Extreme heterogeneity of human mitochondrial DNA from organelles to populations

James B. Stewart^{1,2} & Patrick F. Chinnery^{3,4†}

¹Max Planck Institute for Biology of Ageing, Cologne, Germany.

²Wellcome Centre for Mitochondrial Research, Newcastle University Biosciences Institute, Faculty of Medical Sciences, Framlington Place, Newcastle upon Tyne, UK

³Department of Clinical Neuroscience, University of Cambridge, Cambridge, UK.

⁴Medical Research Council Mitochondrial Biology Unit, University of Cambridge, Cambridge, UK.

†e-mail: pfc25@cam.ac.uk

Abstract

Contrary to the long-held view that most humans harbour only identical mitochondrial genomes (mtDNA), deep re-sequencing has uncovered unanticipated extreme genetic variation. Most, if not all humans, contain multiple mtDNA genotypes (heteroplasmy), specific patterns of variants accumulate in different tissues over time (including cancers), and some are preferentially passed down or suppressed in the maternal germ line. These findings cast light on the origin and spread of mtDNA mutations at multiple scales, from the organelle to the human population, and challenge the conventional view that high percentage levels of a mutation are required before a new variant has functional consequences.

Introduction

Shortly after publication of the first draft sequence of the 16.5 kb human mitochondrial DNA (mtDNA)¹, it became clear that some humans harboured a mixed population of genomes – termed heteroplasmy². Using methods available at the time, heteroplasmy seemed to be rare. Primarily described in patients with severe maternally-inherited mitochondrial diseases, high percentage levels of the mutant form (typically >80%) were linked to a biochemical defect affecting oxidative phosphorylation that compromised the production of adenosine triphosphate (ATP). Subsequent epidemiological studies showed that these mutations included both large-scale deletions and single nucleotide variants (mtSNVs), and affected ~1 in 8,000 individuals³, with higher mutation loads linked to a more severe clinical phenotype. The wide-spread adoption of chain-terminating sequencing methods also identified heteroplasmy in healthy individuals, often involving homopolymeric length

variants in the non-coding sequence that were not obviously causing disease⁴. Although several orders of magnitude more common than the disease-causing mutations, heteroplasmy in protein-coding genes was only detected in ~1 in 25 humans using techniques available at the time⁵.

In less than a decade, our understanding has radically changed. Massively parallel deep re-sequencing has revealed an extraordinary level of mtDNA diversity at multiple levels, with polymorphisms present across the world population reflected within individuals at the cellular – and possibly organellar level. Here we review these recent advances in understanding which cast light on the origins of human mtDNA variation and its relevance for both rare and common human diseases, including cancer, neurodegenerative disorders and even the ageing process itself.

Mitochondrial function, origins and biogenesis

Mitochondria are ubiquitous intracellular organelles that contain ~10% of the entire cell proteome⁶ and enveloped by two membranes. Although often portrayed as discrete ‘bean-like’ structures, in reality they can form a fusing and budding network which changes in response to the energetic requirements of a cell⁷. Highly energy-dependent cells typically contain reticulate mitochondria, with different cellular regions hosting sub-populations servicing local energetic requirements. The outer mitochondrial membrane is largely porous, but the passage of small molecules across the inner membrane is highly regulated and involves over ~100 transporters that control the exchange between the inner matrix from the cytosol. This compartmentalization is central to mitochondrial function, enabling the four respiratory chain complexes to generate a membrane potential which is harnessed by complex V to synthesize ATP through oxidative phosphorylation. Mitochondria also act as hubs for many canonical intracellular signaling networks, and play a central role in intermediary metabolism, the synthesis of critical iron-sulfur cluster enzymes, cell calcium homeostasis, and in cell death through apoptosis^{8,9}.

The vast majority of the ~1,500 mammalian mitochondrial proteins are synthesized from nuclear genes on cytosolic ribosomes in close proximity, allowing transport across the inner membrane using importers that process and deliver the mature peptides to membranes, matrix or inter-membrane space¹⁰. However, 13 proteins essential for oxidative phosphorylation are synthesized within the matrix from the 16.5 kb mitochondrial genome (mtDNA, **Figure 1**). Intra-mitochondrial protein synthesis also requires 22 tRNA and 2 rRNAs encoded by mtDNA, and a complex machinery encoded by the nuclear genome, including translation initiation and elongation factors, key ribosomal components, molecular chaperones and enzyme assembly factors. Many of the nuclear-encoded mitochondrial proteins maintain the structural integrity of the organelle and the maintenance of mtDNA. This complex arrangement has evolved over ~2.5 billion years, following the endosymbiosis of a proteobacterial ancestor, which formed the first mitochondrion. Debate continues

over the precise ancestry of this bacterium¹¹, but most studies show evidence that it was an α -proteobacterium related to modern *Rickettsia*¹². Over time, mitochondria have become more complex and diverse across the animal and plant kingdoms, but in parallel, genes controlling several key functions have been passed to the host cell, driven by replacement of function by nuclear-encoded genes¹³ and the exchange of genes from the mtDNA into the cell nucleus. To this day, we still see the insertion of segments of mtDNA throughout the nuclear genome, so called nuclear-mitochondrial sequences (NuMTs) (**Figure 2, Box 1**)^{14,15}. The subtle but critical interplay between mitochondria and the rest of the cell ultimately reflects the indirect interaction of two genomes that are inherited completely independently, with nuclear genes following Mendel's laws of inheritance, and the mtDNA showing strict uniparental inheritance, which in mammals involves the maternal line.

Molecular origin of mitochondrial DNA mutations

Diploid human cells contain two copies of each autosomal gene, but many more copies of mtDNA (typically 1,000–10,000). Individual mtDNAs are packaged into replicating nucleoids containing an average of ~1.4 molecules that are tethered to the inner mitochondrial membrane¹⁶. MtDNA was thought to be devoid of protective histones and is close to the respiratory chain, which is a potent source of oxygen free radicals and reactive oxygen species (ROS). This finding led to the suggestion that ROS-mediated damage was a major source of mtDNA variation in humans. However, the recent characterization of different trinucleotide mutational signatures^{17,18} suggests that most age-related mtDNA mutations are more likely to be due to errors during mtDNA synthesis¹⁹.

Unlike the nuclear genome, mtDNA is replicated continuously, independent of the cell cycle (*relaxed replication*)²⁰, and has a half-life of ~7–10 days depending on the cell type²¹. MtDNA is replicated by a specific polymerase, pol γ , composed of a catalytic core with proofreading capacity encoded by the nuclear gene *POLG*, and, in mammals two accessory subunits encoded by *POLG2*, which enhance enzyme processivity²². Although pol γ has a high fidelity, the amount of mtDNA replication that is required over the life-time of a cell inevitably introduces base substitution errors (mtDNA single nucleotide variants, mtSNVs)²³. For example, a human post-mitotic neuron containing ~10,000 mtDNA living for 80 years will replace its entire mtDNA ~3,000 times without any parallel nuclear DNA replication. This unusual mode of DNA replication also contributes to the vulnerability of mtDNA²⁴. In quiescent or non-dividing cells, mtDNA replication begins with the leading strand emerging from an RNA primer in the non-coding region (NCR) which includes the D-(displacement)-loop containing the heavy strand origin of replication (O_H) (**Figure 1**). This proceeds ~2/3 of the way around the molecule, exposing single-stranded DNA, before triggering synthesis of the lagging strand from the origin of light strand replication (O_L). This 'strand asymmetric' replication of mtDNA plays a key role in the generation of mtDNA rearrangements through copy-choice recombination during

replication of the light strand²⁵. The extended time spent in a single-stranded state has long been argued to be the source of the observed strand-specific directional mutation bias observed for mtDNA²⁶, and as predicted, the inversion of the control region leads to an inversion of this mutational bias²⁷. However, other models for mtDNA replication exist²⁸, whereby replication is initiated at multiple points along its length (strand-coupled replication), with single stranded ‘bubbles’ replicating heavy and light strands almost simultaneously in both directions^{29,30}. By inference, this mechanism is less likely to cause mtDNA rearrangements or directional mutation biases. However, the existence of different replication modes could provide a selection-free model explaining, for instance, why deletions are less likely to occur in the tissues containing rapidly dividing cells²⁵.

Clonal expansion of mitochondrial DNA mutations

Ultimately, all mtDNA mutations are generated on a single molecule, so to reach detectable levels they must be propagated by one of two mechanisms (**Figure 3**). Individual mtDNA molecules can be ‘clonally expanded’ without cell division through the process of *relaxed replication*. Modelling has shown that, even if mutated and wild-type molecules undergo non-selective replication, this can lead to changes in the proportion of mutated molecules through random genetic drift³¹. However, there is also evidence that some mtDNA mutations undergo preferential replication, either because they copy faster, or because they become available to replicate more frequently. In mammals, this has been shown *in vitro* for mtDNA deletions³², explaining how they accumulate with age in post-mitotic tissues, including the brain and skeletal muscle. However, the same also seems to be the case for mtSNVs in the *Drosophila* D-loop, leading to the ‘selfish replication’ of haplotypes under certain conditions (see below)^{33,34}, or for mtDNA deletions transmitted by various *Caenorhabditis* species³⁵.

Changes in allele frequency can also occur when mtDNAs are partitioned during cell division (*vegetative segregation*), where the unequal distribution of different alleles leads to daughter cells with different heteroplasmy levels. Indirect evidence suggests that mitochondrial sub-compartments can be selectively partitioned into daughter cells^{36,37}, potentially accelerating changes in allele frequency. Together, the combination of relaxed replication and vegetative segregation can lead to changes in heteroplasmy level during life, explaining how mutation accumulate with age³⁸. However, to be visible at the population level, *de novo* mtDNA variants must be transmitted down the maternal line.

