



Deciphering the genetic and epidemiological landscape of mitochondrial DNA abundance

Sara Hägg¹ · Juulia Jylhävä¹ · Yunzhang Wang¹ · Kamila Czene¹ · Felix Grassmann^{1,2} 

Received: 16 September 2020 / Accepted: 15 December 2020 / Published online: 31 December 2020
© The Author(s) 2020, corrected publication 2021

Abstract

Mitochondrial (MT) dysfunction is a hallmark of aging and has been associated with most aging-related diseases as well as immunological processes. However, little is known about aging, lifestyle and genetic factors influencing mitochondrial DNA (mtDNA) abundance. In this study, mtDNA abundance was estimated from the weighted intensities of probes mapping to the MT genome in 295,150 participants from the UK Biobank. We found that the abundance of mtDNA was significantly elevated in women compared to men, was negatively correlated with advanced age, higher smoking exposure, greater body-mass index, higher frailty index as well as elevated red and white blood cell count and lower mortality. In addition, several biochemistry markers in blood-related to cholesterol metabolism, ion homeostasis and kidney function were found to be significantly associated with mtDNA abundance. By performing a genome-wide association study, we identified 50 independent regions genome-wide significantly associated with mtDNA abundance which harbour multiple genes involved in the immune system, cancer as well as mitochondrial function. Using mixed effects models, we estimated the SNP-heritability of mtDNA abundance to be around 8%. To investigate the consequence of altered mtDNA abundance, we performed a phenome-wide association study and found that mtDNA abundance is involved in risk for leukaemia, hematologic diseases as well as hypertension. Thus, estimating mtDNA abundance from genotyping arrays has the potential to provide novel insights into age- and disease-relevant processes, particularly those related to immunity and established mitochondrial functions.

Introduction

Mitochondria (MT), in addition to producing energy through oxidative processes, are involved in heat production, ion storage, apoptosis, intra- and extra-cellular cell signalling, biosynthesis and degradation of important metabolites as well as the processing of therapeutic agents. Depending on the tissue, cells host a dynamic range of multiple mitochondria which in turn contain multiple copies of a tiny circular genome (16,569 base pairs). The human mitochondrial proteome consists of more than 1100 proteins, of which only 13 are encoded in mitochondrial genome (Calvo and Mootha 2010). In contemporary human populations, several haplogroups are defined by ancestral and stable mutations in the

MT genome. Those haplogroups are derived from adaptation to different geographic areas under distinctive selection pressure (Torroni et al. 1996) and may also influence the abundance of mitochondrial DNA (mtDNA) (Liou et al. 2010; Gómez-Durán et al. 2012), although there is conflicting evidence (Hulgan et al. 2019; Pyle et al. 2016). Notably, different haplogroups have been reported to be associated with diseases as well as longevity (Benedictis et al. 1999; Chinnery and Gomez-Duran 2018) and as such contribute to phenotypic variety within human populations. Mitochondria rely on self-replication and are thus prone to cellular stress in aging and diseases. Hence, MT dysfunction and a reduction in MT biogenesis are hallmarks of aging (López-Otín et al. (2013)), and have been associated with most aging-related diseases (Haas 2019) as well as immunological processes (McGuire 2019).

Indeed, prior reports indicate that the amount of mtDNA in the blood decreases with advanced age (McArdle et al. 2013; Moore et al. 2018) as well as with lifestyle factors such as increased smoking behaviour (Cardellach et al. 2003) and higher BMI (McArdle et al. 2013). In addition, males were reported to have a lower amount of mtDNA in their blood

✉ Felix Grassmann
felix.grassmann@ki.se

¹ Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels väg 12A, 171 65 Stockholm, Sweden

² Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK

compared to women (Knez et al. 2016), which may be due to differing levels of white blood cells (Shim et al. 2020) or oestrogen levels (Giordano et al. 2011) between the sexes. However, it remains unclear whether the changes in mtDNA abundance also related to a reduced number of mitochondria in the respective cell population or due to reduced genome copies within the mitochondria.

To our knowledge, there is presently no comprehensive evaluation of aging or other factors influencing mtDNA abundance in nucleated blood cells in a large collection of individuals. In addition, few studies have been conducted to elucidate the inherited genetic contribution to mtDNA abundance (Workalemahu et al. 2017; Guyatt et al. 2019; López et al. 2014; Cai et al. 2015; Curran et al. 2007), most of which include a low number of individuals. Therefore, we aimed to leverage data from the UK Biobank, a cohort of more than 500,000 participants aged 40 and above, to uncover non-modifiable as well as lifestyle factors associated with mtDNA abundance and to further establish a role of inherited genetics in shaping the abundance of MT.

Results

We estimated the amount/abundance of mitochondrial DNA in blood in 295,150 individuals passing quality control (Supplementary Fig. 1 and Table 1) from the weighted intensities of genotyping probes mapped to the MT genome (Supplementary Table S1). We noticed that the individual genotyping batches (each containing around 5000 individuals) were contributing to the variance of the estimated mtDNA abundance (Supplementary Fig. 1). In addition, we found a strong effect of genotyping missingness [slope per SD and 95% confidence intervals (CI) 0.011 (0.008; 0.015), $P < 10^{-10}$] on the abundance of mitochondrial DNA in UK Biobank participants. Hence, we adjusted our analyses by both the genotyping batch as well as genotyping missingness.

Association of mitochondrial DNA abundance with sex, aging and lifestyle factors

Previous studies implicated a strong role of sex and advanced age in mtDNA abundance. Indeed, we found that male sex [slope and 95% CI -0.05 (-0.057 ; -0.043); $P < 10^{-43}$] and advanced age (slope per SD and 95% CI -0.06 (-0.064 ; -0.056); $P < 10^{-238}$) was strongly associated with mtDNA abundance. Since mtDNA abundance may depend on oestrogen levels and thus may be influenced by menopausal status in women, we next investigated whether menopausal status influences mtDNA abundance changes with advanced age. To this end, we used a two-lines test to programmatically determine the optimal age cut-off, which found that the association of age with mtDNA abundance has differential effects in women before and after 55 years of age (Supplementary Fig. 2). We found that mtDNA abundance increases with age in pre-menopausal women and decreases in post-menopausal women. In contrast, mtDNA abundance decreases steadily in men with advanced age and no differences between age groups was observed.

Due to the strong effect of sex on mtDNA abundance as well as the differential effect of advanced age on mtDNA abundance in pre- and post-menopausal women, we present all subsequent results separately for males and females. In addition to sex and increasing age, we also found that more packyears and elevated BMI were negatively associated with MT DNA abundance in both men and women (Q value < 0.05 , Fig. 1, Supplementary Table S2). In addition to those results, we found that mtDNA abundance is negatively associated with the increased frailty as expressed by the frailty index, an accumulation deficit model summarizing self-reported symptoms and diagnoses (Q value < 0.05 , Fig. 1, Supplementary Table S2). In line with those results, individuals with higher mtDNA abundance had a statistically significant better/longer survival time after recruitment.

