Published in final edited form as:

Hum Genet. 2018 March ; 137(3): 203–213. doi:10.1007/s00439-018-1873-4.

Deep Sequencing of the Mitochondrial Genome Reveals Common Heteroplasmic Sites in NADH Dehydrogenase Genes

Chunyu Liu^{1,2,3,*}, Jessica L. Fetterman⁴, Poching Liu⁵, Yan Luo⁵, Martin G. Larson^{2,3}, Ramachandran S. Vasan^{2,4}, Jun Zhu⁶, and Daniel Levy^{1,2,*}

¹Population Sciences Branch, NHLBI/NHI, Bethesda, MD;

²Framingham Heart Study, Framingham, MA;

³Department of Biostatistics, Boston University;

⁴School of Medicine, Boston University, Boston, MA;

⁵DNA Sequencing and Genomics Core, NHLBI/NIH, Bethesda, MD;

⁶System Biology Center, NHLBI/NHI, Bethesda, MD

Abstract

Increasing evidence implicates mitochondrial dysfunction in aging and age-related conditions. But little is known about the molecular basis for this connection. A possible cause may be mutations in the mitochondrial DNA (mtDNA), which are often heteroplasmic – the joint presence of different alleles at a single locus in the same individual. However, the involvement of mtDNA heteroplasmy in aging and age-related conditions has not been investigated thoroughly. We deep-sequenced the complete mtDNA genomes of 356 Framingham Heart Study participants (52% women, mean age 43, mean coverage 4570-fold), identified 2880 unique mutations and comprehensively annotated them by MITOMAP and PolyPhen-2. We discovered 11 heteroplasmic “hot” spots [NADH dehydrogenase (*ND*) subunit 1, 4, 5 and 6 genes, $n=7$; cytochrome *c* oxidase I (*COI*), $n=2$; 16S rRNA, $n=1$; D-loop, $n=1$] for which the alternative-to-reference allele ratios significantly increased with advancing age (Bonferroni correction $p<0.001$). Four of these heteroplasmic mutations in *ND* and *COI* genes were predicted to be deleterious nonsynonymous mutations which may have direct impact on ATP production. We confirmed previous findings that healthy individuals carry many low-frequency heteroplasmy mutations with potentially deleterious effects. We hypothesize that the effect of a single deleterious heteroplasmy may be minimal due to a low mutant-to-wildtype allele ratio, whereas the aggregate effects of many deleterious mutations may cause changes in mitochondrial function and contribute to age-related diseases. The identification of age-related mtDNA mutations is an important step to understand the genetic architecture of age-related diseases and may uncover novel therapeutic targets for such diseases.

*Correspondence: liuc@bu.edu (C.L.), levyd@nhlbi.nih.gov (D.L.).

Introduction

In humans, mitochondria produce energy for most cellular activities through oxidative phosphorylation (OXPHOS) and play essential roles in regulating many cellular processes.^{1,2} Mitochondria contain their own DNA (mtDNA), a circular genome of approximately 16.6 kb that is exclusively inherited maternally. The mtDNA encodes 13 key OXPHOS complex proteins and 22 tRNA and 2 rRNA genes for mitochondrial protein translation.² Due to a limited repair capacity, mutations occur with greater frequency in the mtDNA compared to the nuclear genome (nDNA).³ Cells contain many mitochondria with each mitochondrion containing 5–10 copies of mtDNA. Therefore, mtDNA mutations are often heteroplasmic, a condition whereby different mtDNA variants co-exist within a cell or across cells.⁴

Mitochondrial functions gradually deteriorate during normal aging,⁵ and is considered one of the hallmarks of aging.⁶ Wallace reported the first evidence in humans that mutations in mtDNA may cause human multisystem disorders that parallel aging.¹ High levels of mtDNA mutations have been found in skeletal muscles of old monkeys,⁷ livers of old mice,⁸ and fibroblasts of older adults.⁹ Two similar mouse models with a homozygous knock-in mutation for the error-prone mtDNA polymerase gene (the *PoIG* mice) exhibited greatly increased accumulation of somatic mtDNA mutations, reduced lifespan, and premature aging phenotypes including reduced fertility, cardiomyopathy, and frailty.^{10,11} Several mtDNA mutations have been described to cause rare yet, severe diseases in human.¹² Accumulating evidence has also suggested that mtDNA mutations are involved in age-related common diseases such as cardiovascular disease and cancers.^{1,4,12–18} Replication of associations between homoplasmic mtDNA mutations and common diseases has proven difficult due to small sample sizes in previous studies or lack of replication samples.^{15–17} Establishing the role of heteroplasmic mtDNA variation in relation to complex phenotypes is still in its infancy and the genetic basis for the age-related decline in mitochondrial function remains unclear.^{19,20} With the advent of next generation sequencing and reduced costs of whole-genome sequencing, it is now possible to study a spectrum of rare/low-frequency and heteroplasmic mtDNA variants in large sample sizes.^{21–25}