Germ line transmission of heteroplasmy

Major changes in mtDNA heteroplasmy levels were recognized 30 years ago in Holstein cows³⁹⁻⁴¹, where a new variant emerged and fixed within the maternal lineage over a few generations. This led to the ‘mitochondrial bottleneck’ hypothesis, whereby only a small proportion of mtDNAs are passed

from mother to offspring (**Figure 3**). The bottleneck causes a statistical sampling effect leading to different allele frequencies between siblings. Recent direct measurements have confirmed that there is a dramatic reduction in the cellular mtDNA content in early germ line precursors (primordial germ cells), which is conserved across chordata, including humans^{37,42-46}. Mathematical and statistical modelling has shown that the measured reduction of intracellular mtDNA content is sufficient to cause major shifts in heteroplasmy levels^{42,47,48}, although other mechanisms are likely to contribute, which could include the sub-compartmentalization of mtDNAs into organelles, larger nucleoids, or sampling through the replication of a subpopulation of molecules⁴⁹. At the cellular level, all these mechanisms can be explained by vegetative segregation and relaxed replication of mtDNA on different levels (molecular, organellar, cellular), albeit at an accelerated rate within the female germ line. This leads to the rapid fixation or loss of new alleles within a few generations of a maternal lineage, and provides an explanation for the apparent lack of high-level heteroplasmy in most humans (present at >10% levels). However, the application of deep mtDNA re-sequencing has shown that low-level mtDNA heteroplasmy is extremely common in humans, and at very low levels (0.5–1%), heteroplasmy seems to be a universal finding⁵⁰.

Universal mtDNA heteroplasmy

Deep resequencing allows the detection of variants present at low frequency, and has been used in several contexts, including the detection of rare bacterial species in a mixed microbiome, measuring viral loads, and detecting somatic mutations in the nuclear genome in cancers⁵¹. When applied to human mtDNA, high depth (>1,000-fold) re-sequencing has shown heteroplasmic variants scattered throughout the mitochondrial genome, particularly concentrated in the D-loop⁵²⁻⁵⁵. At ultra-high depth (>5,000-fold), heteroplasmic variants were detected in almost every human studied at allele frequencies between 0.5–1.5%⁵⁰. Higher sensitivity techniques, such as duplex sequencing, or simply sequencing the same sample twice, can improve the sensitivity and specificity of variant detection^{51,56,57}. This has allowed the nucleotide mutational signature of age-related somatic mtDNA variants to be defined in different human tissues¹⁹. Intriguingly, some regions of the genome seem to be devoid of even low-level heteroplasmies⁵⁸. This finding implies that specific variants in these ‘mutation deserts’ cannot be clonally propagated, perhaps because of their impact on mtDNA replication. Importantly, this result indicates that mtDNA mutations can have biological effects and consequences even when the heteroplasmy level is well below the typical high (~80%) threshold associated with the impaired respiratory chain function seen in mitochondrial diseases. Low-level heteroplasmies can also be inherited^{50,55}, and the allele frequency can change between mother and child due to the mtDNA genetic bottleneck⁵⁸. In rare instances, the shift in heteroplasmy can be dramatic, and if pathogenic, higher heteroplasmy levels can cause mitochondrial disease.

Heteroplasmic variants can also become fixed (homoplasmic) within the maternal lineage, and thus contribute to the background polymorphic landscape of mtDNA. However, although paradigm shifting, deep mtDNA sequence data must be interpreted cautiously (**Box 2**).

Population mtDNA variants

The human population has acquired new mtDNA variants from pre-history to the present. Being strictly maternally inherited, there has been negligible intermolecular recombination of mtDNA, so the addition of a new mutation creates a new branch on a complex maternal family tree spanning back over 200,000 years from a theoretical common founder, or 'mitochondrial Eve'⁵⁹. As humans migrated and populated the globe, new variants defined geographically-located 'macro-haplogroups' which differ between continents. Ongoing mutations after the continental arrival initially led to common 'haplogroups', typically present at >5% allele frequency in a given region. Population mtDNA sequencing efforts have defined sub-haplogroups, each with many branches, and we are still discovering new mtDNA variants which decorate the 'leaves' of the mtDNA phylogenetic tree⁶⁰. This presents a challenge to diagnostic labs investigating a new mitochondrial disease, as it can be difficult to separate novel pathogenic mutations from new polymorphisms, but is also remarkable from a genetic perspective. The small 16.5 kb human mitochondrial genome has been acquiring mutations for over 200,000 years, and close to 200,000 modern humans have been sequenced, yet only ~5% of the molecule harbours variants present at >5% population allele frequency^{61,62}. This observation suggests that selective forces have shaped human mtDNA evolution.

Evidence for selection

Migration, population bottlenecks and genetic drift have all contributed to the present distribution of mtDNA haplogroups, with significant differences arising through random or stochastic events. However, there is emerging evidence that the environment has contributed to the current mtDNA landscape, with purifying selection acting at different levels: individuals within populations and during transmission between generations. Until recently, most of the evidence showed selection against specific mtDNA variants, but support is emerging for positive selection in specific genomic regions across several species.

Selection on homoplasmic variants in human populations

Studying randomly ascertained individuals has consistently shown that the ratio of synonymous to non-synonymous variants is >1, implying purifying selection acting on functional variants at the population level⁶³. There also seems to be evidence of selection within large reconstructed phylogenetic trees. In one study, *de novo* non-synonymous and rRNA variants were more common

in the pedigrees than expected based on population data, and the rate and pattern of substitutions differed between the proximal (older) and distal (recent) branches of the tree. In keeping with this finding, the mtDNA substitution frequency calculated from genomes samples across several populations was much lower than the rate measured in the extensive pedigrees⁶⁴. Although one study showed an excess of non-silent variants in arctic populations, implying a role for mtDNA in cold adaptation⁶⁵, similar patterns were also seen in other populations⁶⁶, suggesting that climate was not the main driver, and simply that younger populations had a surplus of non-synonymous substitutions that had not yet been ‘cleaned up’ by the purifying selection evident at the global level.

Further evidence of positive selection emerged from studying patients admitted to the critical care unit (CCU) with severe infection leading to sepsis⁶⁷. Although mtDNA variants did not influence whether or not a person was admitted to the CCU, European mtDNA haplogroup H was associated with a ~2-fold greater chance of survival. Intriguingly, this difference emerged after the first 48 hours, implying that ‘mitochondrial fitness’ was the key to long-term survival. Haplogroup H is considered to be the most recent addition to European mtDNA but is also the most common (~40%), implying a selective pressure driving up the allele frequency. Different European haplogroups have been implicated in other studies^{68,69}, and with different variants in other populations^{68,70}, providing evidence that different haplogroups are associated with different levels of ‘fitness’ that become exposed at times of extreme external or environmental stress. Several examples have been presented to support this notion, including the m.3394T>C *MTND1* variant which is associated with a complex I enzyme defect and the mtDNA disease Leber hereditary optic neuropathy (LHON)⁷¹. m.3394T>C is rare at low elevations but reaches high (~25%) allele frequencies at high altitudes in Tibet on a different mtDNA background. As explained above, the difference in allele frequencies could have arisen by chance, but it is conceivable that the variant has undergone positive selection (or perhaps relaxed selection⁷²) at low oxygen levels. The population and pedigree findings are supported by experimental work, including human cell models and animals. Cytoplasmic hybrid (or cybrid) techniques allow the investigation of the effects of mtDNA genotypes on the same nuclear genetic background, and have revealed differences in protein synthesis, respiratory chain activity and ROS production linked to common mtDNA haplogroups^{73,74}.

Unfortunately, within-cell or within-individual dynamics are insufficient to describe the complexities of the dynamics of mitochondrial heteroplasmy. For instance, mtDNA deletion mutations in *C. briggsae* were demonstrated to have a transmission advantage, and as expected, led to mitochondrial dysfunction and negative selection when the deletion reached high percentage levels³⁵. However, when the number of animals per replicate in the experimental population was increased (therefore increasing the effective population size N_e and the efficiency of natural selection), the deletion frequency within the population was selected against. These dynamics were further modified

by the precise strain used in the experiment, also implying a mito-nuclear interaction as a component in the dynamics of the transmission of these otherwise ‘selfish’ mtDNA deletions³⁵.

Selection on heteroplasmic variants during transmission

The mtDNA genetic bottleneck model predicts the random segregation of mtDNA heteroplasmy during transmission, with heteroplasmy levels equally likely to increase or decrease between mother and offspring. Extensive animal studies support this mechanism, including key work in *Drosophila*, cows, mice and humans transmitting polymorphic variants. However, very high levels of some pathogenic mtDNA mutations were almost never seen, implying purifying selection during mother-to-offspring transmission. Given the challenges in studying cellular and molecular mechanisms in living humans, our understanding of germ line selection has been driven by animal studies, principally in *Drosophila* and mice.

In *Drosophila*, temperature-sensitive selection has been seen against a heteroplasmic thermo-sensitive *mtco1* variant within the germ line (*mtCO1^{ts}*)⁷⁵, which is fatal when homoplasmic. Several mechanisms have been proposed to explain these findings, including an allele-specific reduction in mtDNA replication during early oocyte development, and mitochondrially-targeted autophagy (mitophagy). In one study, damaging mtDNA variants were shown to stabilize PINK1 on the outer mitochondrial membrane. The PINK1 phosphorylated Larp which inhibited local protein synthesis, thus ‘starving’ mitochondria of essential nuclear-encoded replication factors, and preventing the replication of specific mtDNA alleles⁷⁶. Germ line mitophagy seems to be critically dependent on mitochondrial fission induced by a decrease in the mitochondrial fusion protein mitofusin (mfn)⁷⁷. The fragmentation is thought to decrease organellar ATP synthesis, and the activation of Atg1 (the fly homologue of mammalian ULK1) which regulates autophagy, and the mitophagy receptor BNIP3 (closely related to the mammalian orthologue BNIP3L). This is in keeping with earlier observations that the selective elimination of the *mtCO1^{ts}* variant does not require *Parkin*, unlike other forms of mitophagy in somatic tissues⁷⁸.