Table 1 Baseline characteristics of UK Biobank participants by sex

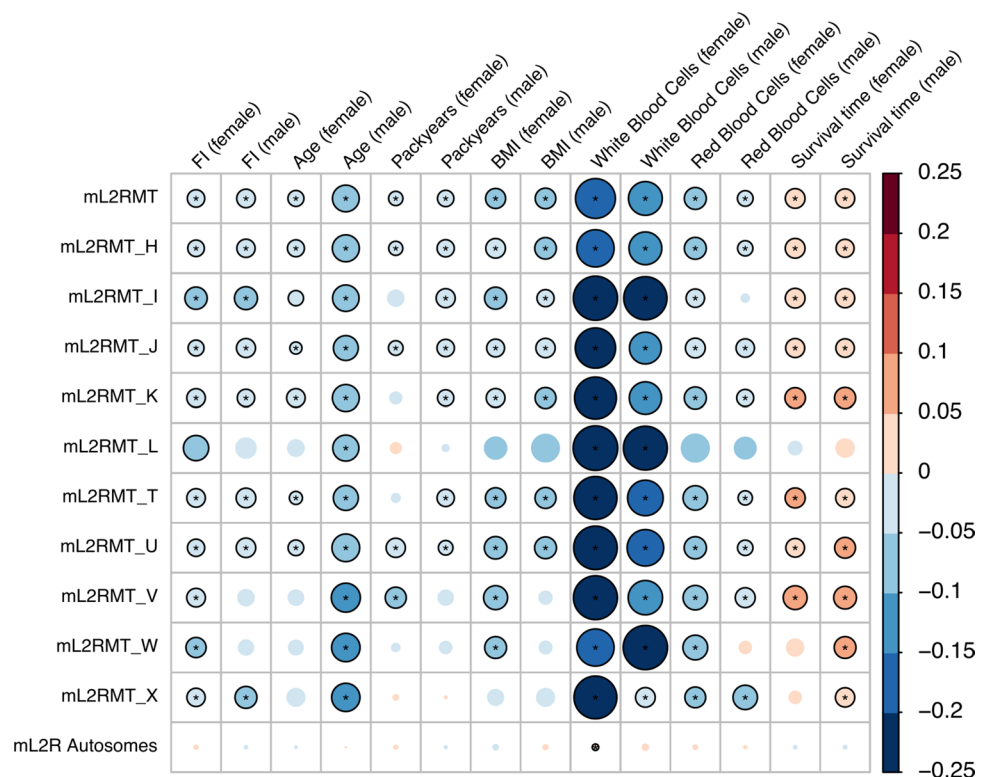
Variable	UK biobank participants		
	Females	Males	Both
Number of individuals	158,762	136,388	295,150
Average age (SD) [years]	56.67 (7.90)	57.14 (8.09)	56.89 (7.99)
Average body mass index (SD) [kg/m ²]	27.03 (5.12)	27.83 (4.21)	27.40 (4.74)
Average packyears smoked (SD)	6.40 (12.89)	11.05 (18.85)	8.52 (16.05)
Average mtDNA abundance (SD) ^a	0.02 (0.99)	-0.03 (0.99)	0.00 (0.99)
Median frailty index (IQR) ^b	0.11 (0.01–0.21)	0.11 (0.01–0.20)	0.11 (0.01–0.21)
Median call rate (range)	1.00 (0.99–1.00)	1.00 (0.99–1.00)	1.00 (0.99–1.00)

SD standard deviation, IQR interquartile range

^aAverage mtDNA abundance (mL2RMT) estimated from the weighted genotyping probe intensities (L2R), scaled to have a mean of zero and a SD of 1.00 in each genotyping plate

^bMedian fraction of reported health related deficits out of 49 investigated deficits

Fig. 1 Association between mitochondrial DNA abundance and different covariates. The size and colour (see colour bar) of the circle represent the slope of the association of seven features with mitochondrial (mL2RMT) as well as autosomal DNA abundance (mL2R Autosomes), stratified by sex (for the precise effect sizes see also Supplementary Table S2). The associations were also analysed separately by major MT haplogroup (labelled H–X). Correlations which were statistically significant (Q value < 0.05) are indicated with an asterisk. Nominally significant associations (P value < 0.05) are shown as circles with black borders. *FI* frailty index, *BMI* body mass index, *mL2RMT* median log 2 ratio of mitochondrial probes; survival time, time till death or end of follow-up



While we found that the abundance of mtDNA was related to the haplogroup of the individuals (Supplementary Figure S3), similar and consistent effect sizes were observed for the most common haplogroups present in the UKB, indicating that the genetic make-up of the MT itself is not a major factor in determining mtDNA abundance changes due to age or other factors (Fig. 1, Supplementary Table S2).

Mitochondrial abundance and blood markers

Next, we investigated whether the amount of DNA that is hybridized to the genotyping arrays may be directly related to the number of nucleated cells per volume of blood. Indeed, we found that the white blood cell count was negatively associated with mtDNA abundance (Fig. 1, Supplementary Table S2). However, we also found that increased red blood cell count was associated with reduced mtDNA abundance, indicating that not only the number of nucleated cells but also other mechanisms seem to play a role in governing mtDNA abundance. This conclusion is supported by the finding that we only observed a weak positive correlation of the aggregated autosomal DNA abundance with white blood cell count, which was estimated from all genotyping probes mapped to autosomes (Fig. 1). We also investigated individual blood markers and their association with mtDNA abundance and found that the mtDNA abundance was negatively linked to neutrophil, eosinophil, basophil and monocyte percentage and positively linked to

lymphocyte percentage and platelet count (Fig. 2a, Supplementary Table S3). In contrast, autosomal DNA abundance was only significantly positively associated with lymphocyte percentage in females.

Mitochondrial abundance and biochemistry markers

When we investigated the association of 29 biochemistry markers (from urine and blood, Fig. 2b, Supplementary Table S4) and mtDNA abundance, we noticed significant associations with markers related to inflammation (C-reactive protein), kidney function (albumin, cystatin c, urea and urate as well as sodium in urine), liver function (alkaline phosphatase, alanine aminotransferase), cholesterol metabolism (LDL cholesterol, total cholesterol, triglycerides, apolipoprotein A) as well as ion homeostasis (calcium and phosphate), vitamin D levels and glucose metabolisms (IGF, glucose, HbA1c). Since the immune system and thus the number of immune cells may well influence the observed associations, we also present the analyses adjusted for the neutrophil and lymphocyte percentage as well as white blood cell count measured in each participant (Fig. 2b, Supplementary Table S4). After adjustment, most of the associated markers remained statistically significantly associated with mtDNA abundance, albeit with lower effect sizes. Of note, markers related to cholesterol metabolisms such as LDL and HDL cholesterol, apolipoprotein A and B as well as total