A few studies recently analyzed mtDNA mutations with whole genome sequencing or mtDNA sequencing and reported that healthy individuals carry many heteroplasmic mutations^{21–25} that increase in prevalence with age²² and are significantly overrepresented at known disease-associated loci.²³ Most reported heteroplasmic mutations, however, are rare and potentially enriched with deleterious mutations.²³ The high coverage and low error rate in sequence analysis of mtDNA is required to accurately identify mtDNA heteroplasmy. Therefore, it is essential to validate and extend previous findings and to discover further evidence for the role of mtDNA variation in aging and health. To that end, we analyzed the complete mtDNA genomes of 356 Framingham Heart Study (FHS) individuals using a deep sequencing method, mitoRCA-seq,²⁶ which utilizes mtDNA-specific rolling circle amplification to enrich the full-length mtDNA in order to increase the coverage and decrease the contamination with nuclear-encoded mtDNA sequences. We aimed to identify and characterize a full spectrum of mtDNA point mutations and to extend findings of recent publications.^{21–25} In addition, we sought to gain new insights into the role of mtDNA heteroplasmy in aging and health.

Methods

Study population

The design and recruitment of FHS three-generation cohorts have been described previously.^{27–29} All study participants have provided written consent for genetic research and cell line creation. The design of this study is detailed in Supplemental Data and illustrated in Supplemental Figure 1. This study used all three generational cohorts to construct maternal lineages¹⁶. We selected 400 individuals from independent lineages for mtDNA sequencing. After quality control procedures, 356 individuals (Supplemental Table 1) remained for analysis of mtDNA mutations with cell line derived DNA.

mtDNA preparation and sequencing

The Offspring cohort participants had their blood drawn between 1998 to 2005 (around examination 7). The Third Generation cohort participants had their blood drawn at examination 1 (between 2002 and 2005). Similar conditions applied to Offspring and Third Generation cohort participants for blood draw, storage, cell line construction, and DNA extraction from cell lines.

The National Heart, Lung, and Blood Institute's DNA Sequencing and Genomics Core Facility performed mtDNA sequencing using the mitoRCA-seq approach.²⁶ This strategy employed a rolling circle amplification (RCA) method to enrich full length mtDNA molecules from genomic DNA for subsequent library construction. The main advantage of the mitoRCA-seq strategy is the enrichment of full length mtDNA, thus minimizing contamination from the nuclear-encoded mtDNA sequences.²⁶ In addition, the mtRCA-seq used a high-fidelity enzyme, Phi29 DNA polymerase, for initial mtDNA enrichment. This enzyme has a low error rate of 1 in 10^6 – 10^7 bases.³⁰ The final mtDNA libraries were sequenced using the Illumina HiSeq-2000 platform.

Quality control

We focused on point mutations in this study. After de-multiplexing, raw sequencing reads were quality filtered with mapping quality score ≥ 30 to minimize potential sequencing errors.²⁶ Then the raw reads were aligned to the human revised Cambridge Reference Sequence (rCRS)³¹ by Burrows-Wheeler Aligner (BWA) using default parameters.³² In parallel, we selected the read with the highest average quality score (defined as best-unique read) for raw reads sharing the same sequence. The resulting best-unique reads were used for downstream mutation calling. In addition, each mutation site was required to be supported by multiple best-unique reads to remove PCR amplification artifacts.²⁶

We applied several additional filters to the mtDNA sequencing data. In an individual, we set the read as missing if a site had <30 mapped reads. We calculated the sample coverage which was the mean coverage of all 16,569 mtDNA sites in an individual. We also calculated the locus coverage which was the mean coverage at a mtDNA site across all individuals. Using the rCRS sequence as the reference, we defined alternative allele frequency (AAF) at a given site, AAF_{site} , which measured the proportion of alleles that were different from the reference allele over all mapped reads in an individual. Unlike the next-generation

sequencing that provides great depth of coverage, earlier genotyping arrays were not designed to discover heteroplasmic mutations. A low (e.g., 15%) or high (e.g., 85%) level heteroplasmic mutation is most likely to be identified as a homoplasmic for the reference allele or the variant allele by genotyping arrays. To compare with mtDNA alleles from previous genotyping, we used a relaxed threshold for assigning homoplasmic and heteroplasmic alleles for QC purposes: a site was deemed to be homoplasmic for the reference allele if the $AAF_{site} \geq 25\%$ at the site; a site to be homoplasmic for the alternative allele if the $AAF_{site} \leq 75\%$; and a site was defined as heteroplasmic if AAF_{site} was between 25% and 75%. We compared sequencing-derived homoplasmic alleles to alleles called on two genotyping arrays. One array included 172 mtDNA variants from the Exome Chip which was designed to investigate rare/low-frequency variants in the protein-coding regions.³³ The other array included 40 mtDNA variants genotyped on a customized platform.¹⁶ Both the Exome Chip and 40-mtDNA arrays used DNA that was extracted from cell lines.

We excluded samples if 1) the sample mean coverage was less than 1000-fold (all mapped reads) or 100-fold (uniquely mapped reads), 2) samples displayed high alternative allele frequencies across many mtDNA loci, and 3) samples showed different homoplasmic alleles at multiple loci by comparing respective alleles called on array-based genotyping and derived on sequencing data.

