-Ethnic Study of Atherosclerosis [MESA], Religious Orders Study and Memory and Aging Project [ROSMAP], Study of Health in Pomerania [SHIP]) and from the UK Biobank (UKB) (Supplemental Table 1). Detailed descriptions of each participating cohort, their quality control practices, study level analyses, and ethic statements are available in the

Supplemental Methods. All study participants provided written informed consent and all centers obtained approval from their institutional review boards.

Methods for Mitochondrial DNA Copy Number Estimation (CHARGE cohorts)

qPCR

mtDNA-CN was determined using a quantitative PCR assay as previously described[32,36]. Briefly, the cycle threshold (Ct) value of a nuclear-specific and mitochondrial-specific probe were measured in triplicate for each sample. In CHS, a multiplex assay using the mitochondrial *ND1* probe and nuclear *RPPH1* probe was used, whereas ALSPAC used a mitochondrial probe targeting the D-Loop and a nuclear probe targeting *B2M*. In CHS, we observed plate effects, as well as a linear increase in ΔCt due to the pipetting order of each replicate. These effects were corrected in the analysis using linear mixed model regression, with pipetting order included as a fixed effect and plate as a random effect to create a raw measure of mtDNA-CN. In ALSPAC, run-to-run variability was controlled using 3 calibrator samples added to every plate, to allow for adjustment by a per-plate calibration factor[32].

Microarray

Microarray probe intensities were used to estimate mtDNA-CN using the Genvisis software package[37] as previously described[10,36]. Briefly, Genvisis uses the median mitochondrial probe intensity across all homozygous mitochondrial SNPs as an initial estimate of mtDNA-CN. Technical artifacts such as DNA input quality, DNA input quantity, and hybridization efficiency were captured through either surrogate variable (SV) or principal component (PC) analyses. SVs or PCs were adjusted for through stepwise linear regression by adding successive components until each successive surrogate variable or principal component no longer significantly improved the model.

182

183 ***Whole Genome Sequencing (ARIC)***

184 Whole genome sequencing read counts were used to estimate mtDNA-CN as previously
185 described[36]. Briefly, the total number of reads in a sample were web scraped from the NCBI
186 sequence read archive. Mitochondrial reads were downloaded directly from dbGaP through
187 Samtools (1.3.1). There was no overlap between ARIC microarray and ARIC whole-genome
188 sequencing samples. A ratio of mitochondrial reads to total aligned reads was used as a raw
189 measure of mtDNA-CN.

190

191 ***Adjusting for Covariates***

192 Each method described above represents a raw measure of mtDNA-CN, adjusted for technical
193 artifacts; however, several potential confounding variables (e.g., age, sex, blood cell
194 composition) have been identified previously[18]. Raw mtDNA-CN values were adjusted for white
195 blood cell count in ARIC, SHIP and CHS (which also adjusted for platelet count), depending on
196 available data. Standardized residuals (mean = 0, standard deviation = 1) of mtDNA-CN were
197 used after adjusting for covariates (Supplemental Table 1.

198

199 **Estimation of Mitochondrial DNA Copy Number (UKB)**

200 Due to the availability of more detailed cell count data, as well as a different underlying
201 biochemistry for the Affymetrix Axiom array compared to the genotyping arrays used in the
202 CHARGE cohorts, mtDNA-CN in the UKB was estimated differently (Supplemental Methods).
203 Briefly, mtDNA-CN estimates derived from whole exome sequencing data, available on ~50,000
204 individuals, were generated first using customized Perl scripts to aggregate the number of
205 mapped sequencing reads and correct for covariates through both linear and spline regression
206 models. Concurrently, mitochondrial probe intensities from the Affymetrix Axiom arrays,
207 available on the full ~500,000 UKB cohort, were adjusted for technical artifacts through principal

208 components generated from nuclear probe intensities. Probe intensities were then regressed
209 onto the whole exome sequencing mtDNA-CN metric, and beta estimates from that regression
210 were used to estimate mtDNA-CN in the full UKB cohort. Finally, we used a 10-fold cross
211 validation method to select the cell counts to include in the final model (Supplemental Table 2).
212 The final UKB mtDNA-CN phenotype is the standardized residuals (mean = 0, standard
213 deviation = 1) from a linear model adjusting for covariates (age, sex, cell counts) as described in
214 the Supplemental Methods.
215
216 **Genome-Wide Association Study**
217 For each individual cohort, regression analysis was performed with residualized mtDNA-CN as
218 the dependent variable adjusting for age, sex, and cohort-specific covariates (e.g., principal
219 components, DNA collection site, family structure, cell composition). Cohorts with multiple
220 mtDNA-CN estimation platforms were stratified into separate analyses. Ancestry-stratified meta-
221 analyses were performed using Metasoft software using the Han and Eskin random effects
222 model to control for unobserved heterogeneity due to differences in mtDNA-CN estimation
223 method[38]. Effect size estimates for SNPs were calculated using a random effect meta-analysis
224 from cohort summary statistics, as the Han and Eskin model relaxes the assumption under the
225 null hypothesis without modifying the effect size estimates that occur under the alternative
226 hypothesis[38]. In total, three complementary analyses were performed in self-identified White
227 individuals, (1) a meta-analysis using all available studies, (2) a meta-analysis of studies with
228 available data for cell count adjustments, and (3) an analysis of UKB-only data. As the vast
229 majority of samples are derived from the UKB study, and given the difficulty in interpreting effect
230 size estimates from a random effects model, further downstream analyses were all performed
231 using effect size estimates from UKB-only data.
232
233 **Identification of Independent GWAS Loci**

To identify the initial genome-wide significant (lead) SNPs in each locus, the most significant SNP that passed genome-wide significance ($p < 5 \times 10^{-8}$) within a 1 Mb window was selected. To avoid Type I error, SNPs were only retained for further analyses if there were either (a) at least two genome-wide significant SNPs in the 1 Mb window or (b) if the lead SNP was directly genotyped. Conditional analyses were performed in UKB, where the lead SNPs from the original GWAS were used as additional covariates in order to identify additional independent associations.

Fine-mapping

The susieR package was used to identify all potential causal variants for each independent locus associated with mtDNA CN[39]. UKB imputed genotype data for unrelated White subjects were used and variants were extracted using a 500 kb window around the lead SNP for each locus with minor allele frequency (MAF) > 0.001. 95% credible sets (CS) of SNPs, containing a potential causal variant within a locus, were generated. The minimum absolute correlation within each CS is 0.5 and the scaled prior variance is 0.01. When the CS did not include the lead SNP identified from the GWAS, some of the parameters were slightly relaxed [minimum absolute correlation is 0.2, estimate prior variance is TRUE]. The SNP with the highest posterior inclusion probability (PIP) within each CS was also identified (Supplemental Table 3). With a few exceptions, final lead SNPs were selected by prioritizing initially identified SNPs unless the SNP with the highest PIP had a PIP greater than 0.2 and was 1.75 times larger than the SNP with the second highest PIP.

Functional Annotation and Gene Prioritization

Functional Annotation

ANNOVAR was used for functional annotation of variants identified in the fine-mapping step[40]. First, variants were converted to an ANNOVAR-ready format using the dbSNP version 150

database[41]. Then, variants were annotated with ANNOVAR using the RefSeq Gene database[42]. The annotation for each variant includes the associated gene and region (e.g., exonic, intronic, intergenic). For intergenic variants, ANNOVAR provides flanking genes and the distance to each gene. For exonic variants, annotations also include likely functional consequences (e.g., synonymous/nonsynonymous, insertion/deletion), the gene affected by the variant, and the amino acid sequence change (Supplemental Table 4).

Co-localization Analyses

Co-localization analyses were performed using the approximate Bayes factor method in the R package *coloc*[43]. Briefly, *coloc* utilizes eQTL data and GWAS summary statistics to evaluate the probability that gene expression and GWAS data share a single causal SNP (colocalize). *Coloc* returns multiple posterior probabilities; H0 (no causal variant), H1 (causal variant for gene expression only), H2 (causal variant for mtDNA-CN only), H3 (two distinct causal variants), and H4 (shared causal variant for gene expression and mtDNA-CN). In the event of high H4, we can assume that the gene is potentially causal for the GWAS phenotype of interest. eQTL summary statistics were obtained from the eQTLGen database[44]. Genes with significant associations with lead SNPs were tested for co-localization using variants within a 500 kb window of the sentinel SNP. Occasionally, some of the eQTLGen p-values for certain SNPs were identical due to R's (ver 4.0.3) limitation in handling small numbers. To account for this, if the absolute value for a SNP's z-score association with a gene was greater than 37.02, z-scores were rescaled so that the largest z-score would result in a p-value of 5×10^{-300}. Additionally, a few clearly co-localized genes did not result in high H4 PPs due to the strong effect for each phenotype of a single SNP (Supplemental Figure 1), possibly due to differences in linkage disequilibrium (LD) between the populations. To account for this, we summed mtDNA-CN GWAS p-values and eQTLGen p-values for each SNP and removed the SNP with the lowest combined p-value. Co-localization

285 analyses were then repeated without the lowest SNP. Genes with H4 greater than 50% were
286 classified as genes with significant evidence of co-localization.

