

## First mitochondrial genome wide association study with metabolomics

Brahim Aboulmaouahib\*<sup>1,12</sup>, Gabi Kastenmüller<sup>2</sup>, Karsten Suhre<sup>3</sup>, Sebastian Zöllner<sup>4,5</sup>, Hansi Weissensteiner<sup>6</sup>, Christian Gieger<sup>7,8,9</sup>, Rui Wang-Sattler<sup>7,8,9</sup>, Peter Lichtner<sup>10</sup>, Konstantin Strauch<sup>1,11,12</sup>, Antònia Flaquer<sup>1,11</sup>

UNCORRECTED MANUSCRIPT

Downloaded from <https://academic.oup.com/mmg/advance-article/doi/10.1093/mmg/ddab312/6408939> by GSF Zentralbibliothek user on 16 December 2021

1. Institute for Medical Information Processing, Biometry and Epidemiology - IBE, LMU, Munich, Germany.
2. Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
3. Department of Physiology and Biophysics, Weill Cornell Medical College in Qatar, State of Qatar.
4. Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America.
5. Department of Psychiatry, University of Michigan, Ann Arbor, Michigan, United States of America.
6. Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, Innsbruck, Austria

7. Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
8. German Center for Diabetes Research (DZD), Neuherberg, Germany.
9. Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
10. Institute of Human Genetics, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
11. Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany.
12. Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany.

**Corresponding author:** Brahim Aboulmaouhib MSc, Institute for Medical Information Processing, Biometry and Epidemiology - IBE, Ludwig-Maximilians-Universität, Marchioninistrasse 15, 81377, Munich, Germany.  
Present address: Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center Mainz, 55101 Mainz, Germany, E-mail: baboulma@uni-mainz.de

## Abstract

**Introduction:** In the era of personalized medicine with more and more patient specific targeted therapies being used, we need reliable, dynamic, faster, and sensitive biomarkers both to track the causes of disease and to develop and evolve therapies during the course of treatment. Metabolomics recently has shown substantial evidence to support its emerging role in disease diagnosis and prognosis. Aside from biomarkers and development of therapies, it is also an important goal to understand the involvement of mitochondrial DNA mtDNA in metabolic regulation, aging, and disease development. Somatic mutations of the mitochondrial genome are also heavily implicated in age-related disease and aging. The general hypothesis is that an alteration in the concentration of metabolite profiles (possibly conveyed by lifestyle and environmental factors) influences the increase of mutation rate in the mtDNA, and thereby contributes to a range of pathophysiological alterations observed in complex diseases.

**Methods:** We performed an inverted mitochondrial genome wide association analysis between mitochondrial nucleotide variants (mtSNVs) and concentration of metabolites. We used 151 metabolites and the whole sequenced mitochondrial genome from 2,718 individuals to identify genetic variants associated with metabolite profiles. Because of the high coverage, next generation sequencing-based analysis of the mitochondrial genome allows for an accurate detection of mitochondrial heteroplasmy and for identification of variants associated with the metabolome.

**Results:** The strongest association was found for mt715<sub>G>A</sub> located in the MT-12SrRNA with the metabolite ratio C2/C10:1 (p-value=6.82\*10<sup>-09</sup>, β=0.909). The second most significant mtSNV was found for mt3714<sub>A>G</sub> located in the MT-ND1 with the metabolite ratio PC aa C42:5/PC aa C44:5 (p-value=1.02\*10<sup>-08</sup>, β=3.631). A large number of significant metabolite ratios were observed involving PC aa C36:6 and the variant mt10689<sub>G>A</sub>, located in the MT-ND4L gene.

**Conclusion:** These results show an important interconnection between mitochondria and metabolite concentrations. Considering that some of the significant metabolites found in this study have been previously related to complex diseases such as neurological disorders and metabolic conditions, these associations found here might play a crucial role for further investigations of such complex diseases. Understanding the mechanisms that control human health and disease, in particular the role of genetic predispositions and their interaction with environmental factors is a prerequisite for the development of safe and efficient therapies for complex disorders.

## Introduction

Genome-wide association studies (GWAS) have identified a number of genetic polymorphisms that convey an increased risk for developing diseases [1]. However, the results of GWAS have not fully responded to initial expectations. The biological effect of many genes found associated with a particular disease cannot be explained so far. More surprising is the fact that, even when combining all available GWAS on a particular disorder, polymorphisms usually associated explain less than 5%–10% of the risk of disease [1].

However, by only associating genotypes with clinical outcomes, little can be inferred about the underlying disease-causing mechanisms. On the other hand, GWAS with metabolic traits as intermediate phenotypes (GWAS-metabolomics) have identified genetically determined variants in metabolic phenotypes that exhibit large effect sizes [2] [3].

Recently, GWAS on the mitochondrial genome (mtGWAS) have also provided some new insights into the mechanisms underlying several complex diseases [4] [5] [6]. Genetic variants that associate with changes in the homeostasis of key lipids, carbohydrates or amino acids are not only expected to display much larger effect sizes due to their direct involvement in metabolite conversion and modification, but may also provide more details on potentially affected pathways and may be more directly related to the etiology of the disease [2].

Mitochondrial DNA (mtDNA), which has a length of ~ 16.6 kb, codes for 13 subunits of the mitochondrial respiratory chain complexes, 2 ribosomal (rRNA) genes, and 22 transfer RNA (tRNA) genes that are required for mitochondrial protein synthesis. Mitochondria consume

oxygen and substrates to generate energy in the form of adenosine triphosphate (ATP) while producing reactive oxygen species (ROS), also known as free radicals, in the process. While ROS have important roles in cell signaling and homeostasis, ROS levels can increase drastically and cause significant damage to the DNA, thereby altering the membrane permeability and calcium homeostasis, along with increasing the heteroplasmic mtDNA. This damage is termed oxidative stress. High levels of oxidative stress may damage cellular structures which, in turn, can lead to mitochondrial dysfunction, possibly even to apoptosis. Mitochondrial mutations can either be somatic or inherited through the maternal line [7]. Since many mtDNA copies are present in a cell and due to their high mutation rate (on average twenty times higher than for nuclear DNA), new somatic mutations may arise only in a subfraction of mtDNA molecules, consequently mutant and wild-type mtDNA can co-exist [8]. This effect is called heteroplasmy. mtDNA heteroplasmy varies among different human tissues and increases with age, suggesting that the majority of mtDNA heteroplasmies are acquired, rather than inherited [9]. Research suggests that the progressive accumulation of mutations in the mitochondria over a person's lifetime may play a central role in aging and in many human diseases.

A relationship that has not been investigated to date is the one between mitochondrial genetic variants and metabolomics (mtGWAS-metabolomics). The most prominent role of mitochondria is not only to generate large quantities of ATP, but also to regulate many metabolic tasks [10] [11] such as signaling through mitochondrial ROS [12], regulation of the membrane potential [11], apoptosis-programmed cell death [13] [14], regulation of cellular metabolism [14], certain heme synthesis reactions [15], and steroid synthesis [16].

The general hypothesis is that an alteration in the concentration of metabolite profiles (possibly conveyed by lifestyle and environmental factors) influences the increase of mutation rate in the mtDNA, and thereby contributes to a range of pathophysiological alterations

observed in complex diseases. Based on this hypothesis, the rationale for the present research is the establishment of mitochondrial and metabolomic signatures as a panel of candidate biomarkers for prediction and early diagnosis of disease as well as monitoring of disease progression.

In this study we conducted association analyses between 151 metabolites and the whole sequenced mitochondrial genome from 1,163 individuals to identify genetic variants influencing metabolite profiles. Because of the high coverage, next generation sequencing (NGS) allows us a deeper and more accurate analysis of the mitochondrial genome for identification of disease-causing variants and heteroplasmy than mitochondrial nucleotide variants (mtSNV) allele intensities measured by genotyping arrays.