Definition of homoplasmic and heteroplasmic mutations

To identify true heteroplasmic mutations, we used a more stringent cutoff to define mtDNA heteroplasmy with best unique reads ≥ 30 . We defined a site to be homoplasmic of the rCRS reference allele if its $AAF_{site} \leq 5\%$ (i.e. no mutation compared to rCRS reference allele) and a site to be homoplasmic for the alternative allele if its $AAF_{site} \geq 95\%$. A site was called as heteroplasmic if its AAF_{site} was between 5% and 95% and the heteroplasmic allele was observed in both forward and reverse strand sequence reads. For simplicity, we refer this cutoff as the “5_95% rule”. We counted the number of heteroplasmic and homoplasmic mutations across the entire mtDNA per individual. We calculated the minor allele frequency (MAF) and the frequency of individuals displaying heteroplasmy (hetF) at each of the 16,569 mtDNA positions.

Bioinformatics analysis

The MITOMAP database provides an extensive compilation of human mtDNA mutations. We used MITOMASTER, a powerful sequence analysis tool, within MITOMAP to analyze and annotate the identified single nucleotide mutations.³⁴ In brief, MITOMASTER aligns mutation nucleotides to the rCRS and GeneBank protein coding sequence data and reports information including mutation type (transition or transversion), gene locus, effect on amino acid change (synonymous, nonsynonymous or termination), any reported disease associated with the mutation. For nonsynonymous mutations, we employed PolyPhen-2 to predict if a variant has deleterious or damaging effects on protein function.³⁵ For all PolyPhen-2 functional predictions, the following UniProtKB accession numbers were used to identify the OXPHOS protein subunits: ND1, P03886; ND2, P03891; ND3, P03897; ND4, P03897; ND4L, P03901; ND5, P03915; ND6, P03923; CYB, P00156; COX1, P00395; COX2, P00403; COX3, P00414; ATP6, P00846; ATP8, P03928.

Association analysis between age and heteroplasmic mutations

We tested if heteroplasmic mutations were related to age in two ways. First, we tested the association between age and the number of heteroplasmic mutations across individuals as a function of age. We performed inverse normal transformation on the number of heteroplasmic mutations in linear regression. Second, we tested if the AAF_{site} as a function of age for common heteroplasmic mutations in >10% individuals. The inverse normal transformed AAF_{site} values were used as the dependent variables. In addition to testing the heteroplasmic mutations in relation to continuous age variable, we displayed the interquartile range of the number of heteroplasmic mutations or AAF_{site} values in three age groups, <40 yrs, 40–59 yrs, and ≥60 yrs.

Several selected mtDNA loci in whole blood derived DNA

Among these 356 individuals, 131 were recently whole-genome sequenced using the whole-blood derived DNA as part of the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. We compared several selected mtDNA loci between the cell line derived DNA and whole blood derived DNA samples in these 131 individuals. Sequencing and control of the mtDNA sequences in these 131 individuals were reported in details in the Supplemental materials. To ensure that not all identified heteroplasmic mutations were near the beginning or the end of sequencing reads, we used Integrative Genomics Viewer (IGV, <http://software.broadinstitute.org/software/igv/>) to examine several DNA samples from both DNA sources for each of these heteroplasmic mutations.

Results

High mtDNA sequencing coverage

After completing all QC procedures, 356 individuals remained eligible for analyses (see Supplemental Data for QC details). All 16,569 mtDNA positions had quality reads to define mutations. The total raw reads was 961,896,903, the mapped reads was 595,007,422 (i.e., the mapping percentage was 61.86%), and the uniquely mapped reads was 594,699,686. The mean coverage was 4570-fold using raw mapped reads and 261-fold using the best-unique reads across the mtDNA genome in 356 individual (Supplemental Figure 2). The AAF_{site} derived from all mapped reads was highly correlated (Pearson correlation=0.98) to that from the best-unique reads. Owing to this high correlation, we report mutations that were based on best-unique reads (n=30).

mtDNA mutations overrepresented by rare heteroplasmic mutations

Based on comparison to rCRS, we identified 2880 mutations with the 5–95% rule, including 1931 (67.0%) heteroplasmic sites, 477 (16.6%) homoplasmic sites, and 472 (16.4%) “mixed” sites that were homoplasmic in some individuals but heteroplasmic in other individuals (Figure 1a). Most (n=1415, 73%) heteroplasmic mutations were singletons, 437 (23%) were observed in 2–10 individuals, 48 (2%) were found in 11–36 individuals; and 31 (2%) were common mutations present in >36 individuals (i.e. 10% of study participants or $hetF > 10\%$) (Figure 1b). A total of 7587 heteroplasmic variants were present at 1931 heteroplasmic sites in 356 individuals. We further investigated the AAF_{site} levels for these

7587 heteroplasmic mutations. Most heteroplasmic mutations were low-level AAF_{site} mutations: 5351 (71%) with $5\% < AAF_{site} \leq 10\%$ and 1403 (19%) with $10\% < AAF_{site} \leq 25\%$. Only 570 (8%) were medium-level ($25\% < AAF_{site} \leq 75\%$) and 263 (4%) were high-level ($75\% < AAF_{site} \leq 95\%$) heteroplasmic mutations (Figure 1c). The singleton heteroplasmic mutations displayed a wide range of AAF_{site} (5–95%). In contrast, 30 common heteroplasmic mutations were low-level ($5\% < AAF_{site} \leq 25\%$) mutations except for T310C in the D-loop, which displayed a wide range of AAF_{site} (9–71%) (Figure 2 and Supplemental Figure 3).