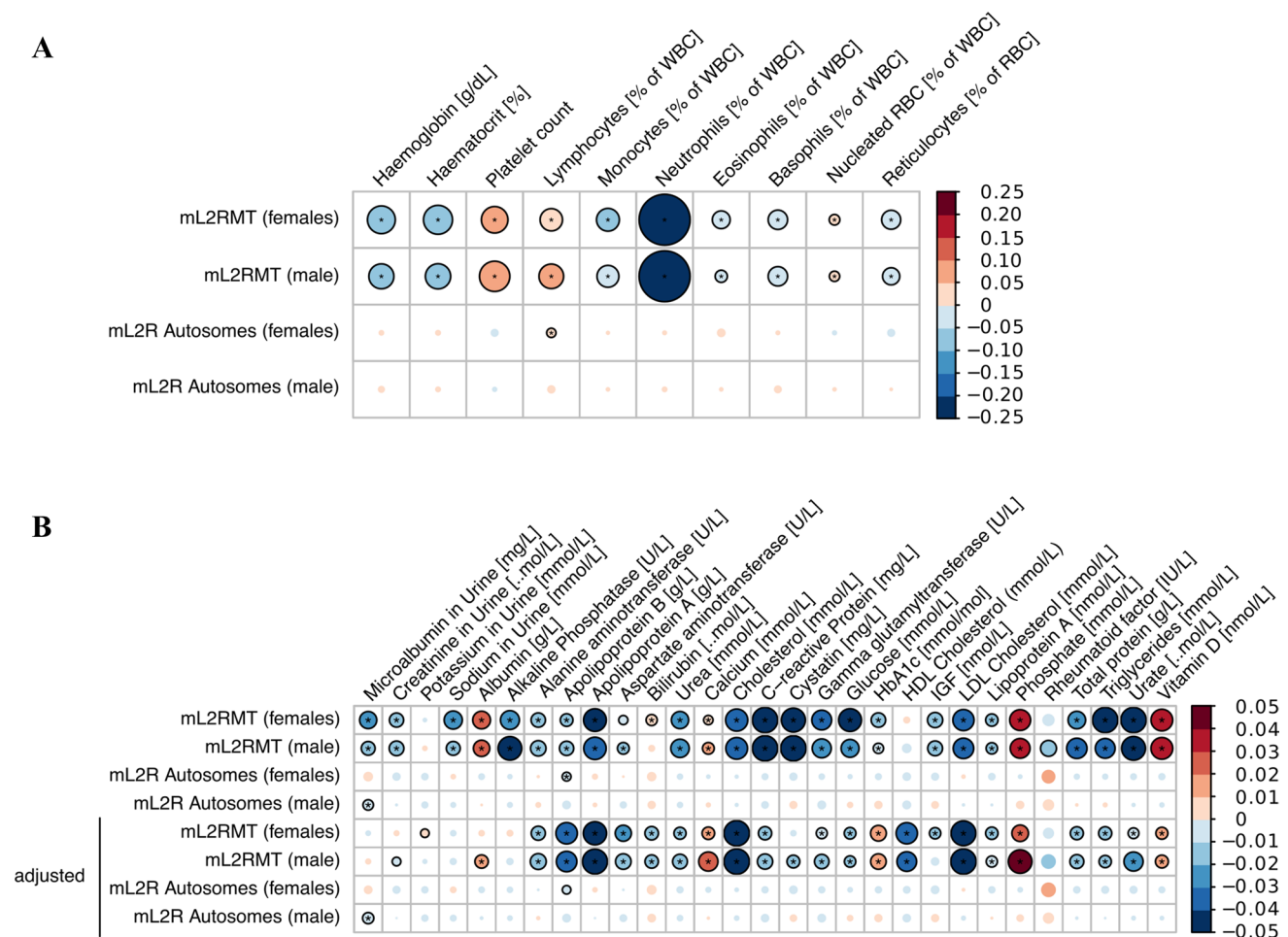


Fig. 2 Association between mitochondrial DNA abundance and individual blood markers. The size and colour of the circle represent the slope of the correlation of blood cell counts or other blood measurements (**a** see also Supplementary Table S3) and biochemistry features (**b** see also Supplementary Table S4) with mitochondrial (mL2RMT) as well as autosomal DNA abundance (mL2R Autosomes), stratified by sex. In **b** the association of biochemistry markers with mtDNA abundance were additionally adjusted for the neutrophil and lym-

phocyte percentage (% of white blood cells) as well as white blood cell count in the last four rows. Correlations which were statistically significant (Q value < 0.05) are indicated with an asterisk. Nominally significant associations (P value < 0.05) are shown as circles with black borders. *WBC* white blood cells, *RBC* red blood cells, *mL2RMT* median log 2 ratio of mitochondrial probes, *NRBC* nucleated red blood cells

cholesterol had even greater effect sizes after adjustment for immune cell count. In contrast, we found few associations of biochemistry markers with autosomal DNA abundance in our analyses. In particular, the only microalbumin in urine was found to be negatively associated to autosomal DNA abundance in men (Q value < 0.05).

Genetic dissection of mitochondrial abundance

To uncover genetic variants associated with mitochondrial abundance, we conducted a genome-wide association study (GWAS) in all individuals passing quality control. In total, we assessed the association of 3,505,788 variants with mtDNA abundance and found 50 regions in the genome with genome-wide significant variants (P value $< 5 \times 10^{-08}$,

Fig. 3). In those regions, a total of 66 independent signals were detected at genome-wide significance (Supplementary Table S5), implicating multiple genes to be involved in mtDNA abundance (Fig. 3). Importantly, by positional mapping of likely functional variants, we identified genes related to the immune system (*CXCL6*, *MEF2C*, *ITPR3*, *UBE2D1*, *C7orf73*, *STIM1*, *PNP*, *CRK* and *SIRPB1*), cancer and cell cycle (*TERT*, *BAK1*, *CDK6*, *CDK10*, *SUFU*, *FANCI*, *MDFIC*, *JMJD1C*, *USP7*, *BIK*) and mitochondrial function (*MFN2*, *TFAM*, *DGUOK*, *USP30*, *CREB5*, *POLG*) to be potentially involved in governing mtDNA abundance in blood (Fig. 3).

Next, we used fastBAT to investigate the burden of mtDNA abundance associated variants within the gene bodies of all protein-coding genes. In total, 227 genes

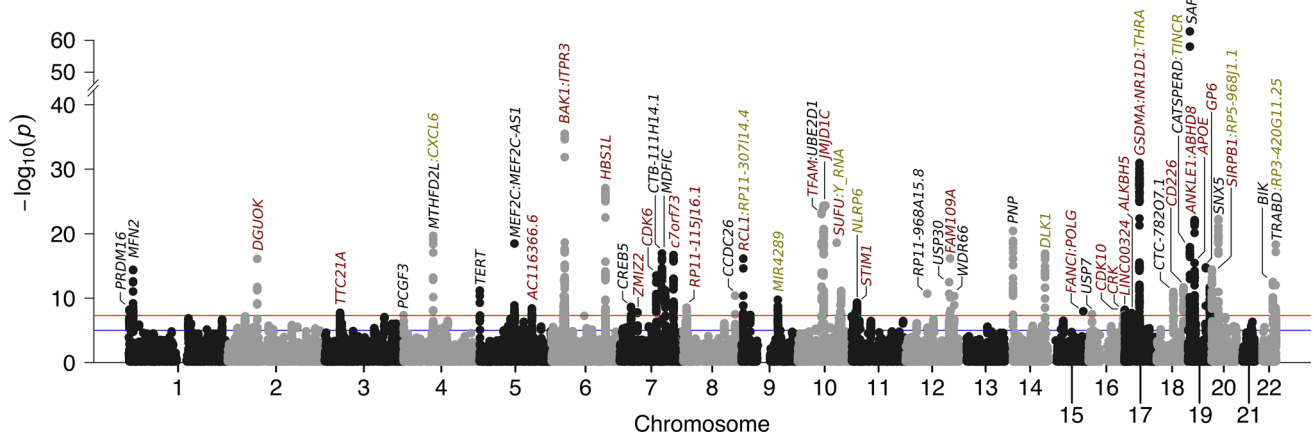


Fig. 3 Summary of the genome-wide association study of mtDNA abundance. Manhattan plot of 3,505,788 variants with minor allele frequency greater 5%. Genes influenced by associated variants are highlighted in different colours: genes in red harbour exonic variants, genes which have associated variants in the intronic region are shown

in black. In case no intronic or exonic variants are mapped to genes in a region, we highlighted the closest gene in green. The red line denotes genome-wide significance ($P < 5.00 \times 10^{-8}$) while the blue line denotes suggestive evidence for association ($P < 1.00 \times 10^{-5}$). The estimated genomic inflation factor λ was 1.147

carrying a statistically significant burden were identified (P value $< 2 \times 10^{-6}$, Supplementary Table S6). We then investigated a potential overrepresentation of pathways within those genes with WebGestaltR. With this approach, an overrepresentation of pathways related to immune activation, cell–cell adhesion, haematopoiesis, apoptosis, platelet production as well as mitochondrial biogenesis and plasma lipoprotein assembly was observed (Supplementary Figure S4 and Supplementary Table S7).

Finally, we used mixed-effects models to estimate the heritability of mtDNA abundance that can be explained by the genotyped variants. Accounting for age, sex, genotyping missingness, genotyping batch, the first ten genotyping PCs, neutrophil and lymphocyte percentage as well as white blood cell count, we estimated the SNP-heritability to be 8.3% (95% CI 7.6–9.0%), strongly implicating a role of inherited genetic variants on mtDNA abundance.

PheWAS of mitochondrial abundance

To estimate the consequence of altered mtDNA abundance on disease risk, we performed a Phenome-wide association study (PheWAS) on ICD10 codes derived from the hospital episode spell data mapped to a total of 1,150 phecodes. We found that higher levels of mtDNA abundance were statistically significantly (Q value < 0.05) associated with increased risk for (chronic lymphoid) leukaemia, myeloproliferative disease as well as diseases of the spleen (Fig. 4 and Supplementary Table S8). In contrast, higher mtDNA abundance resulted in reduced incidence of cardiomegaly, portal hypertension as well as oesophageal bleeding (Fig. 4 and Supplementary Table S8). Finally, prior reports indicated that

the risk increase due to differences in mtDNA abundance may follow a U-shaped distribution (i.e. first and third tertile compared to the second tertile increased the risk for a disease) (Workalemahu et al. 2016; Thyagarajan et al. 2012). To test this hypothesis, we have repeated the PheWAS to compare the lowest or the highest tertile of the mtDNA abundance distribution against the second (middle) tertile. However, we found that this approach yielded similar results as obtained from modelling mtDNA abundance across all tertiles. In addition, we observed no additional diseases statistically significantly associated with mtDNA abundance after correction for multiple testing.