## Results

We conducted a genome-wide association study with metabolite ratios using 151 metabolic traits in a group of 1,163 KORA-F4 individuals with sequenced mitochondria covering 9,172 mtSNVs. Linear regression analysis was also conducted for each of 151 metabolites, but the results were by all individual metabolite not significant.

The significant P-values are plotted in Figure 2. The x axis represents the mitochondrial genome, showing the position and relative size of each of the 13 major mitochondrial genes. A mtSNV was considered significant not only when the p-value was less than  $1.257545 \cdot 10^{-05}$  after  $M_{\text{eff}}$  correction but also when the metabolite ratio p-gain was greater than 151, i.e., the total number of metabolites. In total, we observed 404 mtSNVs that display metabolite ratio associations at a genome-wide significance level. Table 1 shows the information of the 40 most significant results.



The most significant mtSNV, mt715<sub>G>A</sub> located in the MT-12SrRNA, was associated with the metabolite ratio Acetylcarnitine/ Decenoylcarnitine (p-value=6.82\*10<sup>-9</sup>, β=0.909). Following in significance is mt3714<sub>A>G</sub> located in the MT-ND1, which was associated with the metabolite ratio Phosphatidylcholine acyl-alkyl C42:5 / Phosphatidylcholine acyl-alkyl C44:5 (p-value=1.02\*10<sup>-8</sup>, β=3.631).

More details about the estimated model parameters for each significant mtSNV are given in Supplementary Table 1.

It can be observed that a large percentage (15%) of the most significant mtSNVs are located in the MT-ND4L (Figure 3). They are mostly associated with metabolites from the glycerophospholipid class. The variant mt10689<sub>G>A</sub>, located in the ND4L gene is of special interest because it is associated with 16 different metabolite ratios, making it the most common multi-associated mtSNV in our dataset. Moreover, in all these 16 ratios the metabolite Phosphatidylcholine diacyl C36:6 (PC aa C36:6) is involved (Figure 2).

## Discussion

This study was carried out with the aim of investigating the relationship between the metabolite ratios and the genetic variants of the mtDNA. The general hypothesis is that an alteration in the concentration of metabolite profiles (possibly conveyed by lifestyle and environmental factors) influences the increase of mutation rate in the mtDNA, and thereby contributes to a range of pathophysiological alterations observed in complex diseases.

### **Decenoylcarnitine / Acetylcarnitine associated with MT-12SrRNA<sub>mt715G>A</sub> gene**

The strongest association was found for mt715<sub>G>A</sub> located in the MT-12SrRNA with the metabolite ratio Acetylcarnitine (C2) / Decenoylcarnitine (C10:1)

(p-value=6.82\*10<sup>-9</sup>, β=0.909). This association may play a crucial role in the regulation of insulin secretion.

C2 is broken down in the blood by plasma esterases to carnitine which is used by the body to transport fatty acids into the mitochondria for breakdown and thus may have a role in normalizing intracellular lipid metabolism. It is known that C2 levels are low in patients with type 1 and type 2 diabetes mellitus [17]. Furthermore, therapy with C2 ameliorated arterial hypertension, insulin resistance, impaired glucose tolerance, and hypoadiponectinemia in subjects at increased cardiovascular risk [18]. It has also been reported that oral administration of acetylcarnitine increases insulin sensitivity and glucose tolerance in individuals with a low glucose disposal rate [19].

C10:1 is considered to be a fatty ester lipid molecule. This metabolite was found to be associated not only with type 2 diabetes but also with prediabetic states [20]. In addition, C10:1 is

used in the diagnosis of genetic disorders such as fatty acid oxidation disorders [21] in which insulin plays a crucial role.

In humans, MT-12SrRNA is encoded by the MT-RNR1 gene. The MT-RNR1 gene encodes for a protein responsible for regulating insulin sensitivity and metabolic homeostasis [22]. The protein acts as an inhibitor of the folate cycle, thereby reducing de novo purine biosynthesis which leads to the accumulation of the de novo purine synthesis intermediate 5-aminoimidazole-4-carboxamide and the activation of the metabolic regulator 5'-AMP-activated protein kinase.

Interestingly, the concentration of C2 and C10:1 have been observed to increase during physical exercise [23] [24]. It is well known that insulin levels typically decrease during exercise.

In this context, our results show that an alteration of C2/C10:1 levels in serum increases the ratio of heteroplasmy at mt715<sub>G>A</sub> which, in turn, can lead to MT-12SrRNA dysfunction, possibly even to affect insulin regulation.

### **Phosphatidylcholine acyl-alkyl C42:5 (PC ae C42:5) / Phosphatidylcholine acyl-alkyl C44:5 (PC ae C44:5) is associated with the MT-ND1<sub>mt3714A>G</sub> gene**

The second most significant mtSNV was found for mt3714<sub>A>G</sub> located in the MT-ND1 with the metabolite ratio PC ae C42:5 / PC ae C44:5 (p-value=1.02\*10<sup>-08</sup>, β=3.631). This association may play a crucial role in disorders related to damage or disease that affects the brain.

Both metabolites involved in this ratio are Phosphatidylcholines (PC) included in the glycerophospholipid class. Glycerophospholipids are the main component of biological membranes providing them with stability, fluidity, and permeability. Glycerophospholipids are actively catabolized by brain tissue and involved in the regulation of several molecular functions, such

as generation of second messengers, apoptosis, antioxidant and membrane fusion, and regulation of enzyme activities [25]. A marked deficit in glycerophospholipid concentration may be responsible for the neurodegeneration observed in neurological disorders [26]. In particular, alteration in PC metabolism has been observed in Alzheimer, Parkinson, and with an increased risk of dementia [27] [28] [29]. Several studies have used dietary intake of PC in clinical trials treating brain diseases resulting with an improve of the memory, capability of learning, and cognitive performances [30] [31] [28]. Recently, supplementation with PC has been also associated with a reduced risk of dementia [29].

The MT-ND1 gene encodes for the NADH-ubiquinone oxidoreductase chain 1 protein, a subunit of NADH dehydrogenase, which is located in the mitochondrial inner membrane. Genetic mutations of the human MT-ND1 gene have been associated with several inborn genetic disorders such as MELAS, Leigh's syndrome and Leber's hereditary optic neuropathy [32] [33]. All these inborn disorders are related to damage or disease that affects the brain. Also a decreased DNA methylation has been observed in the MT-ND1 gene in cases with early stage of Alzheimer's disease [34].

Based on our results an alteration in PC ae C42:5 and PC ae C44:5 increases heteroplasmy at mt3714<sub>A>G</sub> which, in turn, can lead to MT-ND1 dysfunction. This represents a pathway that may help to understand molecular aspects of neurodegeneration diseases.

#### **MT-ND4L - mitochondrially encoded NADH: ubiquinone oxidoreductase core subunit**

##### **4L**

It should be noted that the largest number of significant associations between metabolite ratios and mtSNVs are located in the MT-ND4L gene (Figure 2 and Figure 3). The variant mt10689<sub>G>A</sub>, located in the ND4L gene, associated with 16 metabolite ratios is the most common multi-associated mtSNV in our dataset. Moreover, in all these 16 ratios the metabolite Phosphatidylcholine diacyl C36:6 (PC aa C36:6) is involved.

Despite the crucial importance of PC for brain functioning, disturbance of PC concentration has also been found to be associated with various metabolic disorders such as atherosclerosis [35], insulin resistance [35] [36], Gaucher disease [37], and obesity in adults [38] [39] as well as in children [40]. Particularly the metabolite PC aa C36:6 has been associated with different patterns of fat concentration in the body such as visceral fat and liver fat content [41]. PC aa C36:6 was also involved in three metabolite ratios previously shown to be associated with Fat Free Mass Index [42].