Common heteroplasmic mutations in the NADH dehydrogenase genes and the D-loop

We further investigated the 31 common heteroplasmic mutations that were shared by more than 10% individuals. Eleven of them were located in NADH dehydrogenase (*ND*) genes, ten in the D-loop (9 in the hypervariable region II), five in cytochrome c oxidase genes ($n=5$), three in rRNAs and two in tRNA genes (Figure 2, Supplemental Table 2). The three most common heteroplasmic mutations in *ND* genes were T10941G (L61R in *ND4*), which was shared by >72% of individuals, A4256T (Q317L in *ND1*), which was shared by 62% of individuals, and T14132C (L599P in *ND5*), which was shared by 48% of individuals. Within the hypervariable region II of the D-loop, C309T and T310C in HVR2 were heteroplasmic in all individuals. Two additional loci in the same area, A302C and G316C, were heteroplasmic in >90% individuals. Two heteroplasmic mutations in rRNA genes were shared by >35% individuals. The heteroplasmic mutations in the cytochrome c oxidase genes and tRNA genes were shared by 10–20% of individuals (Supplemental Table 2).

We investigated the 31 common heteroplasmic mtDNA loci in sequences of 131 participants who also had whole blood derived DNA samples sequenced (mean coverage 2338-fold). The three loci in the D-loop region were heteroplasmic in 46.6% (A302C), 56.6% (G316C), and 100% (T310C) of the 131 individuals of whole blood derived DNA, nine loci were heteroplasmic in smaller numbers of individuals, but the remaining 19 loci were not variable in any of these 131 individuals (Supplemental Table 3). With IGV, we confirmed that these heteroplasmic mutations are likely reliable heteroplasmic mutations (Supplemental Figure 4). However, Caution should be exercised when interpreting their alternative allele frequencies because some alternative alleles were near the beginning or the end of sequencing reads (Supplemental Figure 5).

Heteroplasmic mutations enriched in nonsynonymous deleterious mutations

We comprehensively annotated all 2880 mutations by MITOMAP³⁴ and PolyPhen-2.³⁵ Of all mutations, 321 sites were in the D-loop, 1917 sites in the 13 protein-encoding genes, 212 in 22 tRNAs, 405 in two rRNA genes, and 25 in intergenic regions (Table 1 and Supplemental Table 4).

Of the 1917 mutations in protein-coding genes, 722 (38%) were synonymous mutations, 1137 (59%) were nonsynonymous amino acid substitution mutations, 53 (2.7%) caused protein truncation (that is, a mutation that changes an amino acid codon to a stop codon), and five (0.3%) caused extension (that is, a mutation changes a stop codon to an amino acid codon; Supplemental Table 5). Most of the 53 truncation sites were low-level heteroplasmic

mutations (AAF_{site} : mean $\pm$ SD =10.9 $\pm$ 8.7). Forty of these 53 mutations were singletons (i.e. observed in only one individual), and the rest were shared by two to ten individuals.

We predicted the effects of the 1137 amino acid substitution mutations³⁵: 389 (34%) were predicted to be benign (average prediction score=0.05), 133 (12%) alleles were predicted to be “possibly damaging” (average prediction score=0.77), 609 (53.7%) were predicted to be probably damaging (average prediction score=0.99), and six (0.3%) had unknown effects³⁵ (Supplemental Table 5). We counted the number of possibly or probably damaging mutations in heteroplasmic (681/1931), homoplasmic (36/472) and mixed (19/477) mutation sites. Heteroplasmic mutations harbored a much higher proportion of possibly or probably damaging mutations than homoplasmies (35% versus 8%, Fisher’s exact $p < 2.2 \times 10^{-16}$) or mixed mutation sites (35% versus 4%, Fisher’s exact $p < 2.2 \times 10^{-16}$) (Figure 3a).

Identified mutations enriched with known mtDNA-disease alleles

MITOMAP³⁴ annotation revealed that 3% of the 16,569 mtDNA sites have been reported in association with clinical conditions versus 195 of the 2880 mutations in this study (3% versus 7%, Fisher’s exact $p < 2.2 \times 10^{-16}$; Supplemental Table 6). Furthermore, we found that sites displaying both heteroplasmic and homoplasmic mutations were greatly overrepresented with disease-associated alleles (n=53, 11%) compared to sites displaying heteroplasmy (n=110, 6%, $p = 6.2 \times 10^{-5}$) or homoplasmy (n=32, 7%, $p = 0.05$; Figure 3b). We retrieved “conservation scores” that indicate the degree of conservation for genetic loci across species³⁴ for 2880 loci with an available “conservation score” (scores were not available for loci at D-loop). The average score was 64 for the 195 loci with known clinical conditions, which was lower than the average score of 71 for all identified loci (t-test $p = 0.008$), indicating that the loci with known disease conditions are more mutable.