Intriguingly, there is also evidence of positive selection for specific mtDNA variants in *Drosophila*, even when they have detrimental effects on the cell or organism as a whole (so called selfish transmission)³³. Perhaps the best example is where mtDNA variants harmful to males are transmitted down the female line (the so-called ‘mothers curse’), proposed to occur in human pedigrees transmitting mtDNA mutations associated with LHON⁷⁹. There are also examples where ‘selfish’ mtDNA can affect male and female fitness, as seen in *Drosophila subobscura*⁸⁰, *C. briggsae*⁸¹, and *C. elegans*⁸², where large mtDNA deletions persist down multiple generations despite compromising oxidative phosphorylation, possibly by compensating through mitochondrial biogenesis or activation of the unfolded protein response⁸³.

Recent work mixing mtDNAs from different *Drosophila* strains has shown how different genomes compete when placed together³³. When mixed with mtDNA from a different strain, the *mtCO1^{ts}* variant outcompeted another allele harbouring an *ATP6[1]* variant, eventually leading to death of the fly population. Paradoxically, the *mtCO1^{ts}*-linked non-coding D-loop is much larger than the *ATP6[1]*-linked D-loop. This would usually be associated with a longer time to replicate the mtDNA molecule, allowing the *ATP6[1]*-haplotype to out-compete the *mtCO1^{ts}* haplotype. Further evidence for selfish propagation emerged when the mtDNA was exchanged between *D. melanogaster* and *D. yakuba*. Despite diverging 10 million years ago, completely exchanging the two mtDNAs had no effect. However, when the two mtDNAs were placed in competition, *D. melanogaster* always out-competed *D. yakuba*, even when some deleterious mutations were introduced into the *D. melanogaster* mitochondrial genome. Extending these experiments, it has been possible to construct a hierarchy of *Drosophila* mtDNAs, ranking the best and worst competitors, irrespective of their nuclear genetic background. It remains unclear how one genome out-competes the other, but the data point towards single-base variation in critical regions regulating mtDNA replication.

Given the technical challenges there is less evidence of selection in mammalian genomes, but the case is building. Strong purifying selection against coding-region variants was reported in mice, and several studies have shown selection against high levels of coding region and tRNA gene variants^{47,84-86}. The first studies mixing mouse mtDNA haplogroups used laboratory mouse strains with the BalbC or C57Bl/6NCrl nuclear genetic background, and heteroplasmic mtDNA including the wild-type for that particular strain alongside NZB mtDNA (which has ~90 nucleotide differences from the wild-type)^{87,88}. In early generations, large variations in heteroplasmy or the fixation of one allele were observed over a few generations, resembling the rapid segregation patterns seen recently in studies of mitochondrial replacement therapies in monkeys and human embryos^{89,90}. However, with successive backcrossing, the transmission of the two alleles followed a neutral model of segregation. Intriguingly, an unusual bias selected for the BalbC mtDNA in haematopoietic cells was later found to be regulated by two GTPase of the immunity-associated proteins (the endoplasmic reticulum associated Gimap3 and the lysosomal-associated Gimap5)⁹¹. Conversely, the NZB mtDNA was favoured in the liver, and was linked to undefined variations in a locus named Smdq1, which was identified by ablation of the segregation bias in F2 hybrids carrying the *Mus musculus castaneus* allele⁹². However, more recent work has revealed very different results.

Similar hybrids with NZB and 129S6 mtDNAs were generated that also showed rapid changes in heteroplasmy over just a few generations, with alleles being lost over a few generations. In this study, the heteroplasmy was unstable, and rapidly lost in the mouse lines⁹³. Mice with a mixed C57Bl/6N and either LE or HB mtDNAs showed a subtle haplogroup-specific bias in germ line transmission, also seen in different tissues during the lifetime⁹⁴. By contrast, in heteroplasmic mice

generated from C57BL/6 and NZB/OlaHsd, germ line selection was influenced by the different wild-type mtDNA haplotypes on specific nuclear genetic backgrounds⁹⁵. These included a polymorphic variant in nicotinamide nucleotide transhydrogenase (NNT) and the supercomplex assembly factor SCAF1 (orCox7a21)⁹⁵. Interestingly, the inheritance pattern could be modified by manipulating metabolites including ROS using N-acetylcysteine, casting light on the selection mechanisms and opening up new avenues for preventative therapy. These mouse results are consistent with the work in *Drosophila* discussed above, which shows these unnatural pairings of haplogroups can lead to two healthy, wild-type haplogroups competing in the germ line and the soma, leading to a positive or negative bias for one haplogroup not directly dependent on the gene products of the different mtDNAs. Precisely when and how the selection occurs can vary from mutation to mutation, with some showing selection during germ cell development, and others during transmission to the next generation. In conclusion, the emerging picture from animal studies is complex, with both negative (purifying) selection and selfish replication occurring in different species and for different mutations. It remains to be seen which of these mechanisms is active in the human population.

In humans, large unbiased studies of heteroplasmy transmission are challenging, but several have been performed in recent years. Analysing mtDNA sequences from 228 trios generated through the Dutch Genomes Project, evidence of selection was seen against heteroplasmic variants not previously observed in the human population, particularly involving rRNA and tRNA genes⁹⁶. A similar signature was observed in 96 multigenerational families in the US^{97,98}. However, studying ~10-fold more transmissions revealed a more complex picture in 1,526 mother–offspring pairs in the UK⁵⁸, where D-loop variants were more likely to be transmitted than rRNA gene variants. Interestingly, despite the ~10-fold higher substitution rate in the non-coding D-loop, several sites were completely devoid of any heteroplasmic variants, even at low (~1–2%) levels. The ‘mutation deserts’ included regions known to be involved in mtDNA transcription and replication, but also several sites with no known function, implying that these regions have important roles that are yet to be determined. By contrast, there was evidence of preferential transmission of previously observed variants in other regions of the D-loop⁵⁸, providing some evidence of positive selection reminiscent of the observations made in *Drosophila spp* (see above). Further evidence of selection was seen in primordial germ cells within the developing human germ line⁹⁹. Taken together, these findings imply that the pattern of homoplasmic variants seen around the globe is, at least in part, shaped by the inter-generational transmission of heteroplasmy, and the mechanisms seem to be acting on or within female germ cells.

Nuclear and environmental interactions

What factors drive the selection for and against specific mtDNA variants? All 13 mtDNA-encoded proteins directly interact with nuclear-encoded subunits to form fully functioning enzymes, providing

a platform for the interaction of the two distinctly inherited genomes. Evidence supporting this comes largely from animal studies, although there is nascent evidence in humans.

Initial indications of the importance of nuclear DNA (nDNA)–mtDNA matching came from *Nasonia* wasps, which are normally reproductively isolated due to *Walbachia* species. After curing the wasps of the endosymbiont, hybrids were possible, yet the haploid male F2 offspring showed lethal incompatibilities between specific chromosomal regions and the mitochondrial haplogroup^{100,101}.

More recent experiments focused on natural variation and inter-species introgression in *Drosophila*, where disrupting co-evolved mtDNA-nuclear (n)DNA partnerships in wild-type flies influenced a range of phenotypes including nuclear gene expression, development, ageing phenotypes, levels of activity and organismal fitness. Further evidence for the importance of nDNA–mtDNA matching was provided when different mtDNA haplotypes were injected into the cytoplasm of eggs with the same nuclear background. This experiment showed a hierarchy in haplotype fitness, with concordant effects across several different phenotypes, including key stages in development (egg-to-puparium and egg-to-eclosion). More recently, the experiments have been done the other way around, such that switching nuclear backgrounds had a marked effect on fecundity in association with different mtDNA genotypes involving egg production, and a role for the local environment in modulating the genotype–genotype interactions¹⁰²⁻¹⁰⁵.

Environment-driven co-evolution of mtDNA and nDNA has also been seen in several other species, including nematodes and fish, where local bioenergetic requirements drive mtDNA evolution with nuclear genome co-adaptation¹⁰⁶, but until recently there was limited evidence in mammals. Studying conplastic mice allows the investigation of different mtDNA variants on the same nuclear genetic background. Using this approach to exchange polymorphic mtDNA haplotypes led to differences in mitochondrial proteostasis and reactive oxygen species generation, resulting in late-onset phenotypes including defects of insulin signalling and obesity, which affected longevity¹⁰⁷. More recently, experiments performed on a mouse model of mitochondrial cardiomyopathy caused by knockout of the mitochondrial adenine nucleotide translocase (*SLC25A4/ANT1*^{-/-})¹⁰⁸. *SLC25A4/ANT1*^{-/-} mice harbouring a *MTND6*^{P25L} mtDNA haplotype had a more severe cardiomyopathy than *SLC25A4/ANT1*^{-/-} mice harbouring the *CO1*^{V421A} mtDNA haplotype, despite the fact that the *MTND6*^{P25L} and *CO1*^{V421A} haplotypes do not cause a severe phenotype on their own. These findings provide proof-of-principle that nDNA–mtDNA interactions are important in higher order species, endorsed by the recent description of a spontaneous mutation in the mouse *MT-CTB* gene (*m.G14904A*, *mt-Cyb*^{p.D254N}) that interacts with the nuclear gene variant *Bcs1P*^{p.S78G} to exacerbate the complex III deficiency through the Rieke Fe-S protein¹⁰⁹.