MT-ND4L is a subunit of NADH dehydrogenase, which is the component of the electron transport chain that is responsible for the oxidative phosphorylation (OXPHOS) process. Its dysfunction may cause energy deficiency in cells, resulting in metabolic disorders such as obesity and diabetes [43] [44] [45]. In fact, several variants of human MT-ND4L have been reported to be associated with altered metabolic conditions like BMI and type 2 diabetes [46] [43]. Changes in MT-ND4L gene expression have long-term consequences on energy metabolism and have been suggested to be a major predisposition factor for the development of metabolic syndrome [47].

Based on our results in 16 metabolite ratios where PC aa C36:6 is involved, alteration of its concentration changes significantly the heteroplasmy in the variant mt10689<sub>G>A</sub>, located in the ND4L gene. Put together this association may explain an interesting pathway to understand the development of metabolic conditions.

## **Conclusion**

We presented the results of the first association analysis between genetic positions in the mitochondrial genome and metabolite profiles using sequencing data of the mitochondrial genome and an inverted mtGWAS. Although further analyses are needed to follow up on the present

results, these findings highlight the important role of the mtDNA and the field of metabolomics among the factors that contribute to the balance of the human metabolism and suggest that variants in the mitochondrial genome may be more important than has previously been suspected.

Our results support the idea that a strong relationship exists between mtDNA heteroplasmy and metabolite levels. Both are likely to play a crucial role as discriminating cofactors in the etiology of common multi-factorial diseases. A conceivable hypothesis is that high levels of oxidative stress might be produced when altered levels of certain metabolites (possibly due to environmental factors such as nutrition or life style) alter the permeability of cell membranes, along with increasing the heteroplasmic mtDNA and weakening the mitochondrial defense systems. High levels of oxidative stress damage cellular structures including the mitochondria themselves which, in turn, can lead to mitochondrial dysfunction, possibly even to apoptosis.

In the era of personalized medicine with more and more patient specific targeted therapies being used, we need reliable, dynamic, faster, and sensitive biomarkers both to track the causes of disease and to develop and evolve therapies during the course of treatment. Understanding the mechanisms that control human health and disease, in particular the role of genetic predispositions and their interaction with environmental factors, is a prerequisite for the development of safe and efficient therapies for complex disorders.

We conclude that recent advances in metabolomics and next generation sequencing (NGS) technology along with novel strategies to analyze and understand the metabolic pathways as well as to integrate metabolite networks with mitochondrial genetic data, opens this window of opportunity to identify new biomarkers related to the human mitochondrial DNA, both to track complex diseases and to develop and evolve option of treatment.

## **Materials and Methods**

### *Study design and population*

The Cooperative Health Research in the Region of Augsburg (KORA) study is a series of independent population-based epidemiological surveys and follow-up studies of participants living in the region of Augsburg, in southern Germany, an area with demographic and socioeconomic characteristics roughly reflecting those of an average central European population. All participants are residents of German nationality identified through the registration office and written informed consent was obtained from each participant [48]. The study was approved by the local ethics committee (Bayerische Landesärztekammer). All participants filled in a self-administrated questionnaire and underwent a standardized personal interview and an extensive medical examination. Detailed phenotypes and personal information related to life circumstances, history of disease and medication were recorded in a computer-assisted personal interview. All procedures were subjected to quality assessment. The study design, sampling method, and data collection have been described in detail elsewhere [49]. The participants of this study were selected from the KORA F4 (2006-2008) study including a total number of 3,021 unrelated individuals. KORA F4 is the first follow-up study of the population-based survey KORA S4 (1999-2001). All procedures were subjected to quality assessment. No evidence of population stratification has been found in multiple published analyses using genetic data of the KORA cohort.

### *Metabolite profiles*

Metabolomic measurements have been performed for 3,061 individuals of KORA F4 population-based sample. Men and women were collected in a random order and samples were randomly put on plates to reduce batch effects. Metabolites were measured in serum using the AbsoluteIDQ™ p150 Kit (BIOCRATES Life Sciences AG, Innsbruck, Austria), as described elsewhere [3]. The panel includes 151 metabolites spanning several metabolic classes: 1 hexose, 35 acylcarnitines, 14 amino acids, 14 sphingomyelins, and 87 glycerophospholipids.

### *Sequencing of the mitochondrial genome*

The mitochondrial genome was sequenced for 3,021 KORA-F4 individuals using DNA derived from peripheral blood mononuclear cells (PBMCs). A long-range PCR approach, which refers to the amplification of DNA fragments of a size that may not be amplified using conventional PCR reagents, has been performed using a highly processive polymerase mixture and novel primer pairs to specifically amplify the mitochondrial genome [50]. These samples were subsequently processed with Illumina® Nextera® XT (catalog # FC-131-1096). This NGS library preparation kit employs an engineered Transposome™ to randomly fragment and tag amplicons and small genomes with Illumina® specific adapters. After library preparation, samples were sequenced on the Illumina® MiSeq™. This method generated whole mitochondrial genome next-generation sequencing (NGS) data, which accurately reflects the Sanger sequence. The 16,569 positions have been sequenced with mtSNV covered with 3,500 fold on average. Because of this high coverage, NGS will allow us a much deeper analysis of the mitochondrial genome and heteroplasmy for identification of disease-causing variants than array-based genotyping of mitochondrial SNVs. It has been observed that NGS data are highly reproducible and very reliable, and, overall, it has been found to be superior to the data produced by microarray technology [51].



### *Quality control*

Before analyzing the data, we performed a two-step quality control. In QC step1 we checked the sequencing data while in QC step2 we focused on the quality of the metabolites (see Figure 1).

During QC step1 a total number of 258 out of 3,021 individuals were excluded. 37 individuals were excluded due to DNA contamination and 8 individuals because of an overall mean coverage lower than 1,000. In order to avoid confounding with insulin-dependent diabetes mellitus, 213 individuals diagnosed with type 2 diabetes were not included in the study. The mitochondrial region covered by the PCR primers spanning 49 bases, from position 16,401 to 16,449, had to be excluded from the analysis as it does not contain any reliable information. Position mt3107 was removed due to its missing value in the reference base. In addition, the three known phantom mutations (mt3166, mt3170 and mt16390) were also removed together with the five positions (mt301, mt302, mt310, mt316, and mt16182) in blacklisted sites as recommended by GATK [52]. Further 233 positions were excluded because of mean coverage lower than 1,000. As we were primarily interested in somatic mtDNA variants, 7,106 homoplasmic positions were also excluded since for association analysis they do not provide any type of information. The threshold to call a position as homoplasmic for the reference allele was 0.007(see Supplementary Material 1). Afterwards, we searched for the presence of nuclear-mitochondrial-DNA fragments (NUMTS) using mtDNA-server [53]. Because NUMTS are copies of mtDNA sequences inserted in the nuclear DNA (nDNA) over evolutionary time, these sequences are possible handicaps for data processing, as the sequenced samples might contain some nDNA fragments. If they are falsely counted as mtDNA fragments, the allele frequencies could be erroneous and consequently incorrect heteroplasmy values would be determined. We identified a total number of 343 NUMTS.

However, in our case, a PCR with primers specific to the mtDNA has been performed in the beginning, and so it is very unlikely that the sequenced fragments contain nDNA. Therefore, these NUMTS do not need to be removed from the analysis because they are adding useful information. After QC step1 a total number of 9,172 mtSNVs and 2,763 individuals are considered for QC step2.