287 ***DEPICT***

288 Gene prioritization was performed with Depict, an integrative tool that incorporates gene co-
289 regulation and GWAS data to identify the most likely causal gene at a given locus[45]. Across
290 GWAS SNPs which overlapped with the DEPICT database, we identified SNPs representing
291 119 independent loci with LD pruning defined as $p < 5 \times 10\text{-}8$, $r2 < 0.05$ and > 500 kb from other
292 locus boundaries. Only genes with a nominal p-value of less than 0.05 were considered for
293 downstream prioritization.

294 ***Gene Assignment***

295 To prioritize genes for each identified locus, we utilized functional annotations, eQTL co-
296 localization analyses, and DEPICT gene prioritization results (Supplemental Figure 2). First,
297 genes with missense variants within *SusieR* fine-mapped credible sets were assigned to loci. If
298 loci co-localized with a gene's expression with a posterior probability (PP) of greater than 0.50
299 and there were no other co-localized genes with a PP within 5%, the gene with the highest
300 posterior probability was assigned. If there was still no assigned gene, the most significant
301 DEPICT gene was assigned. If there was no co-localization or DEPICT evidence, the nearest
302 gene was assigned.

303

304 **Gene Set Enrichment Analyses**

305 Using the finalized gene list from the prioritization pipeline, GO and KEGG pathway enrichment
306 analyses were performed using the "goana" and "kegga" functions from the R package *limma*[46],
307 treating all known genes as the background universe[47]. Only one gene per locus was used for
308 "goana" and "kegga" gene set enrichment analysis, prioritizing genes assigned to primary
309 independent hits. If there were multiple assigned genes, one gene was randomly selected to
310 avoid biasing results through loci with multiple genes. To identify an appropriate p-value cutoff,

96 genes were randomly selected from the genome and run through the same enrichment analysis. This permutation was repeated 1000 times to generate a null distribution of the smallest p-values from each permutation. For cluster 5 gene set enrichment analyses, permutation testing used 47 random genes. To ensure robustness of results, gene set enrichment analysis was repeated 50 times with random selection of genes at loci with multiple assigned genes. GO and KEGG terms that passed permutation cutoffs at least 40/50 times were retained.

Gene-based Association Test

We used metaXcan, which employs gene expression prediction models to evaluate associations between phenotypes and gene expression[48]. We obtained pre-calculated expression prediction models and SNP covariance matrices, computed using whole blood from European ancestry individuals in version 7 of the Genotype-Tissue expression (GTEx) database[49]. Using prediction performance p-values of less than 0.05, a total of 6,285 genes were predicted. Of these genes, 74 passed Bonferroni correction of $p < 7.95 \times 10^{-6}$. Gene set enrichment analyses were performed on Bonferroni-significant genes as previously described. REVIGO[50] was used on the "medium" setting (allowed similarity = 0.7) to visualize significantly enriched GO terms.

We used a one-sided Fisher's exact test to test for enrichment of genes that have been previously identified as causal for mtDNA depletion syndromes[51–53].

PHEWAS-based SNP Clustering

mtDNA-CN Phenome-wide Association Study (PHEWAS)

We used the PHEnome Scan ANalysis Tool (PHESANT)[54] to identify mtDNA-CN associated quantitative traits in the UKB. Briefly, we tested for the association of mtDNA-CN with 869 quantitative traits (Supplemental Table 5), limiting analyses to 365,781 White, unrelated

337 individuals (used.in.pca.calculation=1), and excluding individuals with extreme cell count
338 measurements (see Supplementary Methods). Analyses were adjusted for age, sex, and
339 assessment center.
340
341 ***SNP Phenotype Associations***
342 SNP genotypes were regressed on mtDNA-associated quantitative traits using linear
343 regression, adjusted for sex, age with a natural spline (df=2), assessment center, genotyping
344 array, and 40 principal components (provided as part of the UKB data download).
345
346 ***SNP Clustering***
347 To identify distinct clusters of mtDNA-CN GWS SNPs based on phenotypic associations, beta
348 estimates from the SNP phenotype associations were first divided by the beta estimate of the
349 mtDNA-CN SNP-mtDNA-CN association, so that all SNP-phenotype associations are relative to
350 the mtDNA-CN increasing allele and scaled to the effect of the SNP on mtDNA-CN. The
351 adjusted beta estimates were subjected to a dimensionality reduction method, Uniform Manifold
352 and Approximation Projection (UMAP), as implemented in the R package *umap*[55]
353 (random_state=123, n_neighbors=10, n_components=2, n_epochs=5000). SNPs were
354 assigned to clusters using Density Based Clustering of Applications with Noise (DBSCAN) as
355 implemented in the R package *dbscan*[56] (minPts=3). Clusters represent groups of SNPs with
356 similar phenotypic associations.
357
358 To identify reproducible sub-clusters within cluster 5, we ran *umap* 100 times, varying the initial
359 random state (n_neighbors=4, n_epochs=5000, min_dist=0.05, n_components=3). For each
360 *umap* run, sub-clusters were assigned using *dbscan* as described above. Final SNP sub-cluster
361 assignments were determined by identifying SNPs that all mapped to the same cluster >50% of
362 the time (i.e., all pairwise correlations for each SNP within the cluster against all other SNPs in

363 the cluster was >50%). The final umap run for plotting in 2 dimensions used the following
364 parameters: random_state=123, n_neighbors=4, n_epochs=5000, min_dist=0.05.
365
366 ***Phenotype Enrichment and Permutation Testing***
367 To test for enrichment of specific phenotypes within clusters, we compared the median mtDNA-
368 CN scaled phenotype beta estimates within the cluster to the median beta estimates for all
369 SNPs not in the cluster, with significance determined using 20,000 permutations in which cluster
370 assignment was permuted. For multi-test correction of the overall cluster and sub-cluster
371 analyses, we performed 300 permutations of the initial cluster assignment (separately for the
372 cluster and sub-cluster analyses), followed by the comparison of median beta estimates as
373 described above. We retained only the most significant result from across all phenotypes and
374 clusters from each of the 300 permutations, and then selected the 15th most significant value as
375 the study-wide threshold for multi-test corrected significance of $p < 0.05$.
376
377 All statistical analyses were performed using R version 4.0.3
378
379 **Results**
380 **Sample Characteristics**
381 The current study included 465,809 individuals of White (European) ancestry (53.9% female)
382 with an average age of 56.6 yrs (sd = 8.2 yrs) (Supplemental Table 1). Follow-up validation
383 analyses were performed in 4,770 Blacks (60.2% female) with an average age of 61.2 yrs (sd =
384 7.4 yrs). The majority of the data originated from the UKB (93%). The bulk of the DNA used for
385 mtDNA-CN estimation was derived from buffy coat (95.5%) while the rest was derived from
386 peripheral leukocytes (2.2%) or whole blood (2.3%). mtDNA-CN estimated from Affymetrix
387 genotyping arrays consisted of 97.9% of the data while the remainder was derived from qPCR-
388 (1.8%) and WGS (0.3%).

GWAS

Previous work has demonstrated that the method used to measure mtDNA-CN can impact the strength of association[36]. To account for potential differences across studies due to the different mtDNA-CN measurements used, as well the inclusion of blood cell counts in only a subset of the cohorts, we took two approaches. First, we used a random effects model to perform meta-analyses, allowing for different genetic effect size estimates across cohorts. Second, we performed three complementary analyses in individuals who self-identified as White: 1) meta-analysis of all available studies (n = 465,809); 2) meta-analysis of studies with available data for cell count adjustment (n = 456,151); and 3) GWAS of UKB only (n = 440,266). 77 loci were significant in all three meta-analyses, and we identified 93 independent loci, of which 92 were genome-wide significant in the UKB data alone (Supplemental Figure 3, Figure 1). Given that > 90% of the samples come from the UKB study, and the challenge of interpreting effect size estimates from a random effects model, downstream analyses all use effect size estimates from the UKB only (Supplemental Table 6).

The most significant SNP associated with mtDNA-CN was a missense mutation in *LONP1* ($p = 3.00 \times 10^{-141}$), a gene that encodes a mitochondrial protease that can directly bind mtDNA, and has been shown to regulate *TFAM*, a transcription factor involved in mtDNA replication and repair (for review see Gibellini *et al.*)[57].

Fine-mapping and Secondary Hits

To identify additional independent SNPs within novel loci whose effects were masked by the original significant SNP, as well as identify additional loci, we took two approaches. First, a conditional analysis adjusting for the top 93 SNPs from the initial (primary) GWAS run revealed 3 novel loci and 19 additional independent significant SNPs within existing loci. We also performed fine-mapping with SuSiE[39] and discovered an additional 14 independent SNPs within

existing loci. The majority of loci had only one 95% credible set of SNPs; further, twenty of the credible sets contained only one SNP. However, many of the credible sets contained greater than 50 SNPs after fine-mapping, and 12 of the 122 credible sets had a missense SNP as the SNP with the highest PIP in the set. Using these two methods, we identified in total 129 independent SNPs across 96 loci (Supplemental Figure 4).

Associations in African Ancestry (AA) Populations

Examining the 129 SNPs from the Whites-only analysis, 99 were available in the Blacks-only meta-analysis (n = 4770). After multiple testing correction, one of these SNPs was significant (rs73349121, $p = 0.0001$), 9 were nominally significant ($p < 0.05$, with 5 expected), and 58/99 had a direction of effect that was consistent with the White-only analyses (one-sided $p = 0.04$, Figure 2). Despite being under-powered, these results in the Black-only analyses provide evidence for similar genetic effects in a different ancestry.