In humans, recent insights have been gained through the UK 100,000 genomes project. Independently analysing the nuclear genome sequence and the mtDNA sequence in 53,300 humans⁵⁸, it was possible to identify people in whom the two genomes had different ancestral origins (for example, a person with a European nuclear genome who harboured mtDNA originating in Asia that had been introduced by a woman many years previously). New heteroplasmies in this 'mismatched' group were more likely to have been seen before in people with the same nuclear genetic ancestry than in people with the same mtDNA ancestry. These findings imply that novel mtDNA variants are 'tolerated' or 'selected' by the cell nucleus, presumably in response to local environmental constraints. Further evidence for this comes from three laboratories developing mitochondrial transfer to prevent the transmission of pathogenic mtDNA mutations in the clinic^{90,110,111}. Stem cells were derived from embryos after mitochondrial transfer, where the original mtDNA heteroplasmy level had been reduced to <5%. When these lines were followed in culture, there was an unexpected high rate of reversion to the original genotype in ~15% of cases. It was initially thought that the propensity for reversion could be reduced by 'mtDNA haplogroup matching' the donor and recipient, but this no longer seems to be the case¹¹², pointing towards a nuclear genetic mechanism. Finally, one study provided some evidence that increasing discordance between the nuclear and mtDNA ancestry is associated with a reduced amount of mtDNA, perhaps through mildly compromised mtDNA replication¹¹³. These findings need to be replicated but resonate with independent work identifying nuclear genetic loci that seem to regulate the mtDNA content of blood cells¹¹⁴, although these findings have not always been consistent¹¹⁵.

Despite growing evidence that nuclear-mitochondrial interactions are important, beyond the direct physical interaction of respiratory chain protein subunits, the mechanisms are not totally clear. Emerging evidence from animal studies implicate ROS and canonical cell signalling pathways involved in the innate immune system. This is perhaps not surprising, given the origin of mitochondrial as a foreign protobacterium.

Implications for human disease.

mtDNA homoplasmy

Early studies correlating mtDNA haplogroups and homoplasmic mtDNA variants were often underpowered, leading to false-positive results that could not be reproduced¹¹⁶. However, consistent patterns have emerged from much larger studies in recent years, in part building on the success of nuclear genome-wide association studies (GWAS) for complex late-onset disorders. Two-stage discovery-replication studies have shown reproducible associations between mtDNA variants and both Parkinson disease and ischaemic stroke^{117,118}. Several of these findings were reproduced in a

further study of 38,638 cases with 11 different diseases compared to 17,483 controls¹¹⁹, revealing new associations with several autoimmune and psychiatric disorders. Intriguingly, this study showed that the same mtDNA allele can increase the risk of several ostensibly unrelated diseases (allelic pleiotropy), but decrease the risk of others.

The recent completion of national genomics initiatives has added weight to these findings. The CHARGE_{mtDNA}+ consortium measured anthropomorphic and metabolic parameters in 45 cohorts¹²⁰ and found several associations between variants in mtDNA and nuclear mitochondrial genes and cardiometabolic traits that reached nominal significance, implicating their involvement in glucose and insulin metabolism, adipocyte regulation, diabetes mellitus and cardiovascular disease. Imputing mtDNA variants in 147,437 individuals in the Japan Biobank⁶² confirmed the pleiotropic effects of different mtDNA variants and identified novel associations with new complex and quantitative traits, including creatine kinase and kidney function. Similar work has been carried out on 378,670 people in the UK Biobank, confirming the Japanese findings, and identifying new associations with multiple sclerosis, height, longevity and serum biomarkers of liver and kidney function and many blood cell traits (unpublished observations). Although the effect sizes attributable to the individual mtDNA variants were small, these findings illustrate the continuous spectrum of genetic effects, with common variants contributing to the normal physiological variation through to severe mutations causing inherited diseases involving the same organs. Given that many of the mtDNA associations involve late onset diseases, they are unlikely to undergo germ line selection, explaining why they have reached a high frequency in the population. In keeping with this, more recent variants appear to be more likely to increase the disease risk than to decrease disease risk¹¹⁹, consistent with the view that novel disadvantageous variants had not yet been purified at the individual level.

MtDNA heteroplasmy

Heteroplasmic mtDNA mutations accumulate within cells as humans age, driven by *de novo* mutagenesis and clonal expansion. Deletions preferentially accumulate within non-dividing tissues, leading to an age-related accumulation of biochemically deficient cells that may contribute to age-related diseases, including Alzheimer and Parkinson disease^{121,122}. mtSNVs also accumulate throughout life in healthy individuals, with a disproportionate number detected in the D-loop, including some recurrent variant sites reaching very high levels. In both contexts, the process of clonal expansion plays a key role in determining which mutations reach detectable levels.

Recent deep sequencing studies indicate that some of the previously presumed somatic mutations were actually inherited down the maternal line at low levels⁵⁰, subsequently undergoing clonal expansion. Studying this process experimentally is challenging because clonal expansion

takes considerable time; however, key insights have been made in mice with a proofread-deficient polymerase, where inherited and acquired mtSNVs interacted synergistically to cause age-related phenotypes¹²³. The role of mtDNA mutations in human ageing is supported by observations in survivors of HIV infection, in whom previous nucleoside analogue therapy accelerated the process of clonal expansion, leading to levels of biochemically deficient muscle fibers way above those seen in normal ageing¹²⁴.

Although recent population studies have shown the accelerated accumulation of heteroplasmic mtDNA variants in several contexts, including circulating leucocytes cardiometabolic disease, it remains to be determined whether these simply reflect the underlying pathology, or they contribute to the pathogenesis. Intriguingly, specific mutations appear to accumulate in some tissues during life, but not others¹²⁵, potentially influenced by tissue-specific selective forces¹²⁶. The picture is a little clearer for cancer, although not definitively proven, in large part through the work of the International Cancer Genome Consortium (ICGC) and the Cancer Genome Atlas (TCGA) projects as a part of Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium^{18,127}. Deep sequencing a range of tumors and adjacent 'normal' tissue from thousands of patients has shown the diversity of mtDNA mutations in human cancers, but patterns are starting to emerge. Unlike the nuclear genome, mtDNA mutations in cancers show a similar signature, with a marked strand asymmetry with predominant G>A and T>C substitutions and deficient complementary C>T and A>G substitutions on the light (L) strand of the mitochondrial genome, reflecting a common mechanism. Although this could be because the mutations arose in healthy tissue earlier in life and hitch-hiked on an expanding clone of cancer cells. In keeping with this, overall, the ratio of non-synonymous to synonymous SNVs was ~1. However, as seen previously^{128,129}, some cancers showed evidence of selection against non-synonymous variants, and in others, selection in favour of functional variants, and transcriptome analysis showed the disruption of gene networks involved in mitochondrial metabolism, implicating them in the progression of some cancers but not others. Similar studies are long overdue for neurodegenerative diseases, and it is clear that the studies will need to be large to detect effects that may only be relevant for a particular subgroup of mtDNA variants or subgroups of disease. It will be challenging to provide definitive experimental proof of a causal role, but this is important to do given the growing prevalence of these disorders.

Conclusions

Technological advances have driven a step change in our understanding of mtDNA and its ongoing evolution in humans. Selective forces acting directly on the genome, on the organelle, on the cell (particularly in the germ line), and on the organism continue to shape the genome throughout the world, and are influenced by the nuclear genome and exposure to environmental constraints. We are

only just beginning to understand some of the molecular mechanisms, but they are likely to be complex. A mtDNA genotype that is beneficial in one context may be disadvantageous in another, creating the impression of 'selfish' behaviour. Perhaps this explains why some pathogenic mutations are so much more common than others, such as the m.3243A>G mutation which is overwhelmingly the most common heteroplasmic mutation seen in the clinic³. It will be important to understand how and why specific heteroplasmic variants behave in the way they do. In the short term, this will provide insight into the unexplained mtDNA haplotype reversion seen following nuclear transfer^{90,110,111}, which could lead to disease recurrence after an ostensibly effective mtDNA 'clean up'. In the longer term, harnessing these mechanisms has the potential to lead to new treatments (**Box 3**) which suppress the acquisition of mtDNA variants in disease and as we age.

Acknowledgements

JBS is supported by the Max Planck Society and a project grant from the United Mitochondrial Disease Foundation. PFC is a Wellcome Trust Principal Research Fellow (212219/Z/18/Z), and a UK NIHR Senior Investigator, who receives support from the Medical Research Council Mitochondrial Biology Unit (MC_UU_00015/9), the Medical Research Council (MRC) International Centre for Genomic Medicine in Neuromuscular Disease, the Leverhulme Trust (RPG-2018-408), and the National Institute for Health Research (NIHR) Biomedical Research Centre based at Cambridge University Hospitals NHS Foundation Trust and the University of Cambridge. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests

The authors declare no conflicts of interests.

References

1. Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-65 (1981).
2. Holt, I., Harding, A.E. & Morgan-Hughes, J.A. Deletion of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* **331**, 717-719 (1988).
3. Gorman, G.S. *et al.* Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann Neurol* **77**, 753-9 (2015).
4. Bendall, K.E., Macaulay, V.A., Baker, J.R. & Sykes, B.C. Heteroplasmic point mutations in the human mtDNA control region. *American Journal of Human Genetics* **59**, 1276-1287 (1996).
5. Parsons, T.J. *et al.* A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics* **15**, 363-8 (1997).
6. Calvo, S.E. & Mootha, V.K. The mitochondrial proteome and human disease. *Annu Rev Genomics Hum Genet* **11**, 25-44 (2010).
7. Chan, D.C. Mitochondrial Dynamics and Its Involvement in Disease. *Annu Rev Pathol* **15**, 235-259 (2020).
8. Chandel, N.S. Evolution of Mitochondria as Signaling Organelles. *Cell Metab* **22**, 204-6 (2015).
9. Giorgi, C., Marchi, S. & Pinton, P. The machineries, regulation and cellular functions of mitochondrial calcium. *Nat Rev Mol Cell Biol* **19**, 713-730 (2018).
10. Pfanner, N., Warscheid, B. & Wiedemann, N. Mitochondrial proteins: from biogenesis to functional networks. *Nat Rev Mol Cell Biol* **20**, 267-284 (2019).
11. Martijn, J., Vosseberg, J., Guy, L., Offre, P. & Ettema, T.J.G. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* **557**, 101-105 (2018).
12. Fan, L. Mitochondria branch within Alphaproteobacteria. *bioRxiv* **715870**(2019).
13. Gray, M.W. Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. *Proc Natl Acad Sci U S A* **112**, 10133-8 (2015).
14. Hazkani-Covo, E., Zeller, R.M. & Martin, W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet* **6**, e1000834 (2010).
15. Dayama, G., Emery, S.B., Kidd, J.M. & Mills, R.E. The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res* **42**, 12640-9 (2014).
16. Kukat, C. *et al.* Cross-strand binding of TFAM to a single mtDNA molecule forms the mitochondrial nucleoid. *Proc Natl Acad Sci U S A* **112**, 11288-93 (2015).
17. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415-21 (2013).
18. Ju, Y.S. *et al.* Origins and functional consequences of somatic mitochondrial DNA mutations in human cancer. *Elife* **3**(2014).
19. Kennedy, S.R., Salk, J.J., Schmitt, M.W. & Loeb, L.A. Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLoS Genet* **9**, e1003794 (2013).
20. Birky, C.W. Relaxed and stringent genomes: why cytoplasmic genes don't obey Mendel's laws. *Journal of Heredity* **85**, 355-365 (1994).
21. Gross, N.J., Getz, G.S. & Rabinowitz, M. Apparent turnover of mitochondrial deoxyribonucleic acid and mitochondrial phospholipids in the tissues of the rat. *J Biol Chem* **244**, 1552-62 (1969).
22. Krasich, R. & Copeland, W.C. DNA polymerases in the mitochondria: A critical review of the evidence. *Front Biosci (Landmark Ed)* **22**, 692-709 (2017).
23. Zheng, W., Khrapko, K., Coller, H.A., Thilly, W.G. & Copeland, W.C. Origins of human mitochondrial point mutations as DNA polymerase gamma-mediated errors. *Mutat Res* **599**, 11-20 (2006).