During QC step2, metabolites were imputed using multiple imputation by chained equations (MICE) with the method called "predictive mean matching". Missing values are frequently observed in high-throughput mass spectrometry-based metabolomics measurements, which is the approach used by Biocrates. To this end, we have used the same approach as in previous analyses of the Biocrates metabolomics data in KORA [2] [36]. Following the criteria by [54] to avoid spurious false-positive associations due to small sample sizes, at least 80% non-missing values are required for a metabolite to be included, and data points of metabolic traits lying more than three standard deviations off the mean need to be excluded. All of the 151 available metabolites satisfied this criterion.

After quality control, individuals with available mtDNA data and metabolite values were considered in the analysis. Association analyses was performed in a total number of 9,172 mtSNVs and 151 metabolites in a sample of 1,163 KORA-F4 individuals.

### *Statistical methods*

GWAS have been developed as a method to identify genetic loci associated with disease by scanning multiple markers across the genome. The advance in genotyping technology has led to significant advances in the genetics of complex diseases. Recently, NGS has gained popularity through its capacity to analyze a much greater number of markers across the

genome. One of the peculiarities of the mtDNA is the heteroplasmy. Due to heteroplasmy, mtDNA tends to be heterogeneous in the sense that different mitochondria can have different genotypes, such that a genotype at an mtDNA locus may not be restricted to 0, 1, or 2 minor alleles. This issue affects the possibility of estimating genotypes and makes calling algorithms useless. The finest way to approach mitochondrial heteroplasmy is utilize high-throughput sequencing data, which readily provides data of the entire mitochondrial genome and offers the prerequisite for the detection of low-level mitochondrial genetic heteroplasmy.

In our analysis we take the heteroplasmy into account by using the raw sequence reads, i.e., the number of counts for each variant, instead of calling the variants into a few categories.

The best way to assess the association of metabolites with the mtSNV is to apply linear regression analysis. The mtSNV enters the model as the response via the  $\log_2$ -transformed ratio,  $y = \log_2(B/A)$ , where A denotes the number of counts of the reference allele and B the second most frequent allele. In the ratios, to avoid undefined logarithms in those individuals that only carry the reference allele for the mtSNV under study, a term of 2 additional counts was added in the denominator and numerator. Considering that the average number of counts is about 3,500 adding two counts to A and B will not have any impact in the final results.

Sex and age at examination are introduced in the model as covariates. To improve the convergence properties of model estimates, the total coverage (Cov) of the mtSNV (i.e., A+B) was added in the model. In addition, to avoid false positives due to a batch effect during mtDNA sequencing, the effect of the index of the sequencing plate ( $\text{Batch}_{\text{mtDNA}}$ ) was included in the model as well.

It has been suggested several times that the use of ratios of two metabolite concentrations, which serve as proxies for enzymatic reaction rates, not merely reduces the variance and increases the power of GWAS thus yielding robust statistical associations, but as well provides new information about possible metabolic pathways [2] [54] [55]. Therefore, we

considered all possible metabolite concentration ratios ( $MR_1, \dots, MR_{(N*(N-1))/2}$ ), where  $N$  is the number of single metabolites, with the linear model for association analysis with all mtSNVs. This hypothesis-free approach highlights pairs of metabolites that are more likely to be coupled either biochemically or physiologically [3].

The regression model for the  $j$ -th metabolite ratio and the  $i$ -th mtSNV is:

$$y_{ik} = \beta_0 + \beta_1 MR_{jk} + \beta_2 age_k + \beta_3 sex_k + \beta_4 Cov_{ik} + \beta_5 (Batch_{mtDNA})_k + e_{ijk}$$

$$for\ j=1, \dots, m; i=1, \dots, s; k=1, \dots, n,$$

where  $k$  denotes the individual and  $n$  is the sample size,  $m$  the number of metabolite ratios ( $(N*(N-1))/2$ ), and  $s$  the number of mtSNVs to be tested. The aim of this analysis is to evaluate how the mitochondrial heteroplasmy is influenced by the metabolite ratio, representing the phenotype of interest, taking into account the age and sex of a particular person, and therefore the model adjusts for individual differences regarding these two variables in the study population. GWAS has always been referred to analyses that use genetic variance as predictors in models. In our analysis, the general hypothesis is that an alteration in the concentration of metabolite profiles influences the increase of mutation rate in the mtDNA. In this situation it is more appropriate to use the genetic variants as outcome variables. For that reason, we will refer to this approach as “inverted mtGWAS”.

#### *Correction for multiple testing*

P-values are obtained from a Wald-test based on the asymptotic normality of regression coefficient estimates, and corrected for multiple comparisons with the correction factor being equal to the effective number of independent tests ( $M_{eff}$ ) [56]. We calculated the  $M_{eff}$  measure using the Matrix Spectral Decomposition (matSpDlite) from Nyholt et al. [57] with the method of Li and Ji [58]. We used the p-gain statistics [2] to quantify the decrease in p-value for the association with the ratio compared to the p-values of the two corresponding concentrations. This limit is considered a Bonferroni-type conservative cutoff for identifying whether

a ratio between two metabolite concentrations improves the strength of association compared to the two corresponding metabolite concentrations alone. In our case p-gain should be larger than the number of tested metabolic traits ( $p\text{-gain} > 151$ ). To estimate whether deviation from normality of metabolite ratios may have biased our results we tested associations for both untransformed and log-scaled ratios, not detecting substantial differences. Quantile-quantile (Q-Q) plots were used to examine P value distribution, and the lambda ( $\lambda$ ) ranged from 0.93 to 1.26 (see supplementary Figure 1). All statistical analyses were performed using RStudio version 0.98.1103 that uses R version 3.5.2 [59].

### **Acknowledgments**

We thank all members of the field staff involved in the KORA study and we express our appreciation to all study participants. The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München—German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. This work was supported by a grants of the Deutsche Forschungsgemeinschaft (German Research Foundation) [FL 776/6-1].

## References

- [1] J. A. Riancho, "Genome-wide association studies (GWAS) in complex diseases: advantages and limitations," *Reumatol Clin*, vol. 8, no. 2, pp. 56--57, 2012.
- [2] Gieger, Christian and Geistlinger, Ludwig and Altmaier, Elisabeth and De Angelis, Martin Hrabe and Kronenberg, Florian and Meitinger, Thomas and Mewes, Hans-Werner and Wichmann, H-Erich and Weinberger, Klaus M and Adamski, Jerzy and others, "Genetics meets metabolomics: a genome-wide association study of metabolite profiles in human serum," *PLoS Genet*, vol. 4, no. 11, p. e1000282, 2008.
- [3] Illig, Thomas and Gieger, Christian and Zhai, Guangju and Romisch-Margl, Werner and Wang-Sattler, Rui and Prehn, Cornelia and Altmaier, Elisabeth and Kastenmüller, Gabi and Kato, Bernert S and Mewes, Hans-Werner and others, "A genome-wide perspective of genetic variation in human metabolism," *Nature genetics*, vol. 42, no. 2, pp. 137--141, 2010.
- [4] Flaquer, Antonia and Baumbach, Clemens and Kriebel, Jennifer and Meitinger, Thomas and Peters, Annette and Waldenberger, Melanie and Grallert, Harald and Strauch, Konstantin, "Mitochondrial genetic variants identified to be associated with BMI in adults," *PloS one*, vol. 9, no. 8, p. e105116, 2014.
- [5] Hudson, Gavin and Gomez-Duran, Aurora and Wilson, Ian J and Chinnery, Patrick F, "Recent mitochondrial DNA mutations increase the risk of developing common late-onset human diseases," *PLoS Genet*, vol. 10, no. 5, p. e1004369, 2014.
- [6] Flaquer, A and Baumbach, C and Ladwig, KH and Kriebel, J and Waldenberger, M and Grallert, H and Baumert, J and Meitinger, T and Kruse, J and Peters, A and others, "Mitochondrial genetic variants identified to be associated with posttraumatic stress disorder," *Translational psychiatry*, vol. 5, no. 3, pp. e524--e524, 2015.
- [7] Giles, Richard E and Blanc, Hugues and Cann, Howard M and Wallace, Douglas C, "Maternal inheritance of human mitochondrial DNA," *Proceedings of the National academy of Sciences*, vol. 77, no. 11, pp. 6715--6719, 1980.
- [8] Wallace, Douglas C and Chalkia, Dimitra, "Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease," *Cold Spring Harbor perspectives in biology*, vol. 5, no. 11, p. a021220, 2013.
- [9] Floros, Vasileios I and Pyle, Angela and Dietmann, Sabine and Wei, Wei and Tang, Walfred CW and Irie, Naoko and Payne, Brendan and Capalbo, Antonio and Noli, Laila and Coxhead, Jonathan and others, "Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos," *Nature cell biology*, vol. 20, no. 2, pp. 144--151, 2018.
- [10] Kroemer, Guido and Reed, John C, "Mitochondrial control of cell death," *Nature medicine*, vol. 6, no. 5, pp. 513--519, 2000.
- [11] Voet, Donald and Voet, Judith G and Pratt, Charlotte W, *Fundamentals of biochemistry: life at the molecular level*, John Wiley & Sons, 2016.
- [12] Li, Xinyuan and Fang, Pu and Mai, Jietang and Choi, Eric T and Wang, Hong and Yang, Xiao-feng, "Targeting mitochondrial reactive oxygen species as novel therapy for inflammatory diseases and cancers," *Journal of hematology & oncology*, vol. 6, no. 1, pp. 1--19, 2013.
- [13] Green, Douglas R, "Apoptotic pathways: the roads to ruin," *Cell*, vol. 94, no. 6, pp. 695--698, 1998.