Gene Prioritization and Enrichment of mtDNA Depletion Syndrome Genes

We integrated results from three different gene prioritization and functional annotation methods (ANNOVAR[40], COLOC[43] , and DEPICT[45]) so that loci with nonsynonymous variants in gene exons were prioritized first, with eQTL co-localization results considered second (Supplemental Table 7), and those from DEPICT (Supplemental Table 8) were considered last (Supplemental Figure 2). For 20 loci, multiple genes were assigned as analyses could not identify a single priority gene (Supplemental Table 9). We noted the identification of a number of mtDNA depletion syndrome genes in the priority list and tested for enrichment of these known causal genes using a one-sided Fisher's exact test. For this analysis, all genes for loci assigned to multiple genes were used, and genes for all primary and secondary loci were considered. Our gene prioritization approach identified 7 of 16 mtDNA depletion genes (Supplemental Table 10), consistent with a highly significant enrichment (one-sided $p = 3.09 \times 10^{-15}$).

441

442 **Gene Set Enrichment Analyses**

443 To avoid bias from a single locus with multiple functionally related genes contributing to a false-
444 positive signal, only one gene per unique locus was used, prioritizing genes assigned to primary
445 loci. One gene was randomly selected for loci with multiple assigned genes. To test for
446 robustness of gene set enrichment results, random selection was repeated 50 times, and only
447 gene sets that were significantly enriched for at least 40 iterations were retained. In all, a total of
448 96 genes were utilized for GO term and KEGG pathway enrichment analyses. Using a
449 Bonferroni-corrected p-value cutoff, 15 gene sets were significantly enriched for all 50 iterations,
450 including mitochondrial DNA metabolic process, mitochondrial DNA replication, coagulation,
451 hemostasis, amyloid-beta clearance, and mitochondrial genome maintenance (Supplemental
452 Table 11). No KEGG terms were significant across multiple iterations.

453

454 **MetaXcan Gene Expression Analysis**

455 As a complementary approach to single-SNP analyses, we explored the associations between
456 mtDNA-CN and predicted gene expression using MetaXcan[48] MetaXcan incorporates multiple
457 SNPs within a locus along with a reference eQTL dataset to generate predicted gene
458 expression levels. As our study estimated mtDNA-CN derived from blood, we used whole blood
459 gene expression eQTLs from the Gene-Tissue Expression (GTEx) consortium[58] to predict gene
460 expression in the UKB dataset. We identified 6,285 genes that had a predicted performance p-
461 value of less than 0.05 (i.e., they had sufficient data to generate robust gene expression levels)
462 and were tested for association with mtDNA-CN. Of these genes, 74 were significantly
463 associated with mtDNA-CN ($p < 7.95 \times 10^{-6}$, Figure 3), including 8 that were not identified
464 through single-SNP analyses. Many of the significant genes have known mitochondrial
465 functions, notably the mtDNA transcription factor *TFAM* ($p = 1.09 \times 10^{-29}$) and mitochondrial
466 exonuclease *MGME1* ($p = 5.87 \times 10^{-23}$), genes known as causal for mtDNA depletion

syndromes[51,52]. Additionally, *LONP1*, *MRPL43*, and *BAK1*, are all genes with known mitochondrial functions[59–61]. Bonferroni significant MetaXcan genes were used for gene enrichment analysis, finding enrichment for "nucleobase metabolic process" ($p = 1.47 \times 10^{-4}$) and "mitochondrial fusion" ($p = 1.86 \times 10^{-4}$, Supplemental Figure 5).

PHEWAS-based SNP Clustering and Gene Set Enrichment

mtDNA-CN is associated with numerous quantitative and qualitative phenotypes, many of which are relevant to aging-related disease[3–5,9,10,13–16]. We hypothesized that this pleiotropy may reflect different underlying functional domains captured by mtDNA-CN that may be reflected in GWAS-identified SNPs and their likely causal genes. To test this hypothesis, we used the UKB data to identify quantitative traits associated with mtDNA-CN and selected 42 highly significant, non-redundant traits to test for association with the mtDNA-CN GWAS SNPs (Supplemental Table 5, in PHEWAS = 1). We clustered SNPs using the trait effect size (beta) divided by the mtDNA-CN effect size estimate, so that all effects are standardized to the effect of the mtDNA-CN increasing allele for each locus. We identified 5 clusters of SNPs (Figure 4A), with clusters 1, 2, and 3 containing SNPs in which the mtDNA-CN increasing allele is associated with decreased platelet count (PLT) (Figure 4B), increased mean platelet volume (MPV) (Figure 4C), and platelet distribution width (PDW) (Figure 4D), consistent with a role in platelet activation[62]. Cluster 4 is most strongly enriched for SNPs in which the mtDNA-CN increasing allele is associated with increased PLT, plateletcrit (PLTCRIT, a measure of total platelet mass), serum calcium (Figure 4E), serum phosphate, as well as decreased mean corpuscular volume (MCV) and mean spherical cellular volume (Figure 4F) (Supplemental Table 12). The cluster 4 phenotypes, and supported by the genes identified for this cluster, implicate megakaryocyte proliferation and proplatelet formation (*MYB*, *AK3*, *JAK2*)[63], and apoptosis and autophagy (*BAK1*, *BCL2, TYMP*)[64]. Cluster 5 did not yield any specific trait enrichment (all significant results reflected the strong enrichment observed in clusters 1-4); however, gene set enrichment

493 for this cluster identified multiple mtDNA-related gene ontology terms, including mitochondrial
494 genome maintenance, regulation of phospholipid efflux, and amyloid-beta clearance
495 (Supplemental Table 13). Cluster 5 contains a highly diverse set of SNP-trait associations. A
496 secondary clustering only with SNPs mapped to this cluster identified 9 sub-clusters
497 (Supplemental Figure 6A), 3 having clear evidence for enrichment of specific phenotypes.
498 Cluster 6 is enriched for decreased MCV (Supplemental Figure 6B). Cluster 7 is enriched for
499 decreased MPV (Supplemental Figure 6C), PDW, serum phosphate, serum calcium, and
500 PLTCRT. Cluster 10 is enriched for decreased apoA, apoB, total cholesterol (Supplemental
501 Figure 6D), LDL, and HDL, and increased vitamin D, pulse rate, and direct bilirubin. SNPs that
502 were not assigned to a cluster were enriched for decreased fasting blood glucose and maximum
503 workload from fitness testing (Supplemental Table 14).
504
505 **Discussion**
506 We conducted a GWAS for mtDNA-CN using 465,809 individuals from the CHARGE consortium
507 and the UKB. In addition to replicating two previously reported loci, we discovered 94 novel loci
508 and report multiple independent hits for 26 loci. Examining our GWS SNPs in a Black
509 population, we observed a concordant signal, suggesting that the genetic etiology of mtDNA-CN
510 may be broadly similar across populations. Using several functional follow-up methods, genes
511 were assigned for each identified independent hit and significant enrichment was observed for
512 genes involved in mitochondrial DNA metabolism, homeostasis, cell activation, and amyloid-
513 beta clearance. In total, we assigned 124 unique genes to independent GWAS signals
514 associated with mtDNA-CN. We also identified 8 additional genes whose predicted gene
515 expression is associated with mtDNA-CN that could not be mapped back to GWS loci. Finally,
516 using a clustering approach based on SNP associations with various mtDNA-CN associated
517 phenotypes, we were able to functionally categorize SNPs, providing insight into biological
518 pathways that impact mtDNA-CN. We note that during the preparation of this manuscript, a

519 GWAS including 295,150 unrelated individuals from the UK Biobank was published, which
520 reported 50 genome-wide significant regions[33]. While many of our loci overlap with their
521 findings, our study reports twice as many loci largely due to the increased power of our study.
522 We were able to identify a substantial proportion of the genes involved in mtDNA
523 depletion syndromes (7/16, $p = 3.09 \times 10^{-15}$ for enrichment), including *TWNK*, *TFAM*, *DGUOK*,
524 *MGME1*, *RRM2B, TYMP,* and *POLG*. mtDNA depletion syndromes can be broken down into 5
525 subtypes based on their constellation of phenotypes[65] , and with the exception of
526 cardiomyopathic subtypes (associated with mutations in *AGK* and *SLC25A4*), we were able to
527 identify at least 1 gene from the other 4 subtypes, suggesting that our mtDNA-CN measurement
528 in blood-derived DNA can identify genes widely relevant to non-blood phenotypes. This finding
529 is consistent with a large body of work showing that mtDNA-CN measured in blood is
530 associated with numerous aging-related phenotypes for which the primary tissue of interest is
531 not blood (e.g. chronic kidney disease[13], heart failure[11], and diabetes[66]). Also consistent with this
532 finding is recent work demonstrating that mtDNA-CN measured in blood is associated with
533 mtRNA expression across numerous non-blood tissues, suggesting a link between
534 mitochondrial function measured in blood and other tissues[67].
535 In addition to identifying the mtDNA depletion syndrome genes directly linked to
536 mitochondrial DNA metabolic processes, DNA replication, and genome maintenance, we also
537 identify genes which play a role in mitochondrial function. The top GWAS hit is a missense
538 mutation in *LONP1*, which encodes a mitochondrial protease that has been shown to cause
539 mitochondrial cytopathy and reduced respiratory chain activity[68,69]. Interestingly, this missense
540 mutation was recently found to be associated with mitochondrial tRNA methylation levels[70].
541 Additional genes known to impact mitochondrial function include *MFN1*, which encodes a
542 mediator of mitochondrial fusion[71,72], *STMP1*, which plays a role in mitochondrial respiration[73],
543 and *MRPS35*, which encodes a ribosomal protein involved in protein synthesis in the
544 mitochondrion[74,75].