Imbalance transition to transversion ratio across mtDNA

The transition to transversion (ti/tv) ratio for these 2880 mutations was approximately 10. Of note, the ti/tv ratio was highly imbalanced across different regions: ti/tv=4 in the D-loop, ti/tv=12 in protein-coding genes, ti/tv=23 in rRNA genes, ti/tv=13 in tRNA genes, and ti/tv=1.5 in the intergenic regions.

Deleterious nonsynonymous heteroplasmic variation increases with aging

For common heteroplasmic mutations that were shared in >10% of individuals (Supplemental Table 2), we related alternative allele frequencies (i.e. inverse transformed AAF_{site} levels) to age at the time of blood draw. Age was significantly associated with alternative allele frequencies at 11 heteroplasmic loci, including seven sites at *ND* genes, two sites at *COI* genes, one at the D-loop, and one at the 16S rRNA gene ($p < 0.05/31 \sim 0.001$, Table 2). The median and interquartile range of AAF_{site} also displayed increases for these 11 heteroplasmic mutations across age groups <40 yrs, 40–59 yrs, and 60 (Figure 4). Sex was not significantly associated with alternative allele frequencies of the 11 mtDNA loci (Supplemental materials). Four of the 11 mutations were nonsynonymous and were predicted to be “probably or possibly damaging” by PolyPhen-2: three at *ND* genes including T10941G ($p = 0.0002$, L61R in *ND4*), T11678C ($p = 5.2 \times 10^{-9}$, W307R in *ND4*), and T12506A ($p = 1.1 \times 10^{-5}$, L57Q in *ND5*); one at G7065A ($p = 0.0007$, A388T in *COI*) (Table 2).

All individuals in this study harbored heteroplasmic mutations with a median of 26 [inter-quartile range (IQR)=19, 34] (Supplemental Figure 6a). The maximum number of heteroplasmic mutations an individual carried was 105. The inverse-normal transformed value of the number of heteroplasmic mutations was significantly associated with age ($\beta=0.017$, $p=3.4\text{e-}5$) (Supplemental Figure 6b). On average, individuals >60 years of age harbored 36 heteroplasmic mutations and individuals of <40 years old had 26 mutations (Supplemental Figure 7).

Discussion

We used a robust mtDNA-sequencing method, mitoRCA-seq, which employs rolling circle amplification with mtDNA specific primers to enrich the full-length circular mtDNA and to minimize the contamination of nuclear-encoded mtDNA sequence.²⁶ This strategy substantially increased mtDNA sequencing specificity and coverage, and greatly improved the accuracy for detecting heteroplasmic mutations with low level alternative allele frequencies compared to several recent studies that analyzed mtDNA mutations using whole genome sequencing data or shallower mtDNA sequencing.^{21–25} With the mitoRCA-seq method, we acquired high mean coverage (4570-fold) across the entire mtDNA genome.

Our study identified several age-related heteroplasmic “hot” spots in mtDNA protein coding genes, including the *ND* (complex I) and *COI* (complex IV) genes. Strikingly, seven of 11 age-related heteroplasmic variants were in *ND* subunit genes. Complex I is the largest enzyme complex of the mitochondrial OXPHOS respiratory chain, transferring electrons to ubiquinone and generating a proton gradient across the mitochondrial inner membrane to drive ATP production.³⁶ Mutations in the mtDNA-encoded ND subunits lead to Complex I deficiencies, which have been found to cause the most common forms of OXPHOS dysfunctions and mitochondrial DNA diseases.^{37,38} We found two common deleterious nonsynonymous heteroplasmic mutations, T10941G and T11678C, that result in amino acid substitutions (L61R and W307R, respectively), replacing hydrophobic amino acids with the positively charged arginine in *ND4* (Figure 5). Another deleterious heteroplasmy, T12506A, results in L57Q, replacing a hydrophobic amino acid with the polar amino acid glutamine in *ND5*. These three heteroplasmic mutations are located within or close to transmembrane domains of subunits *ND4* (T10941G and T11678C) and *ND5* (T12506A). The mtDNA encoded subunits of Complex I are highly hydrophobic transmembrane proteins that form the proton pumping machinery.^{39–41} The three deleterious nonsynonymous mutations in *ND* genes all replace hydrophobic amino acids with hydrophilic ones. Hence, these mutations likely alter the proton pumping ability of a key enzyme that impacts proton motive force and ATP production efficiency.