Discussion

In this study, we have explored the epidemiological and genetic basis of mtDNA abundance in nucleated blood cells. We found that age as well as multiple lifestyle and health-related factors were associated with the amount of mtDNA in blood. In particular, mtDNA abundance seems to be mostly influenced by cells involved in the innate immune system as well as cholesterol metabolism and ion homeostasis. In addition, our genetic dissection of mtDNA abundance revealed 50 independent GWAS signals as well as estimated the heritability of mtDNA abundance to be around 8%. Finally, we showed that elevated mtDNA abundance is associated with risk for incident leukaemia, hypertensive disorders and oesophageal bleeding.

Due to the low number of variants and thus probes measuring the abundance of the MT genome, we expected that the mtDNA abundance estimates from chip arrays to be

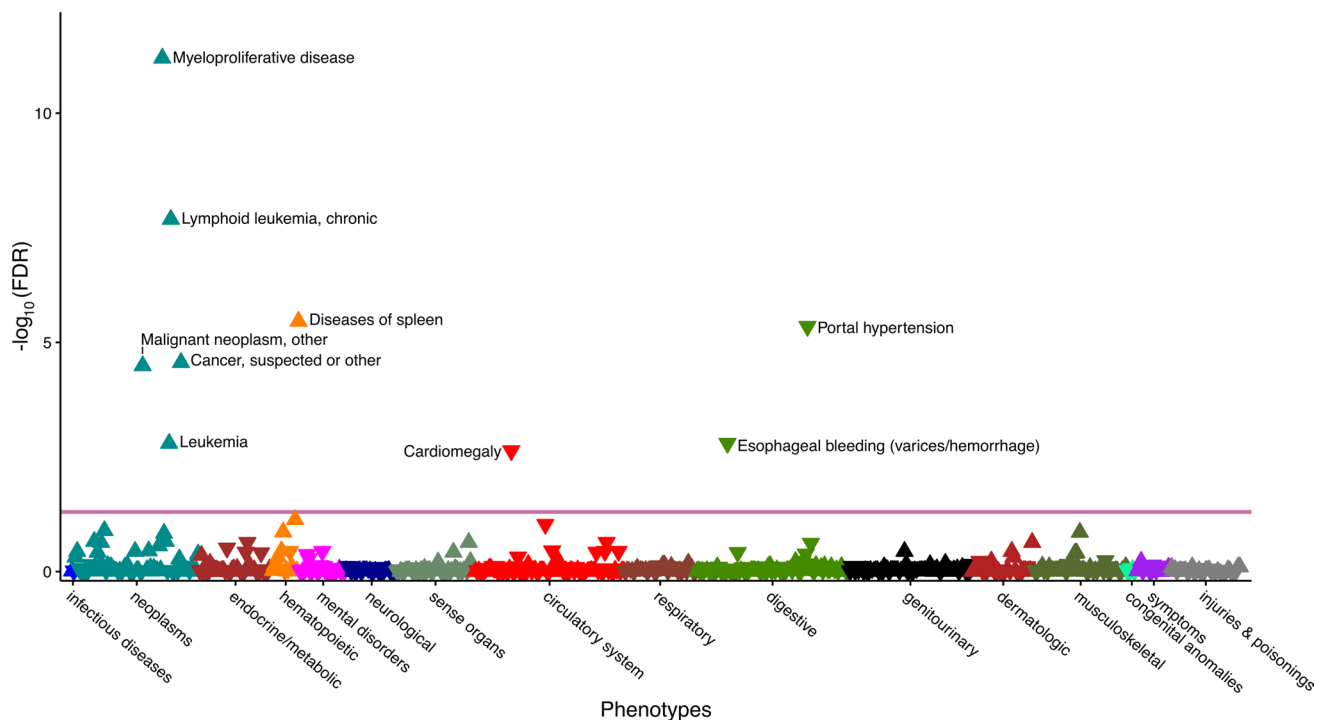


Fig. 4 Phenome-wide association study of mtDNA abundance. The consequence of altered mtDNA abundance on 1150 incident diseases or traits determined from the hospital episode spell data was evaluated in all UK Biobank participants (detailed association results are presented in Supplementary Table S8). ICD10 diagnoses were mapped to phecodes and the association between discretized mtDNA abundance (in tertiles) and the phecodes was estimated with logistic regression, adjusted for age at baseline, genotyping batch, total auto-

somal DNA abundance, fraction of neutrophils, fraction of lymphocytes, white blood cell count, sex, BMI and smoking at baseline as well as the first two principal components of ancestry. Only phecodes with more than 250 cases were considered. Statistically significant associations (Q value < 0.05) are shown above the red line. Triangles facing up indicate increased risk while triangle pointing down indicate reduced risk to due elevated mtDNA abundance

noisy. Indeed, we found that our initial (unweighted) mtDNA abundance values were not associated with known mtDNA correlates such as age, smoking, white blood cell count or BMI in a fashion that is consistent with the literature and was also negatively correlated to the coverage of the MT genome by exome sequencing. Since prior studies suggested that normalized exome coverage provides a robust estimate of mtDNA abundance and also has higher power to detect differences (Zhang et al. 2017; Picardi and Pesole 2012; D’Erchia et al. 2015; Longchamps et al. 2020), we, therefore, weighted the L2R values of individual array probes to more closely resemble the mtDNA abundance estimates from exome sequencing. With this approach, we found that both estimates are moderately positively correlated ($R=0.33$) and that the influence of genotype missingness on the weighted mtDNA abundance estimate was reduced more than tenfold [slope per SD and 95% CI of genotype missingness correlated to the mL2RMT estimate from MoChA: 0.146 (0.143; 0.150), $P < 10^{-250}$]. Since the intensity data does not contain any missing values (i.e. the signal intensity gets recorded for all probes), the decreased influence of genotyping missingness is unlikely to be attributed to the

selection of probes with more complete genotyping data. Furthermore, the resulting weighted mtDNA abundance estimates were associated with most covariates in a fashion that is in line with prior reports while the unweighted estimates were not (Supplementary Fig. 5). Thus, while the weighted mtDNA abundance estimates may not necessarily be as accurate as those determined from qPCR, we can estimate mtDNA abundance in the whole UKB cohort, drastically increasing our power to detect associations.

Although the effect of genotype missingness on MT abundance estimation is strongly reduced with our approach, it could still potentially confound our analyses. As such, we adjusted our analyses for the call rate to account for those differences. Therefore, it is unlikely that the genotyping missingness after adjustment would strongly confound most of our analyses, particularly since genotyping missingness itself was not associated with any of the investigated markers in this study (Q values > 0.05 , data not shown). To estimate the SNP-heritability of mtDNA abundance, the BOLT-LMM algorithm computes the relationship between individuals from all genotyped variants passing QC. Unless genotyping missingness is more strongly pronounced in more

related individuals, the estimation of relatedness should not be majorly confounded, particularly after adjusting for the actual genotyping missingness. Our results thus strongly implicate that the amount of mitochondrial DNA in nucleated blood cells is heritable. Further genetic studies accounting for missingness, DNA quality as well as immune cell count are warranted to uncover the precise genetic changes responsible for alternations in mtDNA abundance.

The disparities in mtDNA abundance by sex can be attributed to multiple factors which differ between the sexes. In this study, we were particularly interested in the effect of oestrogen exposure since it has been hypothesized to influence mtDNA abundance (Giordano et al. 2011). As such, we have investigated whether menopause status plays a role in determining mtDNA abundance. Indeed, the amount of mtDNA increases in pre-menopausal women with advanced age but decreases in post-menopausal women at the same rate as observed in males. This observation also explains the reduced effect sizes of age observed on the mtDNA abundance in women in the sex-stratified analyses. The programmatically determined age cut-off is in line with prior studies which showed that 95% of women are post-menopausal at age 55 (Nejat and Chervenak 2010) and using 55 years as the cut-off for menopause status should be sufficient in most research settings (Phipps et al. 2010). In addition, our results may also provide an explanation for the inconsistent results observed in the associations of mtDNA abundance and breast cancer (Lemnrau et al. 2015; Shen et al. 2015, 2010; Thyagarajan et al. 2013; Campa et al. 2018), which is a largely hormonally driven cancer with a strong influence of menopause status on the resulting tumour characteristics such as receptor status. Similarly, our results may also have implications for the investigation of other diseases which are influenced by changes in hormone levels and/or menopause status.