Using a combination of gene-based tests and gene prioritization using functional annotation, pathway analyses reveal enrichment for numerous mitochondrial related pathways, as well as those involved in regulation of cell activation ($p < 3.65 \times 10^{-5}$), homeostatic processes ($p < 1.82 \times 10^{-5}$), and regulation of immune system processes ($p < 2.75 \times 10^{-5}$) (Supplemental Table 11). These results provide additional evidence for the broad role played by mitochondria in numerous aspects of cellular function. Of particular interest, the GO term for amyloid beta is significantly enriched, reinforcing a link between mtDNA-CN and neurodegenerative disease[76–78]. Previous work from our lab using the UKB has shown that increased mtDNA-CN is associated with lower rates of prevalent neurodegenerative disease, as well as predictive for decreased risk of incident neurodegenerative disease[67]. mtDNA-CN is also known to be decreased in the frontal cortex of Alzheimer's disease (AD) patients[79]. Interestingly, the four GWAS-identified genes driving the enrichment for amyloid-beta clearance are all related to regulation of lipid levels, and lipid homeostasis within the brain is known to play an important role in Alzheimer's disease[80]. *APOE*, one of the most well-known risk genes for Alzheimer's disease, is a cholesterol carrier involved in lipid transport, and the ApoE-ε4 isoform involved in AD pathogenesis is associated with mitochondrial dysfunction and oxidative distress in the human brain[81]; *CD36* is a platelet glycoprotein which mediates the response to amyloid-beta accumulation[82]; *LDLR* is a low-density lipoprotein receptor associated with AD[83]; and *ABCA7* is a phospholipid transporter[84]. *ABCA7* loss of function variants are enriched in both AD and Parkinson's disease (PD) patients[85], suggesting a broad role across neurodegenerative diseases.

Given the integral role of mitochondria in cellular function, not just with ATP formation/energy production, but signaling through ROS, and its key role in apoptosis, there is a strong reason to *a priori* assume that genetic variants associated with mtDNA-CN are likely to be highly pleiotropic. Indeed, mtDNA-CN itself is associated with numerous phenotypes (Supplemental Table 5). Through our PHEWAS-based clustering approach using 42 mtDNA-CN

571 associated phenotypes, we uncovered phenotypic associations between five distinct clusters of
572 GWS mtDNA-CN associated SNPs. Cluster 1-3 were characterized by increased MPV, PDW,
573 and decreased PLT (note that measured MPV and PLT are generally inversely correlated to
574 maintain hemostasis), which are the hallmarks of platelet activation[62]. The link between platelets
575 and mtDNA-CN has typically revolved around platelet count, as platelets have functional
576 mitochondria, but do not have a nucleus. Given that the mtDNA-CN measurement is the ratio
577 between mtDNA and nuclear DNA, increased platelets, all else being equal, would directly
578 equate with increased mtDNA-CN. We note that the mtDNA-CN metric used in this GWAS was
579 adjusted for platelet count, likely increasing the ability to detect variants that impact mtDNA-CN
580 through increased platelet activation. Examining the genes within this cluster suggests role for
581 actin formation/regulation (*CARMIL1*, *TPM4*, *PACSIN2*)[86–88] and vesicular transport/endocytic
582 trafficking (*DNM3*, *EHD3*)[89,90] in platelet activation.
583 Cluster 4 is most strongly enriched for SNPs in which the mtDNA-CN increasing allele is
584 associated with increased PLT/PLTCRIT and serum calcium/phosphate. Examining the genes
585 assigned to the cluster, we implicate megakaryocyte proliferation and proplatelet formation
586 (*MYB*, *AK3*, *JAK2*)[63], and apoptosis and autophagy (*BAK1*, *BCL2, TYMP*)[64]. Megakaryocytes
587 are used to form proplatelets, and the process includes an important role for both intra- and
588 extracellular calcium levels[91]. A role for apoptosis, and specifically *BCL2*, in proplatelet
589 formation and platelet release has been suggested[92,93], however work in mice has suggested
590 that apoptosis does not play a direct role in these processes[94]. Nevertheless, apoptosis is
591 important for platelet lifespan[95].
592 Cluster 5 was particularly challenging to interpret, given that no particular phenotype was
593 enriched relative to the non-cluster 5 SNPs. We note that this cluster appeared to be enriched
594 for the mtDNA depletion syndrome genes, containing 6/7 genes identified in the GWAS, and
595 significantly enriched for GO Terms related mitochondrial DNA. Examining sub-clusters within
596 cluster 5 identified MCV as an important phenotype (while only sub-cluster 6 was formally

associated, looking at Figure 3B suggests more widespread clustering based on MCV). MCV is a measure of the average volume of a red blood corpuscle, and the link between red blood volume and mtDNA-CN is unlikely to be direct, given that red blood cells contain neither nuclei nor mitochondria. Sub-cluster 7 is surprisingly enriched for variants associated with decreased MPV/PDW and serum phosphate/calcium, inverse of what we observe for cluster 4. Finally, sub-cluster 10 is strongly associated with lipids, and includes a number of genes known to be associated with lipid levels (*LIPC*, *CETP*, *LDLR*, *APOE*). While lipids play a role in both energy metabolism (largely through fatty acids) and cellular membrane formation, a link to mtDNA-CN and/or mitochondrial function is not well-established. One potentially interesting result is provided by Olkowicz and colleagues, who demonstrated that $ApoE^{-/-}/LDLR^{-/-}$ mice had increased cardiac mitochondrial oxidative metabolism, with proteomic analysis suggesting increased mitochondrial abundance in mouse hearts[96]. However, we note that our results show an association between decreased lipids and increased mtDNA-CN, rather than the reverse, shown in the Olkowicz study.

Although the question of causality is of great interest, this study was unable to determine the directionality of effect between mtDNA-CN and phenotypes of interest, as we are underpowered for Mendelian randomization (MR), with less than 1% of the variance in mtDNA-CN explained by GWS loci when predicted into ARIC. As an additional limitation, we note that despite the large sample size and numerous loci identified, we are likely missing a great deal of the true signal, as previous studies have estimated mtDNA-CN heritability to be 65%[23]. Finally, while we have adjusted our mtDNA-CN metric for a variety of confounders, it is important to note that mtDNA-CN can be influenced by a variety of environmental factors including smoking[97] and drugs, which have not been adjusted for in these analyses.

In summary, we performed the largest-to-date GWAS for mtDNA-CN, including almost 500,000 individuals. We identified distinct groups of SNPs associated with mtDNA-CN that are related to platelet activation, megakaryocyte formation and apoptotic processes, and showed

clear enrichment for genes involved in mtDNA depletion and nucleotide regulation. Given the role of mtDNA-CN in aging-related disease, this work begins to unravel the many varied underlying mechanisms for genetic control of mtDNA-CN.

Supplemental Data

Supplemental Data include six supplemental figures, fourteen supplemental tables, and supplemental methods.

Declaration of Interests

Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. All other authors declare no competing interests.

Acknowledgements

This work was supported by National Heart, Lung and Blood Institute, National Institutes of Health (NIH) grants R01HL13573 and R01HL144569 (RJL, SYY, CAC, DEA), NIH grant P01-AG027734 (GA, YK, NB, AB) and the National Center for Advancing Translational Sciences, NIH, through Grant KL2TR002490 (LMR). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. LMR was also funded by T32 HL129982.

This research was conducted using data from the Genotype-Tissue Expression (GTEx) project (dbGaP accession: phs000424.**v8**.p2). The GTEx project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

This research was also conducted using the UK Biobank Resource under Application Number 17731.

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Funding support for “Building on GWAS for NHLBI diseases: the U.S. CHARGE consortium” was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). Sequencing was carried out at the Baylor College of Medicine Human Genome Sequencing Center and supported by the National Human Genome Research Institute grants U54 HG003273 and UM1 HG008898. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086, 75N92021D00006; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and U01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National

673 Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant
674 DK063491 to the Southern California Diabetes Endocrinology Research Center.