- [14] McBride, Heidi M and Neuspiel, Margaret and Wasiak, Sylwia, "Mitochondria: more than just a powerhouse," *Current biology*, vol. 16, no. 14, pp. R551--R560, 2006.
- [15] Oh-Hama, Tamiko, "Evolutionary consideration on 5-aminolevulinic synthase in nature," *Origins of Life and Evolution of the Biosphere*, vol. 27, no. 4, pp. 405--412, 1997.
- [16] Rossier, Michel F, "T channels and steroid biosynthesis: in search of a link with mitochondria," *Cell calcium*, vol. 40, no. 2, pp. 155--164, 2006.
- [17] Tamamoğullari, Necdet and Siliğ, Yavuz and İçağasioğlu, Serhat and Atalay, Atila, "Carnitine deficiency in diabetes mellitus complications," *Journal of diabetes and its complications*, vol. 13, no. 5-6, pp. 251--253, 1999.
- [18] Ruggenti, Piero and Cattaneo, Dario and Loriga, Giacomina and Ledda, Franca and Motterlini, Nicola and Gherardi, Giulia and Orisio, Silvia and Remuzzi, Giuseppe, "Ameliorating hypertension and insulin resistance in subjects at increased cardiovascular risk: effects of acetyl-L-carnitine therapy," *Hypertension*, vol. 54, no. 3, pp. 567--574, 2009.
- [19] Ruggenti, Piero and Van Der Meer, Irene M and Remuzzi, Giuseppe, "Oral acetyl-L-carnitine therapy and insulin resistance," *Hypertension*, vol. 55, no. 6, pp. e26--e26, 2010.
- [20] Mai, Manuel and Tonjes, Anke and Kovacs, Peter and Stumvoll, Michael and Fiedler, Georg Martin and Leichter, Alexander Benedikt, "Serum levels of acylcarnitines are altered in prediabetic conditions," *PLoS one*, vol. 8, no. 12, p. e82459, 2013.
- [21] Okun, Jürgen G and Kölker, Stefan and Schulze, Andreas and Kohlmüller, Dirk and Olgemöller, Katharina and Lindner, Martin and Hoffmann, Georg F and Wanders, Ronald JA and Mayatepek, Ertan, "A method for quantitative acylcarnitine profiling in human skin fibroblasts using unlabelled palmitic acid: diagnosis of fatty acid oxidation disorders and differentiation between biochemical phenotypes of MCAD deficiency," *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids*, vol. 1584, no. 2-3, pp. 91--98, 2002.
- [22] "uniprot," "MT-RNR1 - Mitochondrial-derived peptide MOTS-c - Homo sapiens (Human) - MT-RNR1 gene & protein"., [Online]. Available: <https://www.uniprot.org/uniprot/A0A0C5B5G6>. [Accessed 01 02 2021].
- [23] Zhang, Jie and Light, Alan R and Hoppel, Charles L and Campbell, Caitlin and Chandler, Carol J and Burnett, Dustin J and Souza, Elaine C and Casazza, Gretchen A and Huguen, Ronald W and Keim, Nancy L and others, "Acylcarnitines as markers of exercise-associated fuel partitioning, xenometabolism, and potential signals to muscle afferent neurons," *Experimental physiology*, vol. 102, no. 1, pp. 48--69, 2017.
- [24] Meienberg, Fabian and Loher, Hannah and Bucher, Julie and Jenni, Stefan and Krüsi, Marion and Kreis, Roland and Boesch, Chris and Betz, Matthias Johannes and Christ, Emanuel, "The effect of exercise on intramyocellular acetylcarnitine (AcCtn) concentration in adult growth hormone deficiency (GHD)," *Scientific reports*, vol. 9, no. 1, pp. 1--8, 2019.
- [25] Farooqui, AA and Horrocks, LA, "Metabolic and functional aspects of neural membrane phospholipids," *Phospholipids in the nervous system*, vol. 2, pp. 341--348, 1985.
- [26] Farooqui, Akhlaq A and Horrocks, Lloyd A and Farooqui, Tahira, "Glycerophospholipids in brain: their metabolism, incorporation into membranes, functions, and involvement in neurological disorders," *Chemistry and physics of lipids*, vol. 106, no. 1, pp. 1--29, 2000.