Importantly, our results indicate that mtDNA abundance is negatively associated with increased age, higher frailty, increased smoking behaviour as well as increased BMI, in agreement with prior publications. We hypothesized that MT copy number may be diminished due to lower rates of replication with increasing age (McArdle et al. 2013; Moore et al. 2018), due to oxidative stress induced by smoking (Cardellach et al. 2003), by low-grade inflammation and metabolic stress due to higher BMI (McArdle et al. 2013) or due to prior health-related conditions (as expressed by the frailty index) (Ashar et al. 2015; Bennett and Molofsky 2019). Since individuals with a healthier lifestyle tend to have higher mtDNA abundance, we also observed a reduced incidence of hypertensive disorders such as portal hypertension and cardiomegaly as well as lower mortality in those individuals. In our GWAS, genes related to blood pressure (*NLRP6*, *CREB5* and *APOE*) were found in regions associated with mtDNA abundance, providing a potential

explanation between increased mtDNA abundance and reduced risk for hypertensive diseases. Similarly, pathways and genes related to coagulation, platelet production, and count were implicated by our GWAS, thus pointing towards genes that are involved in both mtDNA abundance and risk for oesophageal bleeding. Interestingly, recent reports indicated that higher MT abundance is protective against oesophageal cancer (Xu et al. 2013), which may indicate that the oesophageal tissue is particularly vulnerable to changes in mitochondrial DNA abundance.

To further reinforce the notion that mtDNA abundance is heritable, we conducted a genome-wide association study on over 3 million common variants. We chose to focus on common variants since they are less likely to suffer from missing genotypes and the imputation of missing alleles is more accurate for those variants as well. Our GWAS implicated several genes to be involved in governing mtDNA abundance, particularly those related to immunity, cancer, cell cycle, and mitochondrial function. This is in line with reports that the loss of Y chromosome (mLOY) is related to similar pathways (Thompson et al. 2019), although none of the genes implicated in this study overlap with the genes associated with mLOY. Furthermore, we were able to identify variants significantly associated with mtDNA abundance within a region on chromosome 17 (association signal near *GSDMA*, i.e. genomic locus 40 in Supplementary Table S2), which was previously shown to be involved in mtDNA abundance [estimated from quantitative PCR (Guyatt et al. 2019)]. This region was also described to be involved in blood cell counts, particularly neutrophils (Astle et al. 2016; Okada et al. 2010; Lin et al. 2017). Therefore, our approach does seem to accurately capture mtDNA abundance and provides evidence for additional regions in the genome and thus genes involved in mtDNA abundance.

We observed a strong negative association between mtDNA abundance and the number of neutrophils and a positive correlation with lymphocytes percentage. Since neutrophils are the most abundant leukocyte and thus the most abundant cell carrying DNA, it may seem obvious that the amount of DNA hybridized to the array should be positively correlated. However, our results rather point to other conclusions. First, we did not observe any association of autosomal DNA abundance with the neutrophil count, indicating that the total amount of DNA that was hybridized does not play a major role, potentially due to careful harmonization of DNA concentration used in the genotyping process. Second, we observed differential correlation for mtDNA abundance with myeloblast-derived and lymphoid immune cells. Since both immune cell lineages carry mitochondria, they should similarly contribute to increased mtDNA abundance in case the measured mtDNA abundance is solely dependent on DNA content in blood. Therefore, we conclude that other processes must be responsible for the observed association

between increased mtDNA abundance and reduced (innate) immune cell count.

In our PheWAS analysis, we found increased mtDNA abundance to be associated with increased risk of leukaemia, which is in agreement with prior findings which found higher mtDNA abundance to cause increased leukaemia risk (Lan et al. 2008; Kim et al. 2015). Since an increased number of circulating lymphocytes is a hallmark of lymphoid leukaemia (Kipps et al. 2017) and we found a positive association of mtDNA abundance and lymphocyte count, this association would seem obvious. However, we adjusted our PheWAS by neutrophil and lymphocyte percentage as well as white blood cell count, thus the observed effect should be largely independent of immune cell abundance. Since platelets also contain mitochondria and their count can increase acutely through various stimuli (Rokkam and Kotagiri 2020), we additionally adjusted the PheWAS analysis by platelet count and found virtually the same results (results not shown). Those analyses implicate that other mitochondrial mechanisms independent of white blood cell count are responsible for the increased risk. Importantly, in our GWAS, which was also adjusted for white blood cell counts, we found association signals near genes involved in leukaemia and lymphoma development (*CRK*, *BAK1*, *CDK6*, *MDFIC*, *BIK*) as well as a significant enrichment of pathways related to immune cell proliferation. Thus, further functional studies are necessary to identify the underlying mechanism and responsible (immune) cell population that links the function of those genes to mtDNA abundance and leukaemia.

Of note, the mtDNA estimation was performed at recruitment of the patients and we excluded pre-existing conditions. This allowed us to assess the association of mtDNA abundance with diseases in a prospective manner and thus our results are unlikely influenced by prevalent diseases, which have been shown to alter mtDNA abundance (Ide et al. 2001). The clinical utility of assessing mtDNA estimation remains to be evaluated and would face challenges with our approach: Particularly, the observed effect sizes per tertile are in line with a moderate increase in risk largely independent of known factors influencing mtDNA abundance. Thus, the accuracy to predict the occurrence of those diseases would not be great and required additional risk factors in a comprehensive model. Therefore, further work is necessary to assess the clinical utility of mtDNA estimation, particularly while accounting for known risk factors for those particular diseases. Nevertheless, mtDNA will be a useful marker to add to the increasing number of biomarkers reflecting the human aging process, and as such will be interesting to study further (Jylhävä et al. 2017).

Multiple blood and urine bound biochemistry markers related to cholesterol and triglyceride metabolism as well as ion homeostasis and vitamin D levels were found to be

associated with mtDNA abundance independent of neutrophil and lymphocyte percentage or white blood cell count. This is in line with prior research (Liu et al. 2005) and the proposed function of mitochondria in cells since they are instrumental in the metabolism of cholesterol and other lipids as well as in ion storage, particularly phosphate and calcium.

A recent study presented evidence that cell-free and respiratory competent mitochondria are present in blood and may have a role in cell–cell communication (Al Amir Dache 2020). Although the majority of DNA used for the genotyping experiments should be derived from nucleated cells, our current approach can, however, not distinguish between cellular and cell-free mitochondria. It is therefore possible that some of the observed associations are additionally driven by extracellular mitochondria, which warrants further investigations.

Conclusion

We show that the amount of mtDNA in nucleated blood cells is heritable and associated with different modifiable and non-modifiable markers related to aging, health and lifestyle as well as immunity, cholesterol metabolism and ion homeostasis. Our findings suggest that estimating mtDNA abundance from weighted intensities on genotyping arrays is a promising tool to gain novel insights into disease-relevant mechanisms as well as to study aging.

Methods

UK Biobank study population and DNA extraction

UK Biobank (UKB) is a multi-centre cohort with 502,621 participants recruited in England, Scotland and Wales. Genotype data were available for 488,279 participants (Sudlow et al. 2015). The UK Biobank collected two 10 ml EDTA vacutainers per participant and extracted 850 µl of buffy coat, which was randomly distributed on extraction plates and stored at – 80 °C. The DNA for all participants was extracted from the buffy coat aliquot with the Promega Maxwell® 16 Blood DNA Purification Kit (AS1010). In this method, DNA is purified using magnetic particles which are used to capture, wash and elute DNA. Several quality control measures were evaluated after DNA purification and only plates that passed this step were used in genotyping (Welsh et al. 2017). In addition to the quality control steps by the UKB, we had to apply additional exclusion criteria to the UK Biobank participants, in line with standard practice in genome-wide association studies (Winkler et al. 2014) (Supplementary Figure S1). As such, we further excluded

individuals with documented and inferred sex chromosome abnormalities and those that had a low call rate (less than 99.0%) or excessive heterozygosity as determined by the UKB. Finally, we restricted the analyses to European ancestry (determined by the genotype principal components) as well as unrelated individuals, resulting in a total analytical dataset of 295,150 samples (Table 1). Further quality control measures based on the genetic dosage/intensity estimation are indicated below. The current study was conducted as part of the registered project 22,224. The UK Biobank obtained broad informed consent from its participants and has been approved by IRBs at participating sites. Access to the data implies that the proposed analyses are in agreement with the relevant IRB approvals.