675 The FHS phenotype-genotype analyses were supported by the National Institute of Aging
676 (U34AG051418). This research was conducted in part using data and resources from the
677 Framingham Heart Study of the National Heart Lung and Blood Institute of the National
678 Institutes of Health and Boston University School of Medicine. This work was partially supported
679 by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-
680 HC-25195, HHSN268201500001) and its contract with Affymetrix, Inc for genotyping services
681 (Contract No. N02-HL-6-4278). Genotyping, quality control and calling of the Illumina
682 HumanExome BeadChip in the Framingham Heart Study was supported by funding from the
683 National Heart, Lung and Blood Institute Division of Intramural Research (Daniel Levy and
684 Christopher J. O'Donnell, Principle Investigators). The authors thank the participants for their
685 dedication to the study. The authors are pleased to acknowledge that the computational work
686 reported on in this paper was performed on the Shared Computing Cluster which is
687 administered by Boston University's Research Computing Services. URL:
688 www.bu.edu/tech/support/research/.
689

690 MESA and the MESA SHARe projects are conducted and supported by the National Heart,
691 Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA
692 is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159,
693 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003,
694 N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164,
695 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-
696 95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Funding for SHARe
697 genotyping was provided by NHLBI Contract N02-HL-64278. Genotyping was performed at

698 Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston,
699 Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. The provision
700 of genotyping data was supported in part by the National Center for Advancing Translational
701 Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and
702 Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California
703 Diabetes Endocrinology Research Center.
704
705 ROS/MAP is supported by the Translational Genomics Research Institute and National Institute
706 on Aging (NIA) through grants U01AG46152, U01AG61256, P30AG10161, R01AG17917,
707 RF1AG15819, R01AG30146.
708
709 SHIP and SHIP-TREND are part of the Community Medicine Research net of the University of
710 Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants
711 no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social
712 Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald
713 Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education
714 and Research (grant 03IS2061A).

715 This research was funded in whole, or in part, by the Wellcome Trust [Grant ref: 217065/Z/19/Z].
716 For the purpose of Open Access, the author has applied a CC BY public copyright license to any
717 Author Accepted Manuscript version arising from this submission.

718 **Web Resources**

719 REVIGO was accessed at http://revigo.irb.hr/.

720

721 **Data and Code Availability**

722 All data used in this manuscript is available through either the UKBiobank and CHARGE

723 consortiums. Code and scripts are available in a zipped file at

724 https://www.arkinglab.org/resources/.

725

726 **References**

727 1. Wallace, D.C. (1992). Diseases of the Mitochondrial Dna. Annual Review of Biochemistry *61*,
728 1175–1212.

729 2. Vakifahmetoglu-Norberg, H., Ouchida, A.T., and Norberg, E. (2017). The role of mitochondria
730 in metabolism and cell death. Biochem. Biophys. Res. Commun. *482*, 426–431.

731 3. Herst, P.M., Rowe, M.R., Carson, G.M., and Berridge, M.V. (2017). Functional Mitochondria
732 in Health and Disease. Front Endocrinol (Lausanne) *8*,.

733 4. Dai, D.-F., Rabinovitch, P.S., and Ungvari, Z. (2012). Mitochondria and cardiovascular aging.
734 Circ. Res. *110*, 1109–1124.

735 5. Cui, H., Kong, Y., and Zhang, H. (2012). Oxidative stress, mitochondrial dysfunction, and
736 aging. J Signal Transduct *2012*, 646354.

737 6. Liu, C.-S., Tsai, C.-S., Kuo, C.-L., Chen, H.-W., Lii, C.-K., Ma, Y.-S., and Wei, Y.-H. (2003).
738 Oxidative stress-related alteration of the copy number of mitochondrial DNA in human
739 leukocytes. Free Radic Res *37*, 1307–1317.

740 7. Guha, M., and Avadhani, N.G. (2013). Mitochondrial retrograde signaling at the crossroads of
741 tumor bioenergetics, genetics and epigenetics. Mitochondrion *13*, 577–591.

742 8. Malik, A.N., and Czajka, A. (2013). Is mitochondrial DNA content a potential biomarker of
743 mitochondrial dysfunction? Mitochondrion *13*, 481–492.

744 9. Ashar, F.N., Moes, A., Moore, A.Z., Grove, M.L., Chaves, P.H.M., Coresh, J., Newman, A.B.,
745 Matteini, A.M., Bandeen-Roche, K., Boerwinkle, E., et al. (2015). Association of mitochondrial
746 DNA levels with frailty and all-cause mortality. J. Mol. Med. *93*, 177–186.

747 10. Ashar, F.N., Zhang, Y., Longchamps, R.J., Lane, J., Moes, A., Grove, M.L., Mychaleckyj,
748 J.C., Taylor, K.D., Coresh, J., Rotter, J.I., et al. (2017). Association of Mitochondrial DNA Copy
749 Number With Cardiovascular Disease. JAMA Cardiol *2*, 1247–1255.

750 11. Hong, Y.S., Longchamps, R.J., Zhao, D., Castellani, C.A., Loehr, L.R., Chang, P.P.,
751 Matsushita, K., Grove, M.L., Boerwinkle, E., Arking, D.E., et al. (2020). Mitochondrial DNA Copy
752 Number and Incident Heart Failure: The Atherosclerosis Risk in Communities (ARIC) Study.
753 Circulation *141*, 1823–1825.

754 12. Zhao, D., Bartz, T.M., Sotoodehnia, N., Post, W.S., Heckbert, S.R., Alonso, A.,
755 Longchamps, R.J., Castellani, C.A., Hong, Y.S., Rotter, J.I., et al. (2020). Mitochondrial DNA
756 copy number and incident atrial fibrillation. BMC Medicine *18*, 246.

13. 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., et al. (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.

14. Pyle, A., Anugrha, H., Kurzawa-Akanbi, M., Yarnall, A., Burn, D., and Hudson, G. (2016). Reduced mitochondrial DNA copy number is a biomarker of Parkinson's disease. Neurobiol. Aging *38*, 216.e7-216.e10.

15. Wei, W., Keogh, M.J., Wilson, I., Coxhead, J., Ryan, S., Rollinson, S., Griffin, H., Kurzawa-Akanbi, M., Santibanez-Koref, M., Talbot, K., et al. (2017). Mitochondrial DNA point mutations and relative copy number in 1363 disease and control human brains. Acta Neuropathologica Communications *5*, 13.

16. Reznik, E., Miller, M.L., Şenbabaoğlu, Y., Riaz, N., Sarungbam, J., Tickoo, S.K., Al-Ahmadie, H.A., Lee, W., Seshan, V.E., Hakimi, A.A., et al. (2016). Mitochondrial DNA copy number variation across human cancers. Elife *5*,.

17. Hurtado-Roca, Y., Ledesma, M., Gonzalez-Lazaro, M., Moreno-Loshuertos, R., Fernandez-Silva, P., Enriquez, J.A., and Laclaustra, M. (2016). Adjusting MtDNA Quantification in Whole Blood for Peripheral Blood Platelet and Leukocyte Counts. PLOS ONE *11*, e0163770.

18. Knez, J., Winckelmans, E., Plusquin, M., Thijs, L., Cauwenberghs, N., Gu, Y., Staessen, J.A., Nawrot, T.S., and Kuznetsova, T. (2016). Correlates of Peripheral Blood Mitochondrial DNA Content in a General Population. Am J Epidemiol *183*, 138–146.

19. Kumar, P., Efstathopoulos, P., Millischer, V., Olsson, E., Wei, Y.B., Brüstle, O., Schalling, M., Villaescusa, J.C., Ösby, U., and Lavebratt, C. (2018). Mitochondrial DNA copy number is associated with psychosis severity and anti-psychotic treatment. Sci Rep *8*, 12743.

20. Urata, M., Koga-Wada, Y., Kayamori, Y., and Kang, D. (2008). Platelet contamination causes large variation as well as overestimation of mitochondrial DNA content of peripheral blood mononuclear cells. Ann Clin Biochem *45*, 513–514.

21. Clay Montier, L.L., Deng, J.J., and Bai, Y. (2009). Number matters: control of mammalian mitochondrial DNA copy number. J Genet Genomics *36*, 125–131.

22. Tang, Y., Schon, E.A., Wilichowski, E., Vazquez-Memije, M.E., Davidson, E., and King, M.P. (2000). Rearrangements of Human Mitochondrial DNA (mtDNA): New Insights into the Regulation of mtDNA Copy Number and Gene Expression. Mol Biol Cell *11*, 1471–1485.

23. Xing, J., Chen, M., Wood, C.G., Lin, J., Spitz, M.R., Ma, J., Amos, C.I., Shields, P.G., Benowitz, N.L., Gu, J., et al. (2008). Mitochondrial DNA content: its genetic heritability and association with renal cell carcinoma. J. Natl. Cancer Inst. *100*, 1104–1112.

24. Carling, P.J., Cree, L.M., and Chinnery, P.F. (2011). The implications of mitochondrial DNA copy number regulation during embryogenesis. Mitochondrion *11*, 686–692.

25. Harvey, A., Gibson, T., Lonergan, T., and Brenner, C. (2011). Dynamic regulation of mitochondrial function in preimplantation embryos and embryonic stem cells. Mitochondrion *11*, 829–838.

796 26. Copeland, W.C. (2014). Defects of Mitochondrial DNA Replication. J Child Neurol *29*, 1216–
797 1224.

798 27. Mandel, H., Szargel, R., Labay, V., Elpeleg, O., Saada, A., Shalata, A., Anbinder, Y.,
799 Berkowitz, D., Hartman, C., Barak, M., et al. (2001). The deoxyguanosine kinase gene is
800 mutated in individuals with depleted hepatocerebral mitochondrial DNA. Nat. Genet. *29*, 337–
801 341.

802 28. Wang, L., Limongelli, A., Vila, M.R., Carrara, F., Zeviani, M., and Eriksson, S. (2005).
803 Molecular insight into mitochondrial DNA depletion syndrome in two patients with novel
804 mutations in the deoxyguanosine kinase and thymidine kinase 2 genes. Mol. Genet. Metab. *84*,
805 75–82.