- [27] Whiley, Luke and Sen, Arundhuti and Heaton, James and Proitsi, Petroula and Garcia-Gomez, Diego and Leung, Rufina and Smith, Norman and Thambisetty, Madhav and Kloszewska, Iwona and Mecocci, Patrizia and others, "Evidence of altered phosphatidylcholine metabolism in Alzheimer's disease," *Neurobiology of aging*, vol. 35, no. 2, pp. 271--278, 2014.
- [28] Farmer, Kyle and Smith, Catherine A and Hayley, Shawn and Smith, Jeffrey, "Major alterations of phosphatidylcholine and lysophosphatidylcholine lipids in the substantia nigra using an early stage model of Parkinson's disease," *International journal of molecular sciences*, vol. 16, no. 8, pp. 18865--18877, 2015.
- [29] Ylilauri, Maija and Voutilainen, Sari and Lönnroos, Eija and Virtanen, Heli EK and Tuomainen, Tomi-Pekka and Salonen, Jukka T and Virtanen, Jyrki K, "Associations of dietary choline intake with risk of incident dementia and with cognitive performance: the Kuopio Ischaemic Heart Disease Risk Factor Study," *The American journal of clinical nutrition*, vol. 110, no. 6, pp. 1416--1423, 2019.
- [30] Growdon, John H, "Use of phosphatidylcholine in brain diseases: An overview," *Lecithin*, pp. 121--136, 1987.
- [31] Zhou, Miao-miao and Xue, Yong and Sun, Shu-hong and Wen, Min and Li, Zhao-jie and Xu, Jie and Wang, Jing-feng and Yanagita, Teruyoshi and Wang, Yu-ming and Xue, Chang-hu, "Effects of different fatty acids composition of phosphatidylcholine on brain function of dementia mice induced by scopolamine," *Lipids in health and disease*, vol. 15, no. 1, pp. 1--10, 2016.
- [32] Thorburn, David R and Rahman, Joyeeta and Rahman, Shamima, "Mitochondrial DNA-associated Leigh syndrome and NARP," 2017.
- [33] La Morgia, Chiara and Caporali, Leonardo and Gandini, Francesca and Olivieri, Anna and Toni, Francesco and Nasseti, Stefania and Brunetto, Daniela and Stipa, Carlotta and Scaduto, Cristina and Parmeggiani, Antonia and others, "Association of the mtDNA m. 4171C>A/MT-ND1 mutation with both optic neuropathy and bilateral brainstem lesions," *BMC neurology*, vol. 14, no. 1, pp. 1--5, 2014.
- [34] Blanch, Marta and Mosquera, Jose Luis and Ansoleaga, Belén and Ferrer, Isidre and Barrachina, Marta, "Altered mitochondrial DNA methylation pattern in Alzheimer disease--related pathology and in Parkinson disease," *The American journal of pathology*, vol. 186, no. 2, pp. 385--397, 2016.
- [35] van der Veen, Jelske N and Kennelly, John P and Wan, Sereana and Vance, Jean E and Vance, Dennis E and Jacobs, Rene L, "The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1859, no. 9, pp. 1558--1572, 2017.
- [36] Wang-Sattler, Rui and Yu, Zhonghao and Herder, Christian and Messias, Ana C and Floegel, Anna and He, Ying and Heim, Katharina and Campillos, Monica and Holzapfel, Christina and Thorand, Barbara and others, "Novel biomarkers for pre-diabetes identified by metabolomics," *Molecular systems biology*, vol. 9, no. 1, p. 615, 2012.
- [37] Trajkovic-Bodennec, Selena and Bodennec, Jacques and Futerman, Anthony H, "Phosphatidylcholine metabolism is altered in a monocyte-derived macrophage model of Gaucher disease but not in lymphocytes," *Blood Cells, Molecules, and Diseases*, vol. 33, no. 1, pp. 77--82, 2004.
- [38] Pietiläinen, Kirsi H and Sysi-Aho, Marko and Rissanen, Aila and Seppänen-Laakso, Tuulikki and Yki-Järvinen, Hannele and Kaprio, Jaakko and Oresic, Matej, "Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects--a monozygotic twin study," *PloS one*, vol. 2, no. 2, p. e218, 2007.



- [39] Kim, Ji Young and Park, Ju Yeon and Kim, Oh Yoen and Ham, Bo Mi and Kim, Hyun-Jin and Kwon, Dae Young and Jang, Yangsoo and Lee, Jong Ho, "Metabolic profiling of plasma in overweight/obese and lean men using ultra performance liquid chromatography and Q-TOF mass spectrometry (UPLC- Q-TOF MS)," *Journal of proteome research*, vol. 9, no. 9, pp. 4368--4375, 2010.
- [40] Wahl, Simone and Yu, Zhonghao and Kleber, Michaela and Singmann, Paula and Holzapfel, Christina and He, Ying and Mittelstrass, Kirstin and Polonikov, Alexey and Prehn, Cornelia and Römisch-Margl, Werner and others, "Childhood obesity is associated with changes in the serum metabolite profile," *Obesity facts*, vol. 5, no. 5, pp. 660--670, 2012.
- [41] Boone, Sebastiaan and Mook-Kanamori, Dennis and Rosendaal, Frits and den Heijer, Martin and Lamb, Hildo and de Roos, Albert and le Cessie, Saskia and van Dijk, Ko Willems and de Mutsert, Renee, "Metabolomics: a search for biomarkers of visceral fat and liver fat content," *Metabolomics*, vol. 15, no. 10, pp. 1--12, 2019.
- [42] Jourdan, Carolin and Petersen, Ann-Kristin and Gieger, Christian and Döring, Angela and Illig, Thomas and Wang-Sattler, Rui and Meisinger, Christa and Peters, Annette and Adamski, Jerzy and Prehn, Cornelia and others, "Body fat free mass is associated with the serum metabolite profile in a population-based study," *PloS one*, vol. 6, no. 7, p. e40009, 2012.
- [43] Wang, Bi-Dar and Rennert, Owen M and Su, Yan A, "Mitochondrial oxidative phosphorylation, obesity and diabetes," *Cells Science Reviews*, vol. 4, p. 4, 2008.
- [44] Takamura, Toshinari and Misu, Hirofumi and Matsuzawa-Nagata, Naoto and Sakurai, Masaru and Ota, Tsuguhito and Shimizu, Akiko and Kurita, Seiichiro and Takeshita, Yumie and Ando, Hitoshi and Honda, Masao and others, "Obesity upregulates genes involved in oxidative phosphorylation in livers of diabetic patients," *Obesity*, vol. 16, no. 12, pp. 2601--2609, 2008.
- [45] Hunter, Caroline A and Kartal, Funda and Koc, Zeynep C and Murphy, Tamara and Kim, Jung Han and Denvir, James and Koc, Emine C, "Mitochondrial oxidative phosphorylation is impaired in TALLYHO mice, a new obesity and type 2 diabetes animal model," *The international journal of biochemistry & cell biology*, vol. 16, p. 105616, 2019.
- [46] Dashti, Mohammed and Alsaleh, Hussain and Rodriguez-Flores, Juan L and Eaaswarkhanth, Muthukrishnan and Al-Mulla, Fahd and Thanaraj, Thangavel Alphonse, "Mitochondrial haplogroup J associated with higher risk of obesity in the Qatari population," *Scientific Reports*, vol. 11, no. 1, pp. 1--10, 2021.
- [47] Perks, Kara L and Ferreira, Nicola and Richman, Tara R and Ermer, Judith A and Kuznetsova, Irina and Shearwood, Anne-Marie J and Lee, Richard G and Viola, Helena M and Johnstone, Victoria PA and Matthews, Vance and others, "Adult-onset obesity is triggered by impaired mitochondrial gene expression," *Science advances*, vol. 3, no. 8, p. e1700677, 2017.
- [48] Wichmann, H-E and Gieger, Christian and Illig, Thomas and null for the MONICA/KORA Study Group and others, "KORA-gen-resource for population genetics, controls and a broad spectrum of disease phenotypes," *Das Gesundheitswesen*, vol. 67, no. S 01, pp. 26--30, 2005.
- [49] Holle, Rolf and Happich, Michael and Löwel, Hannelore and Wichmann, Heinz-Erich and null for the MONICA/KORA Study Group and others, "KORA-a research platform for population based health research," *Das Gesundheitswesen*, vol. 67, no. S 01, pp. 19--25, 2005.

- [50] Jia, Haiying and Guo, Yunfei and Zhao, Weiwei and Wang, Kai},, "Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer," *Scientific reports*, vol. 4, no. 1, pp. 1--8, 2014.
- [51] Marioni, John C and Mason, Christopher E and Mane, Shrikant M and Stephens, Matthew and Gilad, Yoav, "RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays," *Genome research*, vol. 18, no. 9, pp. 1509--1517, 2008.
- [52] [Online]. Available: <https://console.cloud.google.com/storage/browser/gatk-best-practices/mitochondria-pipeline>. [Accessed 19 06 2021].
- [53] Weissensteiner, Hansi and Forer, Lukas and Fuchsberger, Christian and Schöpf, Bernd and Kloss-Brandstätter, Anita and Specht, Günther and Kronenberg, Florian and Schönherr, Sebastian, "mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud," *Nucleic acids research*, vol. 44, no. W1, pp. W64--W69, 2016.
- [54] Suhre, Karsten and Shin, So-Youn and Petersen, Ann-Kristin and Mohnhey, Robert P and Meredith, David and Wägele, Brigitte and Altmaier, Elisabeth and Deloukas, Panos and Erdmann, Jeanette and Grundberg, Elin and others, "Human metabolic individuality in biomedical and pharmaceutical research," *Nature*, vol. 477, no. 7362, pp. 54--60, 2011.
- [55] Altmaier, Elisabeth and Ramsay, Steven L and Graber, Armin and Mewes, Hans-Werner and Weinberger, Klaus M and Suhre, Karsten, "Bioinformatics analysis of targeted metabolomics—uncovering old and new tales of diabetic mice under medication," *Endocrinology*, vol. 149, no. 7, pp. 3478--3489, 2008.
- [56] J. M. Cheverud, "A simple correction for multiple comparisons in interval mapping genome scans," *Heredity*, vol. 87, pp. 52--58, 2001.
- [57] Nyholt, Dale R, "A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other," *The American Journal of Human Genetics*, vol. 74, no. 4, pp. 765--769, 2004.
- [58] Li, J and Ji, L, "Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix," *Heredity*, vol. 95, p. 221, 2005.
- [59] Team, R Core and others, "R: A language and environment for statistical computing," 2013.