24. Yasukawa, T. & Kang, D. An overview of mammalian mitochondrial DNA replication mechanisms. *J Biochem* **164**, 183-193 (2018).
25. Persson, O. *et al.* Copy-choice recombination during mitochondrial L-strand synthesis causes DNA deletions. *Nat Commun* **10**, 759 (2019).
26. Faith, J.J. & Pollock, D.D. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics* **165**, 735-45 (2003).
27. Fonseca, M.M., Harris, D.J. & Posada, D. The inversion of the Control Region in three mitogenomes provides further evidence for an asymmetric model of vertebrate mtDNA replication. *PLoS One* **9**, e106654 (2014).
28. Holt, I.J. & Reyes, A. Human mitochondrial DNA replication. *Cold Spring Harb Perspect Biol* **4**(2012).
29. Reyes, A., Yang, M.Y., Bowmaker, M. & Holt, I.J. Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. *J Biol Chem* **280**, 3242-50 (2005).
30. Reyes, A. *et al.* Mitochondrial DNA replication proceeds via a 'bootlace' mechanism involving the incorporation of processed transcripts. *Nucleic Acids Res* **41**, 5837-50 (2013).
31. Chinnery, P.F. & Samuels, D.C. Relaxed replication of mtDNA: a model with implications for the expression of disease. *American Journal of Human Genetics* **64**, 1158-1165 (1999).
32. Diaz, F. *et al.* Human mitochondrial DNA with large deletions repopulates organelles faster than full-length genomes under relaxed copy number control. *Nucleic Acids Res* **30**, 4626-33 (2002).
33. Ma, H. & O'Farrell, P.H. Selfish drive can trump function when animal mitochondrial genomes compete. *Nat Genet* **48**, 798-802 (2016).
34. Gitschlag, B.L. *et al.* Homeostatic Responses Regulate Selfish Mitochondrial Genome Dynamics in *C. elegans*. *Cell Metab* **24**, 91-103 (2016).
35. Phillips, W.S. *et al.* Selfish Mitochondrial DNA Proliferates and Diversifies in Small, but not Large, Experimental Populations of *Caenorhabditis briggsae*. *Genome Biol Evol* **7**, 2023-37 (2015).
36. Davis, A.F. & Clayton, D.A. In situ localisation of mitochondrial DNA replication in intact mammalian cells. *Journal of Cell Biology* **135**, 883-893 (1996).
37. Wai, T., Teoli, D. & Shoubridge, E.A. The mitochondrial DNA genetic bottleneck results from replication of a subpopulation of genomes. *Nat Genet* **40**, 1484-8 (2008).
38. Chinnery, P.F., Samuels, D.C., Elson, J. & Turnbull, D.M. Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism? *Lancet* **360**, 1323-5 (2002).
39. Hauswirth, W.M. & Laipis, P.J. Mitochondrial DNA polymorphism in a maternal lineage of Holstien cows. *Proceedings of the National Academy of Sciences USA* **79**, 4686-4690 (1982).
40. Olivo, P.D., Van de Walle, M.J., Laipis, P.J. & Hauswirth, W.W. Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature* **306**, 400-2 (1983).
41. Hauswirth, W.W. & Laipis, P.J. Transmission genetics of mammalian mitochondria: a molecular model and experimental evidence. in *Achievements and perspectives of mitochondrial research*, Vol. II (ed. Quagliariello, E.) 49-59 (Elsevier, Amsterdam, 1985).
42. Cree, L.M. *et al.* A reduction of mitochondrial DNA molecules during embryogenesis explains the rapid segregation of genotypes. *Nat Genet* **40**, 249-54 (2008).
43. Otten, A.B. *et al.* Differences in Strength and Timing of the mtDNA Bottleneck between Zebrafish Germline and Non-germline Cells. *Cell Rep* **16**, 622-30 (2016).
44. Chinnery, P.F. *et al.* The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? *Trends Genet* **16**, 500-5 (2000).
45. Wilson, I.J. *et al.* Mitochondrial DNA sequence characteristics modulate the size of the genetic bottleneck. *Hum Mol Genet* **25**, 1031-41 (2016).

46. Hellebrekers, D.M. *et al.* PGD and heteroplasmic mitochondrial DNA point mutations: a systematic review estimating the chance of healthy offspring. *Hum Reprod Update* **18**, 341-9 (2012).
47. Freyer, C. *et al.* Variation in germline mtDNA heteroplasmy is determined prenatally but modified during subsequent transmission. *Nat Genet* **44**, 1282-5 (2012).
48. Johnston, I.G. *et al.* Stochastic modelling, Bayesian inference, and new in vivo measurements elucidate the debated mtDNA bottleneck mechanism. *Elife* **4**, e07464 (2015).
49. Burr, S.P., Pezet, M. & Chinnery, P.F. Mitochondrial DNA Heteroplasmy and Purifying Selection in the Mammalian Female Germ Line. *Dev Growth Differ* **60**, 21-32 (2018).
50. Payne, B.A. *et al.* Universal heteroplasmy of human mitochondrial DNA. *Hum Mol Genet* **22**, 384-90 (2013).
51. Salk, J.J., Schmitt, M.W. & Loeb, L.A. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat Rev Genet* **19**, 269-285 (2018).
52. He, Y. *et al.* Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* **464**, 610-4 (2010).
53. Li, M., Schroder, R., Ni, S., Madea, B. & Stoneking, M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc Natl Acad Sci U S A* **112**, 2491-6 (2015).
54. Li, M. *et al.* Detecting heteroplasmy from high-throughput sequencing of complete human mitochondrial DNA genomes. *Am J Hum Genet* **87**, 237-49 (2010).
55. Giuliani, C. *et al.* Transmission from centenarians to their offspring of mtDNA heteroplasmy revealed by ultra-deep sequencing. *Aging (Albany NY)* **6**, 454-67 (2014).
56. Kennedy, S.R. *et al.* Detecting ultralow-frequency mutations by Duplex Sequencing. *Nat Protoc* **9**, 2586-606 (2014).
57. Schmitt, M.W. *et al.* Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A* **109**, 14508-13 (2012).
58. Wei, W. *et al.* Germline selection shapes human mitochondrial DNA diversity. *Science* **364**(2019).
59. Wallace, D.C. Genetics: Mitochondrial DNA in evolution and disease. *Nature* **535**, 498-500 (2016).
60. van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* **30**, E386-94 (2009).
61. Wei, W., Gomez-Duran, A., Hudson, G. & Chinnery, P.F. Background sequence characteristics influence the occurrence and severity of disease-causing mtDNA mutations. *PLoS Genet* **13**, e1007126 (2017).
62. Yamamoto, K. *et al.* Genetic and phenotypic landscape of the mitochondrial genome in the Japanese population. *Commun Biol* **3**, 104 (2020).
63. Cann, R.L., Brown, W.M. & Wilson, A.C. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* **106**, 479-99 (1984).
64. Howell, N. *et al.* The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am J Hum Genet* **72**, 659-70 (2003).
65. Mishmar, D. *et al.* Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci U S A* **100**, 171-6 (2003).
66. Kivisild, T. *et al.* The role of selection in the evolution of human mitochondrial genomes. *Genetics* **172**, 373-87 (2006).
67. Baudouin, S.V. *et al.* Mitochondrial DNA and survival after sepsis: a prospective study. *Lancet* **366**, 2118-21 (2005).
68. Lorente, L. *et al.* Survival and mitochondrial function in septic patients according to mitochondrial DNA haplogroup. *Crit Care* **16**, R10 (2012).