Variables evaluated

The age and BMI of the individuals were recorded at the time of blood draw and smoking status were ascertained at baseline interview via questionnaire. The frailty index (FI) consisting of 49 health-related items, was computed as recently described (Williams et al. 2019) for all participants from questionnaire data. Briefly, the Rockwood accumulation deficit model was applied summarizing the number of self-reported diseases, symptoms and signs as well as overall health assessments in each individual. The FI ranges from zero to one and represents the fraction of the actual number of health deficits reported by an individual divided by the total number of potential health deficits considered (i.e. queried in the questionnaire). Increased frailty of an individual represents a major risk factor for future diseases as well as mortality. Recently, the UKB measured a panel of 29 biomarkers in urine and blood and the data were released to the public after extensive quality control. In addition, platelet, red and white blood cell counts were measured as an absolute number per unit volume and the abundance of individual blood cell classes such as lymphocytes, monocytes, neutrophils, eosinophils, basophils were recorded in percent (%) of white blood cells. Calibration and quality control were performed by the UK Biobank investigators. All-cause mortality and thus survival time after recruitment was ascertained from the date and cause of death reported up until 2018-02-11 for all deceased UK Biobank participants (mean follow-up time 8.90 years).

Computation of mitochondrial abundance in UK Biobank participants

In the UKB participants, we determined somatic mitochondrial DNA abundance [as an approximation of the number of mtDNA copies (Longchamps et al. 2020)] from the intensities of genotyping probes on the MT chromosome. On Affymetrix arrays, the relative amount of DNA hybridized to

the array at each probe is expressed as the log₂ ratio (L2R), which represents the log₂ transformed ratio of the observed genotyping probe intensity divided by the intensity at the same probe observed in a set of reference samples.

Initially, we computed the median L2R values across all 265 variants passing quality control on the MT chromosome with MoChA (MOsaic Chromosomal Alterations caller) (Loh et al. 2020). Similarly, we also computed the median L2R value across 12 probes considered to be probes which are highly predictive of mtDNA abundance on the Affymetrix 6.0 Genome-wide SNP array. This approach is equivalent to the approach used in the MitoPipeline implemented in Genvisis (Tin et al. 2016), however only 12 out of the 25 high-quality probes necessary for the MitoPipeline algorithm are present in the Axiom chips (correlation between both estimates $R = 0.90$).

To compare those array-based estimates to another method of mtDNA abundance determination, we also computed the average coverage of the MT genome normalized to the total genomic coverage as estimated with mosdepth (Pedersen and Quinlan 2018) in around 50,000 individuals with whole-exome sequencing (WES) data in the UK Biobank (Van Hout et al. 2019). Since no capture probes for mtDNA sequences were used for the WES enrichment process, this method quantifies the off-target reads that map to the MT genome. Prior studies indicated that this approach yields accurate mtDNA abundance estimates that are highly correlated to qPCR estimates of mtDNA abundance (Zhang et al. 2017; Picardi and Pesole 2012; D'Erchia et al. 2015). However, we noticed that normalized MT coverage was negatively correlated to the mtDNA abundance values ascertained with MoChA ($R = -0.07$) or with the MitoPipeline approach ($R = -0.06$). This finding indicated that several poorly performing probes were confounding the abundance estimation.

We, therefore, fit a multivariate linear regression model to select intensities that are statistically significantly predicting the normalized MT coverage from exome sequencing data (Supplementary Table S1). We then multiplied the respective L2R value of each probe by the weight in Supplementary Table S1 and computed the median L2R value across those weighted L2R values (mL2RMT), resulting in a single mtDNA abundance estimate for each individual. The distribution was rescaled to have a mean of zero and a standard deviation (SD) of one within each genotyping plate consisting of 96 samples. The weighted mtDNA abundance estimate was positively correlated to the normalized mean mitochondrial exome sequencing coverage with a Spearman rank correlation coefficient of 0.33 (Supplementary Fig. 1), indicating that they should capture the same biological signal, i.e. mtDNA abundance. To assess this and to compare our method of mtDNA abundance quantification to the (unweighted) approaches, we computed the association

of different mtDNA abundance estimates with those established variables (Supplementary Fig. 5) such as age (McArdle et al. 2013; Moore et al. 2018), sex (Knez et al. 2016), white blood cell count (Shim et al. 2020), smoking (Cardellach et al. 2003) as well as BMI (McArdle et al. 2013). We found that the estimates from the weighted computation of mtDNA abundance as well as the normalized MT coverage accurately reflected mtDNA abundance in agreement with prior reports (Supplementary Fig. 5). Importantly, the effect sizes large agreed between those approaches, indicating that they capture the same biological signal that is mtDNA abundance. On the other hand, the unweighted approaches yielded contradictory results and therefore do not agree with prior findings by other groups, suggesting that they are a poor predictor of mtDNA abundance. To further analyse potentially haplogroup specific effects, haplogroups in the UKB participants were estimated with *haplogrep2* (version 2.1.25) (Weissensteiner et al. 2016) from all MT variants genotyped on the chip with standard settings. We only analysed the top-level haplogroups that were present in more than 1000 individuals due to power considerations.

Additional quality control based on the intensity and whole-exome sequencing data in the UKB

Similar to the quality control steps suggested for Illumina genotyping arrays, for each individual, we used MoChA to compute the standard deviation of the L2R values (SDL2R) of all autosomal variants and only included individuals with SDL2R of less than 0.36 (i.e. less than two SDs from the mean SDL2R of all samples), effectively excluding samples with excessive variation in their genotyping intensities. Furthermore, we excluded individuals with an b-allele frequency phase concordance across phased heterozygous sites greater than 0.52 (Vattathil and Scheet 2013). In addition, we also computed the average autosomal DNA abundance from all autosomal L2R values (mL2Rauto) with MoChA. Adjusting for this variable allows us to account for differences in genotyping quality and differences in DNA hybridization to the genotyping chip. Furthermore, we compared the association results from the mtDNA abundance to the overall (autosomal) DNA abundance to identify MT specific effects.

Statistical analyses

We coded the continuous mtDNA abundance (mL2RMT) as the outcome and used linear regression in *R* (function `lm` as implemented in base *R*) to evaluate its correlation with modifiable (life-style) and non-modifiable exposure variables. In the PheWAS, the mtDNA abundance distribution was split into tertiles (cut-offs: -0.43 and 0.39 SD from the mean) and considered the exposure for the diseases (i.e. outcome).

All analyses were adjusted for age at baseline/blood draw, sex (determined from the genotyping data), the first two genotype principal components (PCs), the number of missing genotypes as well as genotyping batch. We also performed some association testing by stratifying patients according to their MT haplogroup. Those analyses were additionally adjusted by the quality (i.e. confidence) of the respective haplotype computation as reported by *haplogrep2*. To determine whether mtDNA abundance is correlated to age in a U-shaped fashion (i.e. differently between pre- and post-menopausal women), we computed a two-lines test. This approach fits two regression models for women above and below a programmatically determined age threshold, respectively, as proposed by Simonsohn (2018). Unless otherwise mentioned, to account for multiple testing (where appropriate), we controlled the false-discovery rate to be less than 5% and thus report results that have a *Q* value of less than 0.05.

Genome-wide association study of mitochondrial abundance

The genome-wide association study (GWAS) of mitochondrial DNA abundance was computed with `plink2` (Chang et al. 2015). In general, rare mutations are harder to call and usually have more missing genotypes. Since mtDNA abundance is associated with missingness, investigating rare variants might result in many false-positive findings, which are reflective of high missingness. We, therefore, restricted the analyses to variants with a minor allele frequency greater than 5% as well as those with less than 2% missing data, a MACH R^2 imputation quality greater than 0.6 as well as not deviating significantly ($P < 10^{-05}$) from Hardy–Weinberg equilibrium.