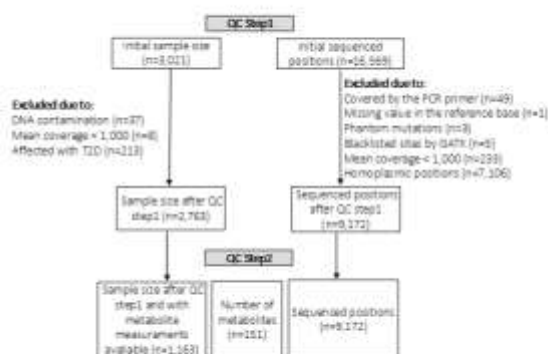


Figure 1: Schema of the quality control performed in the datasets.

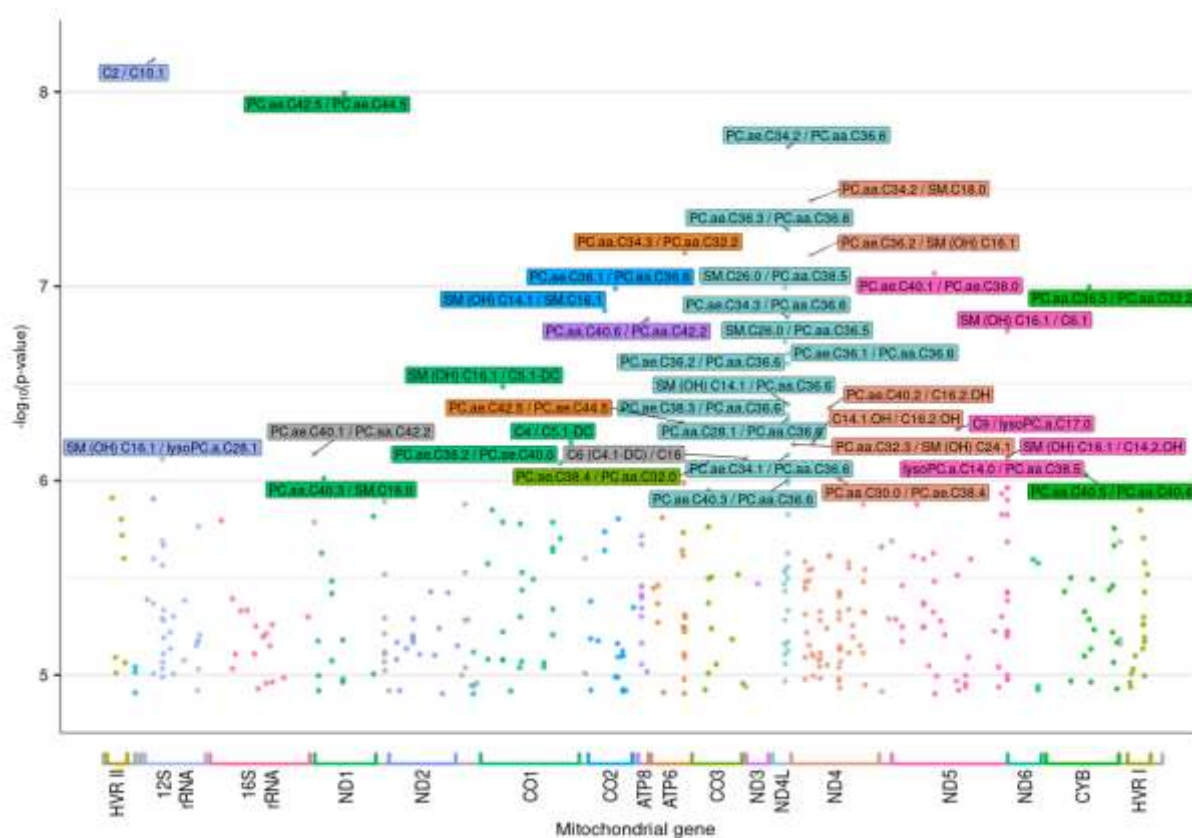


Figure 2: **Mitochondrial genome-wide P values for metabolite ratios.** Legend: On the y axis, p-values transformed into the negative of the base 10 logarithm,  $-\log_{10}(\text{p-value})$ , are shown. The x-axis represents the mitochondrial genome, displaying the position and relative size of each of the 13 major mitochondrial genes, 12S and 16S rRNAs, hypervariable region 1 (HVR I), hypervariable region 2 (HVR II) as well as the position of the 22 tRNAs (gray colour).

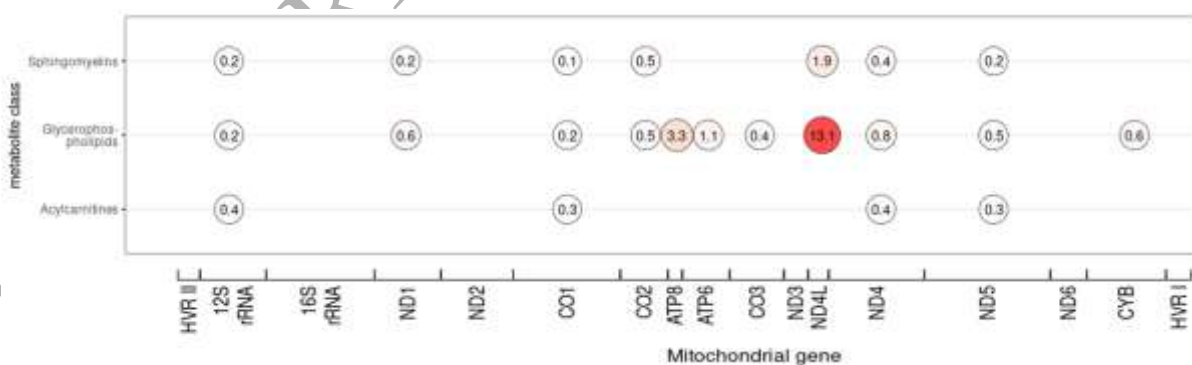


Figure 3: **The distribution of significant metabolites for metabolite classes across mitochondrial genes.** Legend: Given a mitochondrial gene, a circle shows the percentage of metabolites involved in significant ratios with regard to their metabolite class. The 40 most significant mtSNVs are considered. Because the genes have different length, the number of mtSNVs tested in each gene are different. The percentage

has been calculated taking the number of significant metabolite ratios divided by the total number of mtSNVs tested for each gene.