The creation of lymphoblastoid cell lines has provided a renewable source of DNA for many genotyping and sequencing projects, including the 1000 Genomes Project. A previous study found that cell line derived DNA is an appropriate source for sequencing, while cell line transformation may introduce low level *de novo* mutations.⁴² To date, the impact of cell line transformation on heteroplasmic mtDNA mutations has not been well studied. Although three loci in D-loop were common heteroplasmic mutations in both cell line derived DNA and whole blood derived DNA samples, the rest of the common heteroplasmic loci identified

in cell line based DNA samples were not variable or only heteroplasmic in smaller number of 131 individuals with whole blood DNA samples. Lower sequencing coverage of whole blood derived DNA and enrichment of a single white blood cell type in cell line construction may be partly attributable to the observed differences. Nevertheless, our comparisons indicate that DNA sources have a big impact on heteroplasmic mtDNA mutations. Some caution should be exercised in generalizing our findings from cell line derived DNA to whole blood derived or other cell type derived DNA. Future systematic comparison of sequence variations for the entire mitochondrial genome in these 131 individuals with both types of DNA sources will shed light on the impact of cell line transformation on *de novo* mutations in mtDNA.

It is important to note that direct comparison of results between our study and several previous studies is impossible due to differences in DNA sources, sequencing coverage, allele frequency thresholds used to define mutations and heteroplasmy. For example, we identified several heteroplasmic mtDNA “hot” spots in the D-loop and in *ND* subunits, which were not reported in the study with much larger sample size but much lower sequencing coverage by Ding et al.²². Nevertheless, we confirmed several previous findings. First, we found that the transition to transversion (Ti/Tv) ratio of the identified mtDNA mutations was approximate 10, which is similar to that reported in Ding et al.²². The highly variable Ti/Tv ratios across different mtDNA regions, however, are likely to result from mtDNA evolution and natural selection.⁴³ Second, our investigation of the spectrum of mtDNA mutations confirmed that mtDNA mutations are overrepresented by rare heteroplasmic mutations across all regions of the mtDNA genome.^{21–25} Third, we confirmed that the identified heteroplasmic mutations are overrepresented for known disease-associated sites²³ and are greatly enriched in deleterious mutations.³⁵ Fourth, we observed that the number of heteroplasmic mutations per individual increased with advancing age, which is consistent with findings from Ding et al.²² and Kang et al.¹⁸ Based on the fact that healthy individuals carry many low-frequency heteroplasmic mutations with potential deleterious effects, we hypothesize that the effect of a single deleterious heteroplasmy may be minimal due to a low mutant to wild-type ratio and that aggregate effects of many deleterious mutations may result in disease.

The identified mutation sites were overrepresented with known disease-associated sites, which is consistent with Ye et al.²³ The known disease-associated sites showed significantly lower “conservation scores” than other identified loci, and therefore, tend to be more mutable. These findings demonstrate that people harboring homoplasmic or high-level heteroplasmic mtDNA disease-associated mutations may not develop mitochondrial diseases, which suggests that the mitochondrial genetic environment may play a key role compensating for mtDNA disease-associated mutations.^{44,45} The notion that mitochondrial function declines with age^{1,46} is supported by our observation that the number of heteroplasmic variants and their alternative allele frequencies increase with advancing age. The exact mechanism underlying the age-related decline in mitochondrial function, however, remains unclear. Accumulation of heteroplasmic mtDNA mutations as a function of age remains an important area of investigation.

Our study has several limitations. We applied a stringent threshold to define heteroplasmy and found 1931 heteroplasmic mutations with high-confidence, but on the other hand, we likely missed some low-level heteroplasmic mutations. We were unable to investigate mtDNA copy number due to the depletion of nuclear DNA to enrich full-length mtDNA molecules for sequencing specificity. We selected individuals from independent maternal lineages and therefore were not able to investigate germline mutations versus somatic mutations. Despite these limitations, we validated many previous finding and provided new evidence to support the notion that heteroplasmic mutations play important roles in age-related mitochondrial functional decline. Identification of age-related mitochondrial DNA mutations is one of the next steps to understand full genetic architecture of age-related diseases. Such studies may uncover novel therapeutic targets of age-related diseases, leading to strategies that slow down age-related changes and prevent multiple age-related diseases.

URL

MITOMAP: <http://www.mitomap.org/MITOMAP>

PolyPhen-2: <http://genetics.bwh.harvard.edu/pph2/>

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Reference

- Wallace DC A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Annu Rev Genet* 39, 359–407 (2005). [PubMed: 16285865]
- Voet DVJ & Pratt CW Fundamentals of Biochemistry. 2nd Edition John Wiley and Sons, Inc. pp. 547 ISBN 0471214957. (2006).
- Wallace DC Mitochondrial DNA sequence variation in human evolution and disease. *Proc Natl Acad Sci U S A* 91, 8739–46 (1994). [PubMed: 8090716]
- Tuppen HA, Blakely EL, Turnbull DM & Taylor RW Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta* 1797, 113–28 (2010). [PubMed: 19761752]
- Payne BA & Chinnery PF Mitochondrial dysfunction in aging: Much progress but many unresolved questions. *Biochim Biophys Acta* 1847, 1347–53 (2015). [PubMed: 26050973]
- Lopez-Otin C, Blasco MA, Partridge L, Serrano M & Kroemer G The hallmarks of aging. *Cell* 153, 1194–217 (2013). [PubMed: 23746838]
- Schwarze SR et al. High levels of mitochondrial DNA deletions in skeletal muscle of old rhesus monkeys. *Mech Ageing Dev* 83, 91–101 (1995). [PubMed: 8569289]
- Khaidakov M, Heflich RH, Manjanatha MG, Myers MB & Aidoo A Accumulation of point mutations in mitochondrial DNA of aging mice. *Mutat Res* 526, 1–7 (2003). [PubMed: 12714177]
- Michikawa Y, Mazzucchelli F, Bresolin N, Scarlato G & Attardi G Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. *Science* 286, 774–9 (1999). [PubMed: 10531063]
- Kujoth GC et al. Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science* 309, 481–4 (2005). [PubMed: 16020738]
- Trifunovic A et al. Premature ageing in mice expressing defective mitochondrial DNA polymerase. *Nature* 429, 417–23 (2004). [PubMed: 15164064]
- Taylor RW & Turnbull DM Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 6, 389–402 (2005). [PubMed: 15861210]