806 29. Rusecka, J., Kaliszewska, M., Bartnik, E., and Tońska, K. (2018). Nuclear genes involved in
807 mitochondrial diseases caused by instability of mitochondrial DNA. J Appl Genet *59*, 43–57.

808 30. Cai, N., Li, Y., Chang, S., Liang, J., Lin, C., Zhang, X., Liang, L., Hu, J., Chan, W., Kendler,
809 K.S., et al. (2015). Genetic Control over mtDNA and Its Relationship to Major Depressive
810 Disorder. Curr Biol *25*, 3170–3177.

811 31. Workalemahu, T., Enquobahrie, D.A., Tadesse, M.G., Hevner, K., Gelaye, B., Sanchez, S.,
812 and Williams, M.A. (2017). Genetic Variations Related to Maternal Whole Blood Mitochondrial
813 DNA Copy Number: A Genome-Wide and Candidate Gene Study. J Matern Fetal Neonatal Med
814 *30*, 2433–2439.

815 32. Guyatt, A.L., Brennan, R.R., Burrows, K., Guthrie, P.A.I., Ascione, R., Ring, S.M., Gaunt,
816 T.R., Pyle, A., Cordell, H.J., Lawlor, D.A., et al. (2019). A genome-wide association study of
817 mitochondrial DNA copy number in two population-based cohorts. Human Genomics *13*, 6.

818 33. Hägg, S., Jylhävä, J., Wang, Y., Czene, K., and Grassmann, F. (2020). Deciphering the
819 genetic and epidemiological landscape of mitochondrial DNA abundance. Hum Genet.

820 34. Longchamps, R.J., Castellani, C.A., Yang, S.Y., Newcomb, C.E., Sumpter, J.A., Lane, J.,
821 Grove, M.L., Guallar, E., Pankratz, N., Taylor, K.D., et al. (2020). Evaluation of mitochondrial
822 DNA copy number estimation techniques. PLOS ONE *15*, e0228166.

823 35. MitoPipeline: Generating Mitochondrial copy number estimates from SNP array data in
824 Genvisis.

825 36. Han, B., and Eskin, E. (2011). Random-Effects Model Aimed at Discovering Associations in
826 Meta-Analysis of Genome-wide Association Studies. The American Journal of Human Genetics
827 *88*, 586–598.

828 37. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to
829 variable selection in regression, with application to genetic fine mapping. Journal of the Royal
830 Statistical Society: Series B (Statistical Methodology) *82*, 1273–1300.

831 38. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic
832 variants from high-throughput sequencing data. Nucleic Acids Research *38*, e164–e164.

833 39. Sherry, S.T., Ward, M., and Sirotkin, K. (1999). dbSNP-database for single nucleotide
834 polymorphisms and other classes of minor genetic variation. Genome Res *9*, 677–679.

835 40. O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B.,
836 Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq)
837 database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic
838 Acids Res *44*, D733-745.

839 41. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C.,
840 and Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association
841 Studies Using Summary Statistics. PLOS Genetics *10*, e1004383.

842 42. Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H.,
843 Saha, A., Kreuzhuber, R., Kasela, S., et al. (2018). Unraveling the polygenic architecture of
844 complex traits using blood eQTL metaanalysis. BioRxiv 447367.

845 43. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.-J., Wood, A.R., Yang, J., Lui, J.C.,
846 Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide
847 association studies using predicted gene functions. Nature Communications *6*, 5890.

848 44. Smyth, G., Hu, Y., Ritchie, M., Silver, J., Wettenhall, J., McCarthy, D., Wu, D., Shi, W.,
849 Phipson, B., Lun, A., et al. (2021). limma: Linear Models for Microarray Data (Bioconductor
850 version: Release (3.12)).

851 45. Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis
852 for RNA-seq: accounting for selection bias. Genome Biology *11*, R14.

853 46. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M.,
854 Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L., et al. (2018). Exploring the phenotypic
855 consequences of tissue specific gene expression variation inferred from GWAS summary
856 statistics. Nat Commun *9*, 1825.

857 47. Barbeira, A.N., Pividori, M., Zheng, J., Wheeler, H.E., Nicolae, D.L., and Im, H.K. (2019).
858 Integrating predicted transcriptome from multiple tissues improves association detection. PLOS
859 Genetics *15*, e1007889.

860 48. Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and
861 visualizes long lists of gene ontology terms. PLoS One *6*, e21800.

862 49. Stiles, A.R., Simon, M.T., Stover, A., Eftekharian, S., Khanlou, N., Wang, H.L., Magaki, S.,
863 Lee, H., Partynski, K., Dorrani, N., et al. (2016). Mutations in TFAM, encoding mitochondrial
864 transcription factor A, cause neonatal liver failure associated with mtDNA depletion. Mol Genet
865 Metab *119*, 91–99.

866 50. Kornblum, C., Nicholls, T.J., Haack, T.B., Schöler, S., Peeva, V., Danhauser, K., Hallmann,
867 K., Zsurka, G., Rorbach, J., Iuso, A., et al. (2013). Loss-of-function mutations in MGME1 impair
868 mtDNA replication and cause multisystemic mitochondrial disease. Nat Genet *45*, 214–219.

869 51. El-Hattab, A.W., and Scaglia, F. (2013). Mitochondrial DNA depletion syndromes: review
870 and updates of genetic basis, manifestations, and therapeutic options. Neurotherapeutics *10*,
871 186–198.

872 52. Millard, L.A.C., Davies, N.M., Gaunt, T.R., Davey Smith, G., and Tilling, K. (2018). Software
873 Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank.
874 Int J Epidemiol *47*, 29–35.

875 53. Konopka, T. (2020). umap: Uniform Manifold Approximation and Projection.

876 54. Hahsler, M., Piekenbrock, M., Arya, S., and Mount, D. (2019). dbscan: Density Based
877 Clustering of Applications with Noise (DBSCAN) and Related Algorithms.

878 55. Gibellini, L., De Gaetano, A., Mandrioli, M., Van Tongeren, E., Bortolotti, C.A., Cossarizza,
879 A., and Pinti, M. (2020). The biology of Lonp1: More than a mitochondrial protease. Int Rev Cell
880 Mol Biol *354*, 1–61.

881 56. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet *45*,
882 580–585.

883 57. Liu, T., Lu, B., Lee, I., Ondrovicová, G., Kutejová, E., and Suzuki, C.K. (2004). DNA and
884 RNA binding by the mitochondrial lon protease is regulated by nucleotide and protein substrate.
885 J Biol Chem *279*, 13902–13910.

886 58. Sharma, M.R., Koc, E.C., Datta, P.P., Booth, T.M., Spremulli, L.L., and Agrawal, R.K.
887 (2003). Structure of the mammalian mitochondrial ribosome reveals an expanded functional role
888 for its component proteins. Cell *115*, 97–108.

889 59. Shimizu, S., Narita, M., and Tsujimoto, Y. (1999). Bcl-2 family proteins regulate the release
890 of apoptogenic cytochrome c by the mitochondrial channel VDAC. Nature *399*, 483–487.

891 60. Vagdatli, E., Gounari, E., Lazaridou, E., Katsibourlia, E., Tsikopoulou, F., and Labrianou, I.
892 (2010). Platelet distribution width: a simple, practical and specific marker of activation of
893 coagulation. Hippokratia *14*, 28–32.

894 61. PathCards :: Factors involved in megakaryocyte development and platelet production
895 Pathway and related pathways.

896 62. PathCards :: Apoptosis and Autophagy Pathway and related pathways.

897 63. Basel, D. (2020). Mitochondrial DNA Depletion Syndromes. Clin Perinatol *47*, 123–141.

898 64. DeBarmore, B., Longchamps, R.J., Zhang, Y., Kalyani, R.R., Guallar, E., Arking, D.E.,
899 Selvin, E., and Young, J.H. (2020). Mitochondrial DNA copy number and diabetes: the
900 Atherosclerosis Risk in Communities (ARIC) study. BMJ Open Diabetes Res Care *8*,.

901 65. Yang, S.Y., Castellani, C.A., Longchamps, R.J., Pillalamarri, V.K., O'Rourke, B., Guallar, E.,
902 and Arking, D.E. (2020). Blood-derived mitochondrial DNA copy number is associated with gene
903 expression across multiple tissues and is predictive for incident neurodegenerative disease.
904 BioRxiv 2020.07.17.209023.

905 66. Hannah-Shmouni, F., MacNeil, L., Brady, L., Nilsson, M.I., and Tarnopolsky, M. (2019).
906 Expanding the Clinical Spectrum of LONP1-Related Mitochondrial Cytopathy. Front Neurol *10*,
907 981.

908 67. Grainha, T.R.R., Jorge, P.A. da S., Pérez-Pérez, M., Pérez Rodríguez, G., Pereira,
909 M.O.B.O., and Lourenço, A.M.G. (2018). Exploring anti-quorum sensing and anti-virulence
910 based strategies to fight Candida albicans infections: an in silico approach. FEMS Yeast Res
911 *18*,.

912 68. Ali, A.T., Idaghdour, Y., and Hodgkinson, A. (2020). Analysis of mitochondrial m1A/G RNA
913 modification reveals links to nuclear genetic variants and associated disease processes.
914 Communications Biology *3*, 1–11.

915 69. Schrepfer, E., and Scorrano, L. (2016). Mitofusins, from Mitochondria to Metabolism. Mol
916 Cell *61*, 683–694.