**Table 1. Estimates of the model parameters for the 40 most significant mtSNVs**

MT- Gene	Positi on	mtSNV	Alle le	Funcio n	MA F	Metabolite ratio short_name	Bet a	p- value	Metabolite class
12SrR NA	715	-	G > A	-	-	C2 / C10:1	0.9 09	6.82*1 0 <sup>-09</sup>	Acylcarnitine / Acylcarnitine
ND1	3714	rs386828 920	A > G	synony mous	0.000 2	PC ae C42:5 / PC ae C44:5	3.6 31	1.02*1 0 <sup>-08</sup>	Glycerophospholipid / Glycerophospholipid
ND4L	1068 9	rs879102 108	G > A	missens e	0.001 39	PC ae C34:2 / PC aa C36:6	0.6 37	1.92*1 0 <sup>-08</sup>	Glycerophospholipid / Glycerophospholipid
ND4	1105 0	rs160322 3077	T > C	synony mous	0.000 2	PC aa C34:2 / SM:C18:0	2.2 26	3.62*1 0 <sup>-08</sup>	Glycerophospholipid / Sphingolipid
ND4L	1068 9	rs879102 108	G > A	missens e	0.001 39	PC ae C36:3 / PC aa C36:6	0.6 37	5.12*1 0 <sup>-08</sup>	Glycerophospholipid / Glycerophospholipid
ATP6	9071	rs160322 2032	C > T	missens e	0.000 7	PC aa C34:3 / PC aa C32:2	1.2 86	6.75*1 0 <sup>-08</sup>	Glycerophospholipid / Glycerophospholipid
ND4	1105 0	rs160322 3077	T > C	synony mous	0.000 2	PC ae C36:2 / SM (OH) C16:1	2.3 99	6.87*1 0 <sup>-08</sup>	Glycerophospholipid / Sphingolipid
ND5	1299 4	rs160322 3993	G > A	missens e	-	PC ae C40:1 / PC ae C38:0	1.2 55	8.57*1 0 <sup>-08</sup>	Glycerophospholipid / Glycerophospholipid
CYB	1542 8	rs160322 5270	G > A	missens e	-	PC aa C36:5 / PC aa C32:2	0.4 34	1.01*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ND4L	1064 5	-	T > C	-	-	SM:C26:0 / PC aa C38:5	0.6 84	1.01*1 0 <sup>-07</sup>	Sphingolipid / Glycerophospholipid
CO2	7976	rs377368 526	G > A	missens e	-	PC ae C36:1 / PC aa C36:6	0.8 39	1.03*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
CO2	7809	-	T > C	-	-	SM (OH) C14:1 / SM:C16:1	1.1 69	1.33*1 0 <sup>-07</sup>	Sphingolipid / Sphingolipid
ND4L	1068 9	rs879102 108	G > A	missens e	0.001 39	PC ae C34:3 / PC aa C36:6	0.5 89	1.44*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ATP8	8477	rs160322 1517	T > C	missens e	0.000 7	PC aa C40:6 / PC aa C42:2	0.5 56	1.48*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ND5	1414 1	-	T > C	-	-	SM (OH) C16:1 / C6:1	0.8 95	1.69*1 0 <sup>-07</sup>	Sphingolipid/ Acylcarnitine
ND4L	1064 5	-	T > C	-	-	SM:C26:0 / PC aa C36:5	0.4 78	1.93*1 0 <sup>-07</sup>	Sphingolipid / Glycerophospholipid
ND4L	1068 9	rs879102 108	G > A	missens e	0.001 39	PC ae C36:1 / PC aa C36:6	0.7 66	2.22*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ND4L	1068 9	rs879102 108	G > A	missens e	0.001 39	PC ae C36:2 / PC aa C36:6	0.6 37	2.49*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
CO1	6207	-	T > C	-	-	SM (OH) C16:1 / C5:1-DC	0.7 44	3.29*1 0 <sup>-07</sup>	Sphingolipid / Glycerophospholipid
ND4L	1068 9	rs879102 108	G > A	missens e	0.001 39	SM (OH) C14:1 / PC aa C36:6	0.6 2	4.04*1 0 <sup>-07</sup>	Sphingolipids / Glycerophospholipid
ND4	1134 2	-	A > G	-	-	PC ae C40:2 / C16:2- OH	0.8 16	4.22*1 0 <sup>-07</sup>	Glycerophospholipid / Acylcarnitine
ND4L	1068 9	rs879102 108	G > A	missens e	0.001 39	PC ae C38:3 / PC aa C36:6	0.6 88	4.58*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ND4L	1068 9	rs879102 108	G > A	missens e	0.001 39	PC aa C28:1 / PC aa C36:6	0.7 26	4.82*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ATP6	9031	rs155642 3594	C > T	synony mous	0.000 2	PC ae C42:5 / PC ae C44:5	3.2 1	5.03*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ND5	1335 6	rs160322 4159	T > C	synony mous	0.000 6	C9 / lysoPC a C17:0	0.8 53	5.43*1 0 <sup>-07</sup>	Acylcarnitine / Glycerophospholipid
ND4	1108 8	-	T > C	-	-	C14:1-OH / C16:2-OH	0.8 49	6.34*1 0 <sup>-07</sup>	Acylcarnitine / Acylcarnitine
CO1	7279	rs160322 0861	T > C	missens e	0.000 7	C4 / C5:1-DC	0.6 15	6.36*1 0 <sup>-07</sup>	Acylcarnitine / Glycerophospholipid
ND4	1077 5	rs879015 842	G > A	missens e	0.000 2	PC aa C32:3 / SM (OH) C24:1	1.4 85	6.51*1 0 <sup>-07</sup>	Glycerophospholipid / Sphingolipid
tRNA	3241	-	A > G	-	-	PC ae C40:1 / PC aa C42:2	0.9 66	7.30*1 0 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid

ND4L	10689	rs879102108	G > A	missense	0,00139	PC ae C34:1 / PC aa C36:6	0.694	7.37*10 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ND5	14141	-	T > C	-	-	SM (OH) C16:1 / C14:2-OH	0.794	7.60*10 <sup>-07</sup>	Sphingolipid / Acylcarnitine
tRNA	10031	rs200048690	T > C	-	0.00186	C6 (C4:1-DC) / C16	1.221	7.67*10 <sup>-07</sup>	Acylcarnitine / Acylcarnitine
12SrRNA	856	rs1603218502	A > G	-	0.0007	SM (OH) C16:1 / lysoPC a C28:1	0.873	7.76*10 <sup>-07</sup>	Sphingolipid / Glycerophospholipid
CO3	9441	-	C > T	-	-	PC ae C38:4 / PC aa C32:0	0.913	8.07*10 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
CO1	7115	-	C > T	-	-	PC ae C38:2 / PC ae C40:0	1.102	8.10*10 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ND5	12825	-	T > C	-	-	lysoPC a C14:0 / PC aa C38:5	0.762	8.95*10 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
CYB	15373	rs1556424578	A > G	synonymous	-	PC aa C40:5 / PC aa C40:4	2.546	9.36*10 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ND1	3392	-	G > A	-	-	PC aa C40:3 / SM:C16:0	1.075	9.72*10 <sup>-07</sup>	Glycerophospholipid / Sphingolipids
ND4	11493	-	G > A	-	-	PC aa C30:0 / PC ae C38:4	0.754	9.78*10 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid
ND4L	10689	rs879102108	G > A	missense	0.00139	PC ae C40:3 / PC aa C36:6	0.649	9.92*10 <sup>-07</sup>	Glycerophospholipid / Glycerophospholipid

**Legend:** Genomic position in base pairs (bp), alleles, rs\_number, point mutation, and MAF are based on the NCBI dbSNP GRCh38 human genome assembly (rCRS, GeneBank ID NC\_012920.1). Note: MAF refers to the frequency with which the second most common allele occurs as homoplasmy in the population used by NCBI (how many individuals have only the alternative allele). In the current analysis MAF was not used, we used the measure heteroplasmy using individual-level allele frequencies obtained from counts values as proposed by [4]. Alleles are given in terms of major > minor allele. Nominal p-values are provided for each Beta. mtSNV: mitochondrial single nucleotide variant; MT-Gene: mitochondrial gene.