- 659 69. Lorente, L. *et al.* Septic patients with mitochondrial DNA haplogroup JT have higher
660 respiratory complex IV activity and survival rate. *J Crit Care* **33**, 95-9 (2016).
- 661 70. Yang, Y. *et al.* Mitochondrial DNA haplogroup R predicts survival advantage in severe sepsis
662 in the Han population. *Genet Med* **10**, 187-92 (2008).
- 663 71. Ji, F. *et al.* Mitochondrial DNA variant associated with Leber hereditary optic neuropathy and
664 high-altitude Tibetans. *Proc Natl Acad Sci U S A* **109**, 7391-6 (2012).
- 665 72. Lahti, D.C. *et al.* Relaxed selection in the wild. *Trends Ecol Evol* **24**, 487-96 (2009).
- 666 73. Gomez-Duran, A. *et al.* Oxidative phosphorylation differences between mitochondrial DNA
667 haplogroups modify the risk of Leber's hereditary optic neuropathy. *Biochim Biophys Acta*
668 **1822**, 1216-22 (2012).
- 669 74. Gomez-Duran, A. *et al.* Unmasking the causes of multifactorial disorders: OXPHOS differences
670 between mitochondrial haplogroups. *Hum Mol Genet* **19**, 3343-53 (2010).
- 671 75. Klucnika, A. & Ma, H. A battle for transmission: the cooperative and selfish animal
672 mitochondrial genomes. *Open Biol* **9**, 180267 (2019).
- 673 76. Zhang, Y. *et al.* PINK1 Inhibits Local Protein Synthesis to Limit Transmission of Deleterious
674 Mitochondrial DNA Mutations. *Mol Cell* **73**, 1127-1137 e5 (2019).
- 675 77. Lieber, T., Jeedigunta, S.P., Palozzi, J.M., Lehmann, R. & Hurd, T.R. Mitochondrial
676 fragmentation drives selective removal of deleterious mtDNA in the germline. *Nature* (2019).
- 677 78. Narendra, D., Tanaka, A., Suen, D.F. & Youle, R.J. Parkin is recruited selectively to impaired
678 mitochondria and promotes their autophagy. *J Cell Biol* **183**, 795-803 (2008).
- 679 79. Milot, E. *et al.* Mother's curse neutralizes natural selection against a human genetic disease
680 over three centuries. *Nat Ecol Evol* **1**, 1400-1406 (2017).
- 681 80. Beziat, F., Morel, F., Volz-Lingenhol, A., Saint Paul, N. & Alziari, S. Mitochondrial genome
682 expression in a mutant strain of *D. subobscura*, an animal model for large scale mtDNA
683 deletion. *Nucleic Acids Res* **21**, 387-92 (1993).
- 684 81. Tsang, W.Y. & Lemire, B.D. Stable heteroplasmy but differential inheritance of a large
685 mitochondrial DNA deletion in nematodes. *Biochem Cell Biol* **80**, 645-54 (2002).
- 686 82. Gitschlag, B.L. *et al.* Homeostatic Responses Regulate Selfish Mitochondrial Genome
687 Dynamics in *C. elegans*. *Cell Metabolism* **24**, 91-103 (2016).
- 688 83. Lin, Y.F. *et al.* Maintenance and propagation of a deleterious mitochondrial genome by the
689 mitochondrial unfolded protein response. *Nature* **533**, 416-9 (2016).
- 690 84. Fan, W. *et al.* A mouse model of mitochondrial disease reveals germline selection against
691 severe mtDNA mutations. *Science* **319**, 958-62 (2008).
- 692 85. Stewart, J.B. *et al.* Strong purifying selection in transmission of mammalian mitochondrial
693 DNA. *PLoS Biol* **6**, e10 (2008).
- 694 86. Kauppila, J.H.K. *et al.* A Phenotype-Driven Approach to Generate Mouse Models with
695 Pathogenic mtDNA Mutations Causing Mitochondrial Disease. *Cell Rep* **16**, 2980-2990 (2016).
- 696 87. Jenuth, J., Peterson, A.C., Fu, K. & Shoubridge, E.A. Random genetic drift in the female germ
697 line explains the rapid segregation of mammalian mitochondrial DNA. *Nature Genetics* **14**,
698 146-151 (1996).
- 699 88. Meirelles, F. & Smith, L.C. Mitochondrial genotype segregation in a mouse heteroplasmic
700 lineage produced by embryonic karyoplast transplantation. *Genetics* **145**, 445-451 (1997).
- 701 89. Lee, H.S. *et al.* Rapid mitochondrial DNA segregation in primate preimplantation embryos
702 precedes somatic and germline bottleneck. *Cell Rep* **1**, 506-15 (2012).
- 703 90. Kang, E. *et al.* Mitochondrial replacement in human oocytes carrying pathogenic mitochondrial
704 DNA mutations. *Nature* **540**, 270-275 (2016).
- 705 91. Jokinen, R. *et al.* Gimap3 regulates tissue-specific mitochondrial DNA segregation. *PLoS Genet*
706 **6**, e1001161 (2010).
- 707 92. Battersby, B.J., Loredó-Osti, J.C. & Shoubridge, E.A. Nuclear genetic control of mitochondrial
708 DNA segregation. *Nat Genet* **33**, 183-6 (2003).

93. Sharpley, M.S. *et al.* Heteroplasmy of mouse mtDNA is genetically unstable and results in altered behavior and cognition. *Cell* **151**, 333-343 (2012).
94. Burgstaller, J.P. *et al.* MtDNA segregation in heteroplasmic tissues is common in vivo and modulated by haplotype differences and developmental stage. *Cell Rep* **7**, 2031-2041 (2014).
95. Latorre-Pellicer, A. *et al.* Regulation of Mother-to-Offspring Transmission of mtDNA Heteroplasmy. *Cell Metab* **30**, 1120-1130 e5 (2019).
96. Li, M. *et al.* Transmission of human mtDNA heteroplasmy in the Genome of the Netherlands families: support for a variable-size bottleneck. *Genome Res* **26**, 417-26 (2016).
97. Barrett, A. *et al.* Pronounced somatic bottleneck in mitochondrial DNA of human hair. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190175 (2020).
98. Zaidi, A.A. *et al.* Bottleneck and selection in the germline and maternal age influence transmission of mitochondrial DNA in human pedigrees. *Proc Natl Acad Sci U S A* **116**, 25172-25178 (2019).
99. Floros, V.I. *et al.* Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos. *Nat Cell Biol* **20**, 144-151 (2018).
100. Niehuis, O., Judson, A.K. & Gadau, J. Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids of *Nasonia giraulti* and *N. vitripennis*. *Genetics* **178**, 413-26 (2008).
101. Ellison, C.K., Niehuis, O. & Gadau, J. Hybrid breakdown and mitochondrial dysfunction in hybrids of *Nasonia* parasitoid wasps. *J Evol Biol* **21**, 1844-51 (2008).
102. Rand, D.M. & Mossman, J.A. Mitonuclear conflict and cooperation govern the integration of genotypes, phenotypes and environments. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190188 (2020).
103. Mossman, J.A., Ge, J.Y., Navarro, F. & Rand, D.M. Mitochondrial DNA Fitness Depends on Nuclear Genetic Background in *Drosophila*. *G3 (Bethesda)* **9**, 1175-1188 (2019).
104. Vaught, R.C. *et al.* Interactions between cytoplasmic and nuclear genomes confer sex-specific effects on lifespan in *Drosophila melanogaster*. *J Evol Biol* **33**, 694-713 (2020).
105. Nagarajan-Radha, V., Rapkin, J., Hunt, J. & Dowling, D.K. Interactions Between Mitochondrial Haplotype and Dietary Macronutrient Ratios Confer Sex-Specific Effects on Longevity in *Drosophila melanogaster*. *J Gerontol A Biol Sci Med Sci* **74**, 1573-1581 (2019).
106. Blier, P.U., Dufresne, F. & Burton, R.S. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends in Genetics* **17**, 400-406 (2001).
107. Latorre-Pellicer, A. *et al.* Mitochondrial and nuclear DNA matching shapes metabolism and healthy ageing. *Nature* **535**, 561-5 (2016).
108. McManus, M.J. *et al.* Mitochondrial DNA Variation Dictates Expressivity and Progression of Nuclear DNA Mutations Causing Cardiomyopathy. *Cell Metab* **29**, 78-90 e5 (2019).
109. Purhonen, J. *et al.* A spontaneous mitonuclear epistasis converging on Rieske Fe-S protein exacerbates complex III deficiency in mice. *Nat Commun* **11**, 322 (2020).
110. Hyslop, L.A. *et al.* Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease. *Nature* **534**, 383-6 (2016).
111. Paull, D. *et al.* Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. *Nature* **493**, 632-7 (2013).
112. Hudson, G., Takeda, Y. & Herbert, M. Reversion after replacement of mitochondrial DNA. *Nature* **574**, E8-E11 (2019).
113. Zaidi, A.A. & Makova, K.D. Investigating mitonuclear interactions in human admixed populations. *Nat Ecol Evol* **3**, 213-222 (2019).
114. Cai, N. *et al.* Genetic Control over mtDNA and Its Relationship to Major Depressive Disorder. *Curr Biol* **25**, 3170-7 (2015).
115. Guyatt, A.L. *et al.* A genome-wide association study of mitochondrial DNA copy number in two population-based cohorts. *Hum Genomics* **13**, 6 (2019).