13. Seyfried TN Cancer as a mitochondrial metabolic disease. *Front Cell Dev Biol* 3, 43 (2015). [PubMed: 26217661]
14. Seyfried TN, Flores R, Poff AM, D'Agostino DP & Mukherjee P Metabolic therapy: a new paradigm for managing malignant brain cancer. *Cancer Lett* 356, 289–300 (2015). [PubMed: 25069036]
15. Saxena R et al. Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *Am J Hum Genet* 79, 54–61 (2006). [PubMed: 16773565]
16. Liu C et al. Association of genetic variation in the mitochondrial genome with blood pressure and metabolic traits. *Hypertension* 60, 949–56 (2012). [PubMed: 22949535]
17. Byrne EM et al. Family-based mitochondrial association study of traits related to type 2 diabetes and the metabolic syndrome in adolescents. *Diabetologia* 52, 2359–68 (2009). [PubMed: 19760390]
18. Kang E et al. Age-Related Accumulation of Somatic Mitochondrial DNA Mutations in Adult-Derived Human iPSCs. *Cell Stem Cell* 18, 625–36 (2016). [PubMed: 27151456]
19. Stewart JB & Chinnery PF The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet* 16, 530–42 (2015). [PubMed: 26281784]
20. Stefano GB & Kream RM Mitochondrial DNA heteroplasmy in human health and disease. *Biomed Rep* 4, 259–262 (2016). [PubMed: 26998260]
21. He Y et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* 464, 610–4 (2010). [PubMed: 20200521]
22. Ding J et al. Assessing Mitochondrial DNA Variation and Copy Number in Lymphocytes of ~2,000 Sardinians Using Tailored Sequencing Analysis Tools. *PLoS Genet* 11, e1005306 (2015). [PubMed: 26172475]
23. Ye K, Lu J, Ma F, Keinan A & Gu Z Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc Natl Acad Sci U S A* 111, 10654–9 (2014). [PubMed: 25002485]
24. Li M et al. Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet* 87, 237–49 (2010). [PubMed: 20696290]
25. Payne BA et al. Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet* 22, 384–90 (2013). [PubMed: 23077218]
26. Ni T et al. MitoRCA-seq reveals unbalanced cytosine to thymine transition in Polg mutant mice. *Sci Rep* 5, 12049 (2015). [PubMed: 26212336]
27. Dawber TR, Meadors GF & Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health* 41, 279–81 (1951). [PubMed: 14819398]
28. Feinleib M, Kannel WB, Garrison RJ, McNamara PM & Castelli WP The Framingham Offspring Study. Design and preliminary data. *Prev Med* 4, 518–25 (1975). [PubMed: 1208363]
29. Splansky GL et al. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol* 165, 1328–35 (2007). [PubMed: 17372189]
30. Esteban JA, Salas M & Blanco L Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J Biol Chem* 268, 2719–26 (1993). [PubMed: 8428945]
31. Andrews RM et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23, 147 (1999). [PubMed: 10508508]
32. Li H & Durbin R Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–60 (2009). [PubMed: 19451168]
33. Grove ML et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* 8, e68095 (2013). [PubMed: 23874508]
34. Lott MT et al. mtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr Protoc Bioinformatics* 44, 1 23 1–26 (2013). [PubMed: 25489354]
35. Adzhubei I, Jordan DM & Sunyaev SR Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet Chapter 7, Unit7* 20 (2013).
36. Berg JM, Tymoczko JL & Stryer L *Biochemistry*. 5th edition., (2002).