917 70. Ishihara, N., Eura, Y., and Mihara, K. (2004). Mitofusin 1 and 2 play distinct roles in
918 mitochondrial fusion reactions via GTPase activity. J Cell Sci *117*, 6535–6546.

919 71. Zhang, D., Xi, Y., Coccimiglio, M.L., Mennigen, J.A., Jonz, M.G., Ekker, M., and Trudeau,
920 V.L. (2012). Functional prediction and physiological characterization of a novel short trans-
921 membrane protein 1 as a subunit of mitochondrial respiratory complexes. Physiol Genomics *44*,
922 1133–1140.

923 72. Cavdar Koc, E., Burkhart, W., Blackburn, K., Moseley, A., and Spremulli, L.L. (2001). The
924 small subunit of the mammalian mitochondrial ribosome. Identification of the full complement of
925 ribosomal proteins present. J Biol Chem *276*, 19363–19374.

926 73. Márquez-Jurado, S., Díaz-Colunga, J., das Neves, R.P., Martinez-Lorente, A., Almazán, F.,
927 Guantes, R., and Iborra, F.J. (2018). Mitochondrial levels determine variability in cell death by
928 modulating apoptotic gene expression. Nat Commun *9*, 389.

929 74. Dölle, C., Flønes, I., Nido, G.S., Miletic, H., Osuagwu, N., Kristoffersen, S., Lilleng, P.K.,
930 Larsen, J.P., Tysnes, O.-B., Haugarvoll, K., et al. (2016). Defective mitochondrial DNA
931 homeostasis in the substantia nigra in Parkinson disease. Nat Commun *7*, 13548.

932 75. Chen, C., Turnbull, D.M., and Reeve, A.K. (2019). Mitochondrial Dysfunction in Parkinson's
933 Disease-Cause or Consequence? Biology (Basel) *8*,.

934 76. Pinto, M., and Moraes, C.T. (2014). Mitochondrial genome changes and neurodegenerative
935 diseases. Biochim Biophys Acta *1842*, 1198–1207.

936 77. Rodríguez-Santiago, B., Casademont, J., and Nunes, V. (2001). Is mitochondrial DNA
937 depletion involved in Alzheimer's disease? Eur J Hum Genet *9*, 279–285.

938 78. Chew, H., Solomon, V.A., and Fonteh, A.N. (2020). Involvement of Lipids in Alzheimer's
939 Disease Pathology and Potential Therapies. Front Physiol *11*, 598.

940 79. Yin, J., Reiman, E.M., Beach, T.G., Serrano, G.E., Sabbagh, M.N., Nielsen, M., Caselli,
941 R.J., and Shi, J. (2020). Effect of ApoE isoforms on mitochondria in Alzheimer disease.
942 Neurology *94*, e2404–e2411.

943 80. El Khoury, J.B., Moore, K.J., Means, T.K., Leung, J., Terada, K., Toft, M., Freeman, M.W.,
944 and Luster, A.D. (2003). CD36 mediates the innate host response to beta-amyloid. J Exp Med
945 *197*, 1657–1666.

946 81. Lämsä, R., Helisalmi, S., Herukka, S.-K., Tapiola, T., Pirttilä, T., Vepsäläinen, S., Hiltunen,
947 M., and Soininen, H. (2008). Genetic study evaluating LDLR polymorphisms and Alzheimer's
948 disease. Neurobiol Aging *29*, 848–855.

949 82. Tomioka, M., Toda, Y., Mañucat, N.B., Akatsu, H., Fukumoto, M., Kono, N., Arai, H., Kioka,
950 N., and Ueda, K. (2017). Lysophosphatidylcholine export by human ABCA7. Biochim Biophys
951 Acta Mol Cell Biol Lipids *1862*, 658–665.

952 83. Nuytemans, K., Maldonado, L., Ali, A., John-Williams, K., Beecham, G.W., Martin, E., Scott,
953 W.K., and Vance, J.M. (2016). Overlap between Parkinson disease and Alzheimer disease in
954 ABCA7 functional variants. Neurol Genet *2*, e44.

955 84. Yang, C., Pring, M., Wear, M.A., Huang, M., Cooper, J.A., Svitkina, T.M., and Zigmond, S.H.
956 (2005). Mammalian CARMIL inhibits actin filament capping by capping protein. Dev Cell *9*, 209–
957 221.

958 85. Crabos, M., Yamakado, T., Heizmann, C.W., Cerletti, N., Bühler, F.R., and Erne, P. (1991).
959 The calcium binding protein tropomyosin in human platelets and cardiac tissue: elevation in
960 hypertensive cardiac hypertrophy. Eur J Clin Invest *21*, 472–478.

961 86. Kostan, J., Salzer, U., Orlova, A., Törö, I., Hodnik, V., Senju, Y., Zou, J., Schreiner, C.,
962 Steiner, J., Meriläinen, J., et al. (2014). Direct interaction of actin filaments with F-BAR protein
963 pacsin2. EMBO Rep *15*, 1154–1162.

964 87. Sever, S. (2002). Dynamin and endocytosis. Curr Opin Cell Biol *14*, 463–467.

965 88. Cai, B., Giridharan, S.S.P., Zhang, J., Saxena, S., Bahl, K., Schmidt, J.A., Sorgen, P.L.,
966 Guo, W., Naslavsky, N., and Caplan, S. (2013). Differential roles of C-terminal Eps15 homology
967 domain proteins as vesiculators and tubulators of recycling endosomes. J Biol Chem *288*,
968 30172–30180.

969 89. Di Buduo, C.A., Moccia, F., Battiston, M., De Marco, L., Mazzucato, M., Moratti, R., Tanzi,
970 F., and Balduini, A. (2014). The importance of calcium in the regulation of megakaryocyte
971 function. Haematologica *99*, 769–778.

972 90. De Botton, S., Sabri, S., Daugas, E., Zermati, Y., Guidotti, J.E., Hermine, O., Kroemer, G.,
973 Vainchenker, W., and Debili, N. (2002). Platelet formation is the consequence of caspase
974 activation within megakaryocytes. Blood *100*, 1310–1317.

975 91. Josefsson, E.C., James, C., Henley, K.J., Debrincat, M.A., Rogers, K.L., Dowling, M.R.,
976 White, M.J., Kruse, E.A., Lane, R.M., Ellis, S., et al. (2011). Megakaryocytes possess a
977 functional intrinsic apoptosis pathway that must be restrained to survive and produce platelets. J
978 Exp Med *208*, 2017–2031.

979 92. Josefsson, E.C., Burnett, D.L., Lebois, M., Debrincat, M.A., White, M.J., Henley, K.J., Lane,
980 R.M., Moujalled, D., Preston, S.P., O'Reilly, L.A., et al. (2014). Platelet production proceeds
981 independently of the intrinsic and extrinsic apoptosis pathways. Nat Commun *5*, 3455.

982 93. McArthur, K., Chappaz, S., and Kile, B.T. (2018). Apoptosis in megakaryocytes and
983 platelets: the life and death of a lineage. Blood *131*, 605–610.

984 94. Olkowicz, M., Tomczyk, M., Debski, J., Tyrankiewicz, U., Przyborowski, K., Borkowski, T.,
985 Zabielska-Kaczorowska, M., Szupryczynska, N., Kochan, Z., Smeda, M., et al. (2021).
986 Enhanced cardiac hypoxic injury in atherogenic dyslipidaemia results from alterations in the
987 energy metabolism pattern. Metabolism *114*, 154400.

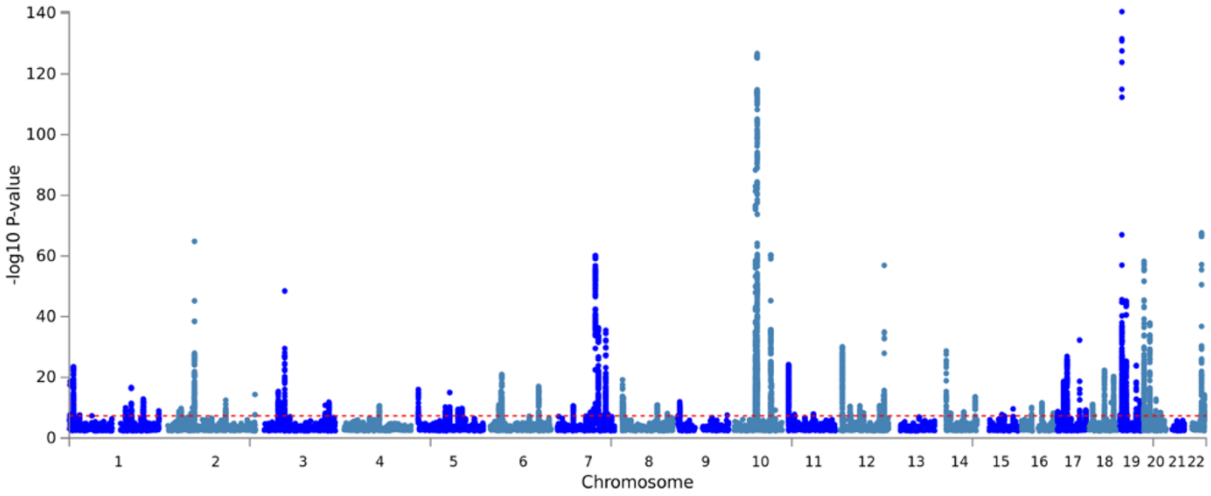
988 95. Vyas, C.M., Ogata, S., Reynolds, C.F., Mischoulon, D., Chang, G., Cook, N.R., Manson,
989 J.E., Crous-Bou, M., De Vivo, I., and Okereke, O.I. (2020). Lifestyle and behavioral factors and
990 mitochondrial DNA copy number in a diverse cohort of mid-life and older adults. PLoS One *15*,
991 e0237235.