116. Samuels, D.C., Carothers, A.D., Horton, R. & Chinnery, P.F. The power to detect disease associations with mitochondrial DNA haplogroups. *Am J Hum Genet* **78**, 713-20 (2006).
117. Chinnery, P.F., Elliott, H.R., Syed, A., Rothwell, P.M. & Oxford Vascular, S. Mitochondrial DNA haplogroups and risk of transient ischaemic attack and ischaemic stroke: a genetic association study. *Lancet Neurol* **9**, 498-503 (2010).
118. Hudson, G. *et al.* Two-stage association study and meta-analysis of mitochondrial DNA variants in Parkinson disease. *Neurology* **80**, 2042-8 (2013).
119. Hudson, G., Gomez-Duran, A., Wilson, I.J. & Chinnery, P.F. Recent mitochondrial DNA mutations increase the risk of developing common late-onset human diseases. *PLoS Genet* **10**, e1004369 (2014).
120. Kraja, A.T. *et al.* Associations of Mitochondrial and Nuclear Mitochondrial Variants and Genes with Seven Metabolic Traits. *Am J Hum Genet* **104**, 112-138 (2019).
121. Bender, A. *et al.* High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat Genet* (2006).
122. Simon, D.K. *et al.* Somatic mitochondrial DNA mutations in cortex and substantia nigra in aging and Parkinson's disease. *Neurobiol Aging* **25**, 71-81 (2004).
123. Ross, J.M. *et al.* Germline mitochondrial DNA mutations aggravate ageing and can impair brain development. *Nature* **501**, 412-5 (2013).
124. Payne, B.A. *et al.* Mitochondrial aging is accelerated by anti-retroviral therapy through the clonal expansion of mtDNA mutations. *Nat Genet* **43**, 806-10 (2011).
125. Samuels, D.C. *et al.* Recurrent tissue-specific mtDNA mutations are common in humans. *PLoS Genet* **9**, e1003929 (2013).
126. Hubner, A. *et al.* Sharing of heteroplasmies between human liver lobes varies across the mtDNA genome. *Sci Rep* **9**, 11219 (2019).
127. Yuan, Y. *et al.* Comprehensive molecular characterisation of mitochondrial genomes in human cancers. *Nature Genetics In press*(2020).
128. Stewart, J.B. *et al.* Simultaneous DNA and RNA Mapping of Somatic Mitochondrial Mutations across Diverse Human Cancers. *PLoS Genet* **11**, e1005333 (2015).
129. Grandhi, S. *et al.* Heteroplasmic shifts in tumor mitochondrial genomes reveal tissue-specific signals of relaxed and positive selection. *Hum Mol Genet* **26**, 2912-2922 (2017).
130. Kuhl, I. *et al.* POLRMT does not transcribe nuclear genes. *Nature* **514**, E7-11 (2014).
131. Albayrak, L. *et al.* The ability of human nuclear DNA to cause false positive low-abundance heteroplasmy calls varies across the mitochondrial genome. *BMC Genomics* **17**, 1017 (2016).
132. Maude, H. *et al.* NUMT Confounding Biases Mitochondrial Heteroplasmy Calls in Favor of the Reference Allele. *Front Cell Dev Biol* **7**, 201 (2019).
133. Luo, S. *et al.* Biparental Inheritance of Mitochondrial DNA in Humans. *Proc Natl Acad Sci U S A* **115**, 13039-13044 (2018).
134. Wei, W. *et al.* Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans. *Nat Commun* **11**, 1740 (2020).
135. Wei, W. & Chinnery, P.F. Inheritance of mitochondrial DNA in humans: implications for rare and common diseases. *J Intern Med* **287**, 634-644 (2020).
136. van den Amele, J., Li, A.Y.Z., Ma, H. & Chinnery, P.F. Mitochondrial heteroplasmy beyond the oocyte bottleneck. *Semin Cell Dev Biol* **97**, 156-166 (2020).
137. Dou, X. *et al.* The strand-biased mitochondrial DNA methylome and its regulation by DNMT3A. *Genome Res* **29**, 1622-1634 (2019).
138. Saini, S.K., Mangalhar, K.C., Prakasam, G. & Bamezai, R.N.K. DNA Methyltransferase1 (DNMT1) Isoform3 methylates mitochondrial genome and modulates its biology. *Sci Rep* **7**, 1525 (2017).

139. van der Wijst, M.G., van Tilburg, A.Y., Ruiters, M.H. & Rots, M.G. Experimental mitochondria-targeted DNA methylation identifies GpC methylation, not CpG methylation, as potential regulator of mitochondrial gene expression. *Sci Rep* **7**, 177 (2017).
140. Rebelo, A.P., Williams, S.L. & Moraes, C.T. In vivo methylation of mtDNA reveals the dynamics of protein-mtDNA interactions. *Nucleic Acids Res* **37**, 6701-15 (2009).
141. Calvo, S.E., Clauser, K.R. & Mootha, V.K. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. *Nucleic Acids Res* **44**, D1251-7 (2016).
142. Ratel, D. *et al.* Undetectable levels of N6-methyl adenine in mouse DNA: Cloning and analysis of PRED28, a gene coding for a putative mammalian DNA adenine methyltransferase. *FEBS Lett* **580**, 3179-84 (2006).
143. Douvlataniotis, K., Bensberg, M., Lentini, A., Gylemo, B. & Nestor, C.E. No evidence for DNA N (6)-methyladenine in mammals. *Sci Adv* **6**, eaay3335 (2020).
144. Hao, Z. *et al.* N(6)-Deoxyadenosine Methylation in Mammalian Mitochondrial DNA. *Mol Cell* **78**, 382-395 e8 (2020).
145. Ameer, A. *et al.* Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins. *PLoS Genet* **7**, e1002028 (2011).
146. Kaupila, J.H.K. *et al.* Base-excision repair deficiency alone or combined with increased oxidative stress does not increase mtDNA point mutations in mice. *Nucleic Acids Res* **46**, 6642-6669 (2018).
147. Stewart, J.B. & Chinnery, P.F. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet* **16**, 530-42 (2015).
148. Salas, A., Schonherr, S., Bandelt, H.J., Gomez-Carballa, A. & Weissensteiner, H. Extraordinary claims require extraordinary evidence in asserted mtDNA biparental inheritance. *Forensic Sci Int Genet* **47**, 102274 (2020).
149. Santibanez-Koref, M. *et al.* Assessing mitochondrial heteroplasmy using next generation sequencing: A note of caution. *Mitochondrion* **46**, 302-306 (2019).
150. Just, R.S., Irwin, J.A. & Parson, W. Questioning the prevalence and reliability of human mitochondrial DNA heteroplasmy from massively parallel sequencing data. *Proc Natl Acad Sci U S A* **111**, E4546-7 (2014).
151. Li, M., Schroeder, R., Ko, A. & Stoneking, M. Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs. *Nucleic Acids Res* **40**, e137 (2012).
152. Calvignac, S., Konecny, L., Malard, F. & Douady, C.J. Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion* **11**, 246-54 (2011).
153. Fu, K. *et al.* A novel heteroplasmic tRNA^{leu(UUR)} mtDNA point mutation in a sporadic patient with mitochondrial encephalomyopathy segregates rapidly in muscle and suggests an approach to therapy. *Human Molecular Genetics* **5**, 1835-1840 (1996).
154. Clark, K. *et al.* Correction of a mitochondrial DNA defect in human skeletal muscle. *Nature Genetics* **16**, 222-224 (1997).
155. Gammage, P.A. *et al.* Genome editing in mitochondria corrects a pathogenic mtDNA mutation in vivo. *Nat Med* **24**, 1691-1695 (2018).
156. Bacman, S.R. *et al.* MitoTALEN reduces mutant mtDNA load and restores tRNA(Ala) levels in a mouse model of heteroplasmic mtDNA mutation. *Nat Med* **24**, 1696-1700 (2018).
157. Mok, B.Y. *et al.* A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**, 631-637 (2020).
158. Steele, H. *et al.* Metabolic effects of bezafibrate in mitochondrial disease. *EMBO Mol Med* **12**, e11589 (2020).
159. Pirinen, E. *et al.* Niacin Cures Systemic NAD(+) Deficiency and Improves Muscle Performance in Adult-Onset Mitochondrial Myopathy. *Cell Metab* (2020).

Glossary

Allele frequency

(also known as Variant Allele Frequency). The proportion of alleles carrying a specific variant. This term can be used to describe the proportion of individuals with a homoplasmic mtDNA variant in a population group; or the proportion of mtDNA molecules within an individual person, organ, cell or mitochondrion. The latter is the same as the heteroplasmy fraction, or heteroplasmy burden.

Clonal expansion

When a mtDNA molecule containing a specific variant or haplotype increases in allele frequency

Cybrid

A cellular model system usually used to study the effects of mtDNA variants or mutations on the same nuclear genetic background.

Displacement (D) loop

~1.1Kb of non-coding mtDNA in humans involved in the initiation of transcription and mtDNA replication (see Figure 1).

Heteroplasmy

A mixture of wild-type and mutant mtDNA. The proportion of mutant and wild-type molecules is often referred to as the heteroplasmy percentage, heteroplasmy frequency, or heteroplasmy burden (see allele frequency).

Homoplasmy

When all mtDNA molecules are identical.

Mitochondrial dynamics

Fusion and fission of mitochondria

Mitophagy

A specific form of autophagy where mitochondria are engulfed by a lysosome and destroyed.

NuMTs

891 Nuclear-mitochondrial sequences - segments of mtDNA that have translocated into nuclear
892 chromosomes of the cell.

893

894 ***Relaxed replication***

895 Replication of mtDNA throughout the cell cycle and is independent of nuclear division.

896

897 ***Replicative advantage***

898 When one mtDNA molecule is preferentially copied over another molecule.

899

900 ***Selfish replication***

901 When a genotype is preferentially propagated, despite have a deleterious effect on the organism or
902 population

903

904 ***Variant allele frequency***

905 See allele frequency

906

907 ***Vegetative segregation***

908 Changes in heteroplasmy that occur through the unequal partitioning of different mitochondrial
909 genotypes during cell division.