The association between mtDNA abundance and the additive genotypes in the genome-wide association study was adjusted for by the first 10 genotype PCs, age at baseline, sex, genotyping batch, genotyping missingness/call rate as well as neutrophil percentage, lymphocyte percentage and white blood cell count. The genomic inflation factor λ was calculated with the *GenABEL* library version 1.8-0 implemented in *R*. We considered variants associated with mtDNA abundance with a *P* value lower than 5×10^{-08} to be genome-wide significantly correlated to mtDNA abundance. Functional mapping of associated variants to genes within genome-wide significant loci was done with FUMA (Watanabe et al. 2017). Independent signals within each locus were defined as additional association signals at genome-wide significance and with a correlation (R^2) of less than 0.1 to the variant with the lowest *P* value. In each locus, we considered the variant (as well as correlated variants at $R^2 > 0.6$) within gene bodies with the highest CADD (Rentzsch et al. 2019) score to be the most likely functional variant and also recorded the affected gene (i.e. labelled each locus according

to the gene name). Similarly, for variants outside of the gene body, we extracted the variant with the lowest RegulomeDB (Boyle et al. 2012) score as well as the closest gene to that variant.

We used fastBAT implemented in the Complex-Traits Genetics Virtual Lab (Cuellar-Partida et al. 2019) using the GWAS summary statistics to assess the genetic burden of variants associated with mtDNA abundance within the gene body of all genes. The statistical significance threshold was set at P value $< 2.0 \times 10^{-6}$ which is the Bonferroni correction threshold for 20,000 protein-coding genes. Genes with a statistically significant burden were analysed in a pathway overrepresentation analysis (ORA) with Webgestalt as implemented in WebGestaltR (Liao et al. 2019) in R. Pathways for the ORA were retrieved from MSigDB (Liberzon et al. 2011) version 7.1 and we report pathways with a size of between 25 and 1000 genes and with a Q value of less than 0.05.

Heritability estimation of mitochondrial DNA abundance

The phenotypic variance explained by additive contributions from SNPs (SNP heritability) of mtDNA abundance was estimated with the genomic restricted maximum likelihood estimation method implemented in the genome-wide complex trait analysis in BOLT-LMM (Loh et al. 2018), with standard settings and adjusted for the same covariates as described for the genome-wide association analysis.

PheWAS analysis

Diseases were extracted from the hospital episode spell data of all individuals passing QC and the ICD10 codes were mapped to phecodes (Wu et al. 2019) using the *createPheWASTable* function from the *PheWAS* package (Carroll et al. 2014) implemented in R. This approach combines and maps ICD codes from electronic health records to clinically relevant outcomes, which were manually defined by clinical experts. The hierarchical approach also defines related diseases which are built-in exclusion criteria to prevent contamination of control individuals with cases that have related diseases (Wu et al. 2019). We only considered incident diseases, i.e. diseases that occurred at least once after blood draw/recruitment. We also excluded all patients with prevalent diseases from the respective incident data set so only individuals without pre-existing conditions were evaluated within each respective phecode or phecode grouping which consists of related diseases (Wei et al. 2017). In each association analysis, the controls were individuals that did not have the respective disease or a related disease at baseline nor did they develop the diseases over the follow-up period.

The PheWAS was performed with the *pheWAS* function and we only considered phecodes which occurred in more than 250 cases. In order to reduce the impact of extreme mL2R values and to enable comparison with prior studies (Lan et al. 2008; Kim et al. 2015), we discretized the variable according to tertiles and used this variable as the exposure in the PheWAS. The analyses were adjusted for the same covariates as the genome-wide association study. Results were plotted with the *pheWASManhattan* function.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00439-020-02249-w>.

Acknowledgements The computations and data handling were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at UPPMAX, Uppsala University, partially funded by the Swedish Research Council through grant agreement no. 2018-05973. This research has been conducted using the UK Biobank Resource under Application Number 22224.

Author contributions Conception and design: FG, SH. Financial support: KC, SH. Collection and assembly of data: FG, JJ, SH. Data analysis and interpretation: FG, JJ, YW, SH. Manuscript writing: FG, KC, SH, JJ. Final approval of manuscript: all authors.

Funding Open Access funding provided by Karolinska Institute. Open Access funding provided by Karolinska Institute. This work was financed by the Swedish Research Council (Grant 2018-02547, 2015-03255, 2019-01272, 2018-02077), the Swedish Cancer Society (Grants CAN 2016/684), the Stockholm County Council (Grant no. 20170088), the Karolinska Institutet's Research Foundation (Grant 2018-02146), Karolinska Institutet's Strategic Research Program in Epidemiology, King Gustaf V:s and Queen Victoria's Foundation of Freemasons, and the Åke Wibergs Foundation (M19-0294). FG was a Leopoldina Postdoctoral Fellow (Grant no. LPDS 2018-06) funded by the Academy of Sciences Leopoldina.

Data availability The data were exclusively retrieved from the UK Biobank and can be accessed upon request from the UK Biobank. The mitochondrial DNA abundance as computed in this manuscript will be reported back to the UK Biobank upon publication. The scripts to compute the weights and the weighted mtDNA abundance in the UKB dataset will be published at https://github.com/GrassmannLab/MT_UKB.

Compliance with ethical standards

Conflict of interest The authors affirm that there is no conflict of interest to declare.

Ethics approval UK Biobank has obtained Research Tissue Bank (RTB) approval from its ethics committee that covers approved uses of the Resource.

Informed consent Written informed consent forms were signed by all participants.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source,

provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Al Amir Dache Z, Otandault A, Tanos R et al (2020) Blood contains circulating cell-free respiratory competent mitochondria. *FASEB J* 34:3616–3630
- Ashar FN, Moes A, Moore AZ et al (2015) Association of mitochondrial DNA levels with frailty and all-cause mortality. *J Mol Med (Berl)* 93:177–186
- Astle WJ, Elding H, Jiang T et al (2016) The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* 167:1415–1429.e19
- Bennett FC, Molofsky AV (2019) The immune system and psychiatric disease: a basic science perspective. *Clin Exp Immunol* 19:294–307
- Boyle AP, Hong EL, Hariharan M et al (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 22:1790–1797
- Cai N, Li Y, Chang S et al (2015) Genetic control over mtDNA and its relationship to major depressive disorder. *Curr Biol* 25:3170–3177
- Calvo SE, Mootha VK (2010) The mitochondrial proteome and human disease. *Annu Rev Genom Hum Genet* 11:25–44
- Campa D, Barrdahl M, Santoro A et al (2018) Mitochondrial DNA copy number variation, leukocyte telomere length, and breast cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Breast Cancer Res* 20:29
- Cardellach F, Alonso JR, López S et al (2003) Effect of smoking cessation on mitochondrial respiratory chain function. *J Toxicol Clin Toxicol* 41:223–228
- Carroll RJ, Bastarache L, Denny JC (2014) R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics* 30:2375–2376
- Chang CC, Chow CC, Tellier LC et al (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7
- Chinnery PF, Gomez-Duran A (2018) Oldies but goldies mtDNA population variants and neurodegenerative diseases. *Front Neurosci* 12:682
- Cuellar-Partida G, Lundberg M, Kho PF, D'Urso S, Gutierrez-Mondragon LF (2019) Complex-Traits Genetics Virtual Lab: a community-driven web platform for post-GWAS analyses. <https://doi.org/10.1101/518027>
- Curran JE, Johnson MP, Dyer TD et al (2007) Genetic determinants of mitochondrial content. *Hum Mol Genet* 16:1504–1514
- De Benedictis G, Rose G, Carrieri G et al (1999) Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans. *FASEB J* 13:1532–1536
- D'Erchia AM, Atlante A, Gadaleta G et al (2015) Tissue-specific mtDNA abundance from exome data and its correlation with mitochondrial transcription, mass and respiratory activity. *Mitochondrion* 20:13–21
- Giordano C, Montopoli M, Perli E et al (2011) Oestrogens ameliorate mitochondrial dysfunction in Leber's hereditary optic neuropathy. *Brain* 134:220–234
- Gómez-Durán A, Pacheu-Grau D, Martínez-Romero I et al (2012) Oxidative phosphorylation differences between mitochondrial DNA haplogroups modify the risk of Leber's hereditary optic neuropathy. *Biochim Biophys Acta* 1822:1216–1222
- Guyatt AL, Brennan RR, Burrows K et al (2019) A genome-wide association study of mitochondrial DNA copy number in two population-based cohorts. *Hum Genom* 13:6
- Haas RH (2019) Mitochondrial dysfunction in aging and diseases of aging. *Biology (Basel)* 8:48
- Hulgan T, Kallianpur AR, Guo Y et al (2019) Peripheral blood mitochondrial DNA copy number obtained from genome-wide genotype data is associated with neurocognitive impairment in persons with chronic HIV infection. *J Acquir Immune Defic Syndr* 80:e95–e102
- Ide T, Tsutsui H, Hayashidani S et al (2001) Mitochondrial DNA damage and dysfunction associated with oxidative stress in failing hearts after myocardial infarction. *Circ Res* 88:529–535
- Jylhävä J, Pedersen NL, Hägg S (2017) Biological age predictors. *EBioMedicine* 21:29–36
- Kim C, Bassig BA, Seow WJ et al (2015) Mitochondrial DNA copy number and chronic lymphocytic leukemia/small lymphocytic lymphoma risk in two prospective studies. *Cancer Epidemiol Biomark Prev* 24:148–153
- Kipps TJ, Stevenson FK, Wu CJ et al (2017) Chronic lymphocytic leukaemia. *Nat Rev Dis Prim* 3:16096
- Knez J, Winkelmanns E, Plusquin M et al (2016) Correlates of peripheral blood mitochondrial DNA content in a general population. *Am J Epidemiol* 183:138–146
- Lan Q, Lim U, Liu C-S et al (2008) A prospective study of mitochondrial DNA copy number and risk of non-Hodgkin lymphoma. *Blood* 112:4247–4249
- Lemnrau A, Brook MN, Fletcher O et al (2015) Mitochondrial DNA copy number in peripheral blood cells and risk of developing breast cancer. *Cancer Res* 75:2844–2850
- Liao Y, Wang J, Jaehnig EJ et al (2019) WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res* 47:W199–W205
- Liberzon A, Subramanian A, Pinchback R et al (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27:1739–1740
- Lin BD, Carnero-Montoro E, Bell JT et al (2017) 2SNP heritability and effects of genetic variants for neutrophil-to-lymphocyte and platelet-to-lymphocyte ratio. *J Hum Genet* 62:979–988
- Liou C-W, Lin T-K, Chen J-B et al (2010) Association between a common mitochondrial DNA D-loop polycytosine variant and alteration of mitochondrial copy number in human peripheral blood cells. *J Med Genet* 47:723–728
- Liu C-S, Kuo C-L, Cheng W-L et al (2005) Alteration of the copy number of mitochondrial DNA in leukocytes of patients with hyperlipidemia. *Ann N Y Acad Sci* 1042:70–75
- Loh P-R, Kichaev G, Gazal S et al (2018) Mixed-model association for biobank-scale datasets. *Nat Genet* 50:906–908
- Loh P-R, Genovese G, McCarroll SA (2020) Monogenic and polygenic inheritance become instruments for clonal selection. *Nature* 584:136–141
- Longchamps RJ, Castellani CA, Yang SY et al (2020) Evaluation of mitochondrial DNA copy number estimation techniques. *PLoS ONE* 15:e0228166
- López S, Buil A, Souto JC et al (2014) A genome-wide association study in the genetic analysis of idiopathic thrombophilia project suggests sex-specific regulation of mitochondrial DNA levels. *Mitochondrion* 18:34–40
- López-Otín C, Blasco MA, Partridge L et al (2013) The hallmarks of aging. *Cell* 153:1194–1217
- McArdle MA, Finucane OM, Connaughton RM et al (2013) Mechanisms of obesity-induced inflammation and insulin resistance:

- insights into the emerging role of nutritional strategies. *Front Endocrinol (Lausanne)* 4:52
- McGuire PJ (2019) Mitochondrial dysfunction and the aging immune system. *Biology (Basel)* 8:26
- Moore AZ, Ding J, Tuke MA et al (2018) Influence of cell distribution and diabetes status on the association between mitochondrial DNA copy number and aging phenotypes in the InCHIANTI study. *Aging Cell* 17(1):e12683
- Nejat EJ, Chervenak JL (2010) The continuum of ovarian aging and clinicopathologies associated with the menopausal transition. *Maturitas* 66:187–190
- Okada Y, Kamatani Y, Takahashi A et al (2010) Common variations in PSMD3-CSF3 and PLCB4 are associated with neutrophil count. *Hum Mol Genet* 19:2079–2085
- Pedersen BS, Quinlan AR (2018) Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* 34:867–868
- Phipps AI, Ichikawa L, Bowles EJA et al (2010) Defining menopausal status in epidemiologic studies: a comparison of multiple approaches and their effects on breast cancer rates. *Maturitas* 67:60–66
- Picardi E, Pesole G (2012) Mitochondrial genomes gleaned from human whole-exome sequencing. *Nat Methods* 9:523–524
- Pyle A, Anugraha H, Kurzawa-Akanbi M et al (2016) Reduced mitochondrial DNA copy number is a biomarker of Parkinson's disease. *Neurobiol Aging* 38:216.e7–216.e10
- Rentzsch P, Witten D, Cooper GM et al (2019) CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 47:D886–D894
- Rokkam VR, Kotagiri R (2020) Secondary thrombocytosis. In: *StatPearls* [Internet]. StatPearls Publishing, Treasure Island (FL). <https://www.ncbi.nlm.nih.gov/books/NBK560810/>
- Shen J, Platek M, Mahasneh A et al (2010) Mitochondrial copy number and risk of breast cancer: a pilot study. *Mitochondrion* 10:62–68
- Shen J, Wan J, Song R et al (2015) Peripheral blood mitochondrial DNA copy number, length heteroplasmy and breast cancer risk: a replication study. *Carcinogenesis* 36:1307–1313
- Shim HB, Arshad O, Gadawska I et al (2020) Platelet mtDNA content and leukocyte count influence whole blood mtDNA content. *Mitochondrion* 52:108–114
- Simonsohn U (2018) Two lines: a valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Adv Methods Pract Psychol Sci* 1:538–555
- Sudlow C, Gallacher J, Allen N 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
- Thompson DJ, Genovese G, Halvardson J et al (2019) Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* 575:652–657
- Thyagarajan B, Wang R, Barcelo H et al (2012) Mitochondrial copy number is associated with colorectal cancer risk. *Cancer Epidemiol Biomark Prev* 21:1574–1581
- Thyagarajan B, Wang R, Nelson H et al (2013) Mitochondrial DNA copy number is associated with breast cancer risk. *PLoS ONE* 8:e65968
- Tin A, Grams ME, Ashar FN 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
- Torrioni A, Huoponen K, Francalacci P et al (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835–1850
- Van Hout CV, Tachmazidou I, Backman JD et al (2020) Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* 586:749–756
- Vattathil S, Scheet P (2013) Haplotype-based profiling of subtle allelic imbalance with SNP arrays. *Genome Res* 23:152–158
- Watanabe K, Taskesen E, van Bochoven A et al (2017) Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 8:1826
- Wei W-Q, Bastarache LA, Carroll RJ et al (2017) Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS ONE* 12:e0175508
- Weissensteiner H, Pacher D, Kloss-Brandstätter A et al (2016) HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* 44:W58–63
- Welsh S, Peakman T, Sheard S et al (2017) Comparison of DNA quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. *BMC Genom* 18:26
- Williams DM, Jylhävä J, Pedersen NL et al (2019) A frailty index for UK biobank participants. *J Gerontol A Biol Sci Med Sci* 74:582–587
- Winkler TW, Day FR, Croteau-Chonka DC et al (2014) Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 9:1192–1212
- Workalemahu T, Enquobahrie DA, Yohannes E et al (2016) Placental telomere length and risk of placental abruption. *J Matern Fetal Neonatal Med* 29:2767–2772
- Workalemahu T, Enquobahrie DA, Tadesse MG et al (2017) Genetic variations related to maternal whole blood mitochondrial DNA copy number: a genome-wide and candidate gene study. *J Matern Fetal Neonatal Med* 30:2433–2439
- Wu P, Gifford A, Meng X et al (2019) Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR Med Inform* 7:e14325
- Xu E, Sun W, Gu J et al (2013) Association of mitochondrial DNA copy number in peripheral blood leukocytes with risk of esophageal adenocarcinoma. *Carcinogenesis* 34:2521–2524
- Zhang P, Lehmann BD, Samuels DC et al (2017) Estimating relative mitochondrial DNA copy number using high throughput sequencing data. *Genomics* 109:457–462

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.