992

993 **Figures**

994 **Figure 1. Manhattan plot of GWS loci from UKB-only analyses.**

995

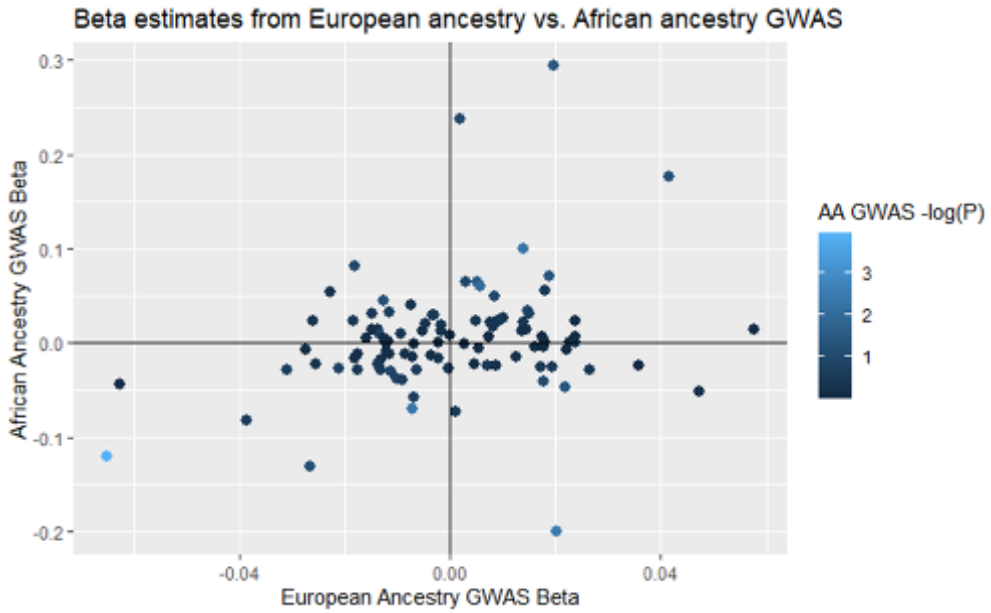


996

997 Manhattan plot showing genome-wide significant loci for the UK Biobank-only analyses.

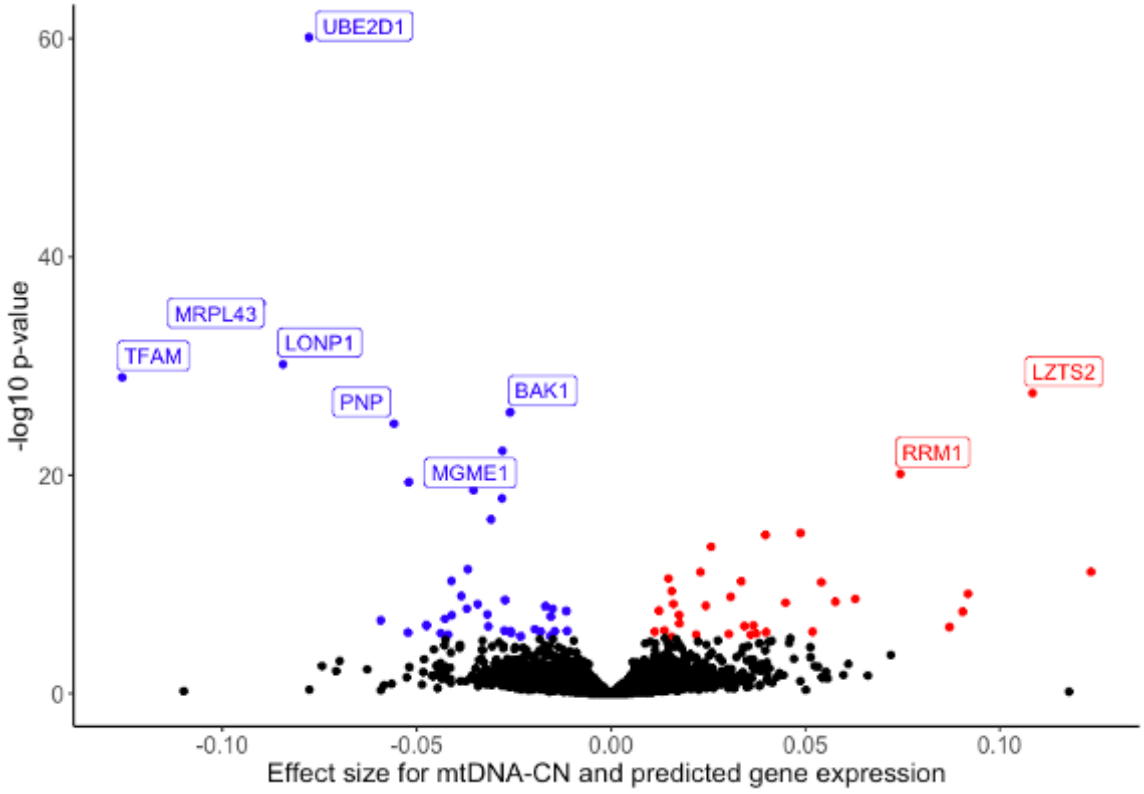
998

Figure 2. Scatterplot displaying effect size estimates between Whites and Blacks GWAS results for the 129 loci identified in the Whites analyses.



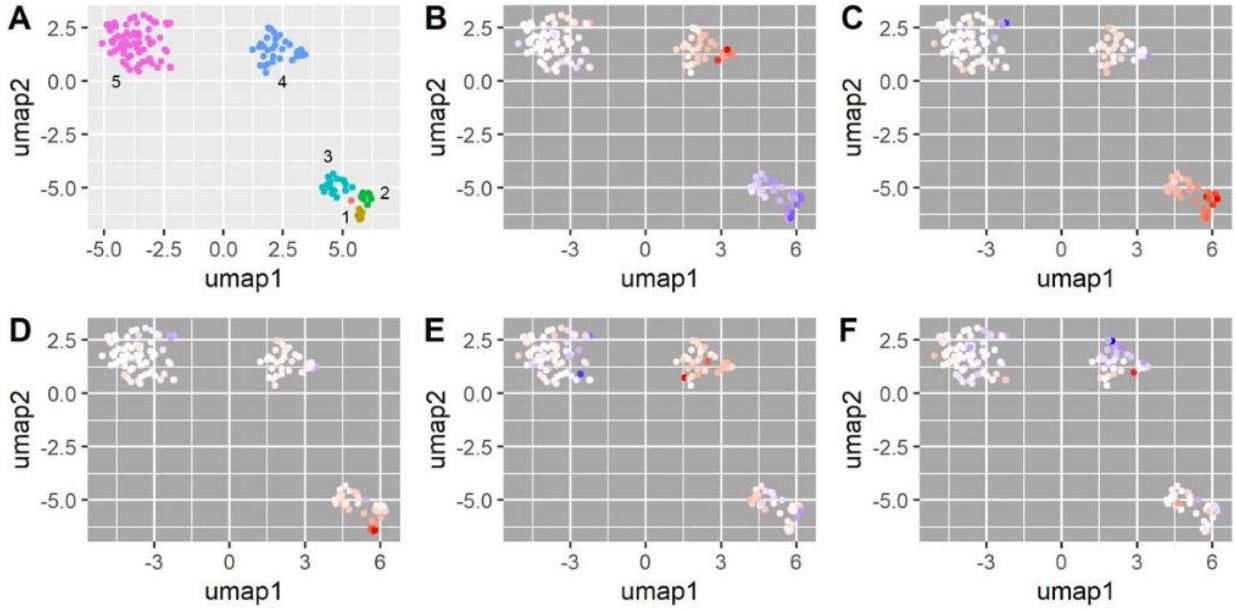
Scatterplot showing comparison between effect size estimates for White (European ancestry) and Black (African Ancestry) individuals. Color represents significance of effect for each locus in Blacks GWAS analyses.

Figure 3. Volcano plot of genes whose predicted gene expression is significantly associated with mtDNA-CN.



Volcano plot showing genes whose predicted gene expression is significantly associated with mtDNA-CN. Red indicates positive associations, blue indicates negative associations. Three genes (ARRDC1, EHMT1, PNPLA7) had extreme effect size estimates greater than 0.3 but were non-significant and removed from the plot for readability.

Figure 4. PHEWAS-based clustering of mtDNA-CN associated SNPs.



UMAP clusters created from PHEWAS associations for mtDNA-CN associated SNPs. (A) Five clusters were identified as labeled in the panel; orange indicates no cluster. (B-F) SNPs are colored based on their effect estimate size, standardized to the effect on mtDNA-CN (red = positive, blue = negative estimates), for (B) platelet count, (C) mean platelet volume, (D) platelet distribution width, (E) serum calcium levels, (F) mean spherical cellular volume.