910

Figures

Figure 1 | Human mitochondrial DNA. The human mitochondrial genome (mtDNA) is a 16,569 base pair circle of double stranded DNA with an inner 'light' (L)-strand, and an outer 'heavy' (H)-strand reflecting a higher guanine content. MtDNA genes code for protein subunits of the mitochondrial respiratory chain include *MTND1–MTND6* and *MTND4L* (complex I); *MTCYB* (complex III); *MTCO1–3* (complex IV); and *MTATP6* and *MTATP8* (complex V); 22 tRNA and 2 rRNA genes required for intra-mitochondrial protein synthesis; and the 1.1Kb non-coding non-coding region (NCR) which includes the displacement (D-) loop involved in the regulation of transcription and mtDNA synthesis. Transcription of human mtDNA relies on at least three transcription promoters: Hsp1, Hsp2 and the LSP. Hsp1 allows transcription of the two ribosomal RNAs. Hsp2 leads to transcription of the rest of the heavy strand in one large polycistronic transcript, which is subsequently spliced into functional tRNA, rRNA and mRNAs. The LSP promotes transcription of the light strand as either one long transcript or as several small primer transcripts by RNase mitochondrial RNA processing (MRP). Transcription initiation involves several proteins, including the mitochondrial RNA polymerase and the mitochondrial transcription factors A and B2¹³⁰. MtDNA replication is described in the text. There is an ongoing debate as to whether mtDNA is methylated or not (**Text Box 1**)

Figure 2 | Nuclear mitochondrial sequences (NuMTs) & the paternal inheritance of mtDNA. Segments of mtDNA have become embedded throughout the nuclear genome throughout evolution, and are often referred to as nuclear-mitochondrial sequences (NuMTs). Most NuMTs are found in the non-coding space and range in size from very small to extremely large, including a representation of the entire mtDNA on chromosome 1. NUMTs have recently attracted interest because for two reasons. First, they can confound mtDNA analysis, particularly when using short-read sequencing technology where the primers can anneal to both mtDNA and the similar or identical NuMT sequence. This can lead to the impression of low-level mtDNA heteroplasmy, where the minor allele is actually a NuMT variant^{131,132}. Second, they have been proposed as an explanation for the apparent paternal transmission of mtDNA recently described in three families of Chinese origin¹³³. Recent whole genome sequencing has revealed the presence of rare NuMTs transmitted from fathers which resemble paternally inherited mtDNA on short read sequencing¹³⁴. Long-read sequencing has shown concatenated mtDNA sequences within the nuclear genome that can explain why the NUMT allele frequencies can reach high (>25%) levels. Figure adapted from¹³⁵.

Figure 3 | **How mitochondrial DNA heteroplasmy levels change.** Mitochondrial DNA (mtDNA) heteroplasmy levels can change during life through vegetative segregation in dividing cells, and relaxed replication in non-dividing and dividing cells. In non dividing (post mitotic) cells, relaxed replication during life can lead to an increase in the amount of mutant mtDNA, which if it exceeds a critical threshold, leads to a biochemical defect of oxidative phosphorylation. The germ line genetic bottleneck accelerates segregation between the generations, leading to major changes in the heteroplasmy level. In both contexts, high levels of a pathogenic mutation can cause a cellular biochemical defect of ATP synthesis, but there is emerging evidence that even very low level heteroplasmy levels can have physiological consequences. Heteroplasmy can also be introduced through somatic and germ-line *de novo* mutations. Figure adapted from ¹³⁶.

Figure 5 | **The role of mtDNA in human diseases.** Low level mtDNA heteroplasmy is common and the level increases during life, potentially contributing to common late-onset diseases. Intermediate level heteroplasmy is uncommon, and associated with severe mitochondrial diseases that are rare in the population. Homoplasmic mtDNA variants can be either phenotypically neutral, or contribute to common multifactorial diseases in combination with other genetic and environmental factors. Some classical pathogenic mutations sit in between 'pathogenic mutations' and 'homoplasmic polymorphisms'. For example, Leber Hereditary Optic Neuropathy (LHON) mutations are homoplasmic and found in ~1/300 of the population⁵⁸ but cause disease in the context of a specific mtDNA background, in response to environmental toxins (cigarette smoking and heavy alcohol use), and in concert with predisposing nuclear genetic variants (although these have yet to be defined). Figure adapted from¹³⁵.

Box 1 | Is mtDNA methylated?

The debate as to whether mtDNA is methylated has resurfaced recently, particularly in cancer biology. The presence of both 5-methylcytosine and N⁶-methyldeoxyadenosine have been implicated to be present in mtDNA, though the functional relevance, or even confidence that methylation of mtDNA is truly present, remains unclear. A few studies have shown that manipulations of known nuclear DNA cytosine-5-methyltransferase enzymes DNMT3A or a DNMT1 isoform^{137,138} lead to alteration in the methylation patterns of mtDNA sequences, usually measured with indirect detection methods such as bisulfide conversion or methyl-sensitive restriction assays. A more systematic study sought to uncover the possible consequences of methylation on mtDNA function by transgenic mitochondrial-targeting of viral M.CviPI (a GpC methyltransferase) or bacterial M.SssI (a CpG methyltransferase) to the mitochondria. They reported a clear increase in mtDNA methylation levels, but only GpC methylation was reported to alter mitochondrial gene expression in cells, but did not alter mtDNA copy number in their study¹³⁹. These results appear contradictory to another study in cell culture, where mtDNA methylation was simply used as a marker in protection assays to observe protein binding on the mtDNA¹⁴⁰. However, a recent study of neurodegenerative syndromes associated with DNMT1 mutations failed to detect mitochondrial matrix localization, and instead showed that mitochondria were indirectly deregulated via AMPK-dependent/mTORC1 signalling. In support of this finding, both the Mitocarta compendium of mitochondrial proteins¹⁴¹ and a recent BioID analysis failed to find any enrichment of DNMT1 or DNMT3A in the mitochondrial matrix (BioRxiv doi:<https://doi.org/10.1101/2020.04.01.020479>). From a technical perspective, Mechta et al. describe that with bisulfide detection, the topology of compacted mtDNA lead to an aberrant signal that can be misinterpreted as 5-methylcytosine protection in the bisulfide conversion assay. When enriched mtDNA was linearized with a restriction enzyme, the methylation signal all but vanished, leading them to conclude many estimates of methylation signal may be technical artifacts. Similarly, a technical debate over detection underlines the claims regarding the presence or absence of N⁶-methyldeoxyadenosine in mtDNA¹⁴²⁻¹⁴⁴. Further complicating these analyses, NuMTs (Figure 2) from heterochromatic regions may carry 5-methylcytosine, and their inadvertent amplification — especially in blood cells with low mtDNA copy number — may also contribute to the false attribution of methylation to mtDNA. Even standard density gradient preparations of mtDNA maintain nuclear DNA fragments¹⁴⁵, necessitating the use of highly pure mitochondria and harsh treatment with DNase before DNA extraction^{19,146}. Long-read sequencing platforms such as PacBio and Oxford Nanopore have long claimed to be able to detect modified DNA bases with their sequencing platforms. Hopefully such direct detection methods on highly purified mtDNA will finally resolve these conflicting reports.

Box 2 | Pitfalls of deep re-sequencing of mtDNA

Deep re-sequencing of mtDNA (reviewed previously¹⁴⁷) has revealed the remarkable extent of mtDNA heteroplasmy in humans, but the data needs to be carefully interpreted¹⁴⁸⁻¹⁵⁰. In addition to the risks of cross-contamination, sequencing errors can lead to artifacts which may be indistinguishable from mtDNA mutations occurring *in vivo*, particularly when DNA polymerase-based amplification methods are used. The amplification of nuclear mitochondrial sequences (NuMTs) can also create the impression of heteroplasmy (see Figure 2), particularly with short-read sequencing techniques. When NuMT sequencing reads align to the mtDNA reference sequence this leads to a false positive ‘heteroplasmy’ call¹³¹. Conversely, when a true mtDNA read aligns to the NuMT sequence, this leads to a false negative call which can lower the true heteroplasmy level¹³². Molecular and bioinformatic approaches have been developed to minimise the impact of NuMTs, but it is difficult to absolutely exclude their impact. Different approaches include the pre-amplification of mtDNA with long-range PCR (which can itself introduce polymerase errors), stringent variant quality calling, and only analysing heteroplasmies above a conservative threshold (eg. >1%)^{58,151,152}.

Box 3 | Current strategies for treating mtDNA heteroplasmy

Several strategies have been developed to treat heteroplasmy. Gene-based approaches involve addition or replacement with wild-type mtDNA, or targeting of specific mutations. Adding wild-type mtDNA can reduce the overall percentage of mutant genomes to below the threshold level, correcting the biochemical defect. This approach has been used in skeletal muscle enrolling muscle stem cells (called satellite cells), which typically have a low mutation level, either using muscle-toxic drugs or with exercise training. Drug-induced satellite cell activation has been shown to work in experimental medicine studies^{153,154}, but has not been shown to have clinical benefits in patients. Although exercise training has been shown to improve stamina and reduce fatigue in patients, this was not associated with any change in the muscle heteroplasmy level. mtDNA replacement has also been developed to prevent the inheritance of mtDNA diseases through pronuclear or chromosome spindle transfer in oocytes or embryos^{90,110}. Nuclease targeting to degrade mtDNA mutations show promise *in vivo* and in pre-clinical animal studies^{155,156}, but have not been tested in patients. The recent report of mtDNA base editing by a bacterial cytidine deaminase¹⁵⁷ opens up new therapeutic opportunities, but requires considerable pre-clinical development and may only be possible for a small number of pathogenic mutations. The allotopic expression of mitochondrial proteins in the cell nucleus has also been proposed as a therapeutic strategy for mtDNA diseases, with several gene therapy studies underway against Leber hereditary optic neuropathy (LHON) (gensight.biologics.com). If these prove to be successful, then a similar approach could be possible for heteroplasmic protein-coding gene mutations. All the above approaches are only likely to be useful in patients with a single specific

mtDNA mutation, and not the multiple mtDNA mutations associated with late-onset neurodegenerative diseases and ageing.

Several gene-agnostic approaches are also being developed to treat heteroplasmic mtDNA diseases, potentially harnessing endogenous pathways to tackle heteroplasmic mtDNA mutations. One approach involves inducing mitochondrial biogenesis using re-purposed drugs that could rapidly translate into clinical care. The aim here is to increase the amount of wild-type mtDNA within cells, tipping the balance away from the mutant. However, although beneficial as a compensatory mechanism in the short term, mitochondrial proliferation precedes apoptosis in skeletal muscle, and one clinical study raised concerns based on serum biomarkers so patients will need to be carefully monitored for signs of toxicity¹⁵⁸. Recent evidence supports the use of niacin adult-onset mitochondrial myopathy with associated mitochondrial biogenesis¹⁵⁹. Several clinical trials are underway globally using novel agents aimed at promoting mitochondrial biogenesis, which are regularly updated on [ClinicalTrials.Gov](https://clinicaltrials.gov), which includes other strategies aimed at treating mitochondrial disorders that could be beneficial for heteroplasmic mtDNA diseases.