37. Loeffen JL et al. Isolated complex I deficiency in children: clinical, biochemical and genetic aspects. *Hum Mutat* 15, 123–34 (2000). [PubMed: 10649489]
38. Scaglia F et al. Clinical spectrum, morbidity, and mortality in 113 pediatric patients with mitochondrial disease. *Pediatrics* 114, 925–31 (2004). [PubMed: 15466086]
39. Potluri P, Yadava N & Scheffler IE The role of the ESSS protein in the assembly of a functional and stable mammalian mitochondrial complex I (NADH-ubiquinone oxidoreductase). *Eur J Biochem* 271, 3265–73 (2004). [PubMed: 15265046]
40. Zickermann V et al. Architecture of complex I and its implications for electron transfer and proton pumping. *Biochim Biophys Acta* 1787, 574–83 (2009). [PubMed: 19366614]
41. Wirth C, Brandt U, Hunte C & Zickermann V Structure and function of mitochondrial complex I. *Biochim Biophys Acta* 1857, 902–14 (2016). [PubMed: 26921811]
42. Londin ER et al. Whole-exome sequencing of DNA from peripheral blood mononuclear cells (PBMC) and EBV-transformed lymphocytes from the same donor. *BMC Genomics* 12, 464 (2011). [PubMed: 21943378]
43. Duchene S, Ho SY & Holmes EC Declining transition/transversion ratios through time reveal limitations to the accuracy of nucleotide substitution models. *BMC Evol Biol* 15, 36 (2015). [PubMed: 25886870]
44. Gaweda-Walerych K & Zekanowski C The impact of mitochondrial DNA and nuclear genes related to mitochondrial functioning on the risk of Parkinson's disease. *Curr Genomics* 14, 543–59 (2013). [PubMed: 24532986]
45. Kern AD & Kondrashov FA Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. *Nat Genet* 36, 1207–12 (2004). [PubMed: 15502829]
46. Wallace DC Mitochondrial DNA mutations in disease and aging. *Environ Mol Mutagen* 51, 440–50 (2010). [PubMed: 20544884]

Figure 1. Mitochondrial DNA Mutations Overrepresented by Rare Heteroplasmic Mutations. a. Most identified mutations were heteroplasmic mutations; b. Most heteroplasmic mutations were singleton mutations; c. Most heteroplasmic mutations displayed low-level alternative allele frequencies.

Figure 2.
The Position, Alternative Allele Frequency, and Sample Frequency of the 1931 Mitochondrial DNA Heteroplasmic mutations.

Figure 3. Mitochondrial DNA Heteroplasmic mutations Enriched with Deleterious Nonsynonymous Mutations and Known Mitochondrial Disease Loci. a. Heteroplasmic mutations enriched with deleterious nonsynonymous mutations; b. Identified mutations enriched with known mitochondrial diseases.

Figure 4.

The Alternative Allele Frequencies of Eleven Mitochondrial DNA Heteroplasmic Mutations Significantly Associated with Advancing Age. The eleven loci included seven sites at NADH dehydrogenase genes, two sites at cytochrome oxidase I genes, one at the D-loop, and one at the 16s rRNA gene.

Figure 5.
Common Heteroplasmic Mutations Located within or Close to the Transmembrane Domain
of the NADH Dehydrogenase Subunits

Table 1.

Distribution of the 2880 Mitochondrial DNA Mutations

mtDNA region	Number of mutations (n=2880)
D-loop	319
13 Protein coding genes	1917
2 rRNA genes	406
22 tRNA genes	213
Intergenic	25

We deep sequenced 356 Framingham Heart Study individuals and identified 2880 mutations

Table 2.

The Alternative Allele Frequencies of Common Heteroplasmic mutations Significantly Associated with Aging

mtDNA position	Ref/Alt	HetF	Beta	SE	p	Locus	Effect ^a
469	C/T	0.12	0.020	0.004	2.2E-06	<i>D-loop</i>	Non-coding
2378	C/T	0.37	0.022	0.004	1.2E-07	<i>16S</i>	rRNA
3543	C/T	0.36	0.030	0.004	3.7E-13	<i>ND1</i>	SYN
6338	A/T	0.21	0.026	0.004	3.5E-10	<i>COI</i>	SYN
7065	G/A	0.21	0.014	0.004	6.5E-04	<i>COI</i>	A388T ^b
10941	T/G	0.73	0.015	0.004	2.4E-04	<i>ND4</i>	L61R ^b
11272	A/T	0.17	0.023	0.004	3.7E-08	<i>ND4</i>	SYN
11678	T/C	0.26	0.024	0.004	5.2E-09	<i>ND4</i>	W307R ^b
12506	T/A	0.13	0.018	0.004	1.1E-05	<i>ND5</i>	L57Q ^c
13455	C/T	0.47	0.031	0.004	2.5E-14	<i>ND5</i>	SYN
14536	A/G	0.11	0.022	0.004	1.1E-07	<i>ND6</i>	SYN

The alternative allele frequency of common heteroplasmic mutations was transformed by rank-based inverse normal transformation and used as dependent variable.

^aFunction or amino acid change caused by a mutation;

^aProbably damaging (average prediction score =0.99) nonsynonymous mutation predicted by PolyPhen-2;

^bPossibly damaging (average prediction score =0.77) nonsynonymous mutation predicted by PolyPhen-2; SYN, synonymous mutation. Ref/Alt, the revised Cambridge Reference Sequence allele and the observed alternative allele; HetF, the frequency of the heteroplasmy in 356 individuals; *ND1/3/5/6*, NADH dehydrogenase subunit 1, 3, 5, and 6. *COI*, cytochrome oxidase I; Amino acids – A, alanine; T, threonine; L, leucine; R, arginine; W, tryptophan; Q, glutamine.