Article

Thermodynamic analysis of DNA hybridization signatures near mitochondrial DNA deletion breakpoints

Lakshmi Narayanan Lakshmanan, Zhuangli Yee, Barry Halliwell, Jan Gruber, Rudiyanto Gunawan

rgunawan@buffalo.edu

HIGHLIGHTS

We analyzed DNA hybridization signatures of 9921 mtDNA deletions in 5 species

A large fraction of mtDNA deletions exhibit short ≤ 5-nt misalignment signatures

Our results point to erroneous non-homologous DNA end-joining repair in mutagenesis

The agreement of results across species suggests similar deletion mutagenesis

Article

Thermodynamic analysis of DNA hybridization signatures near mitochondrial DNA deletion breakpoints

Lakshmi Narayanan Lakshmanan,[1,2] Zhuangli Yee,[3] Barry Halliwell,[4] Jan Gruber,[4,5] and Rudiyanto Gunawan[6,7,*]

SUMMARY

Broad evidence in the literature supports double-strand breaks (DSBs) as initiators of mitochondrial DNA (mtDNA) deletion mutations. While DNA misalignment during DSB repair is commonly proposed as the mechanism by which DSBs cause deletion mutations, details such as the specific DNA repair errors are still lacking. Here, we used DNA hybridization thermodynamics to infer the sequence lengths of mtDNA misalignments that are associated with mtDNA deletions. We gathered and analyzed 9,921 previously reported mtDNA deletion breakpoints in human, rhesus monkey, mouse, rat, and *Caenorhabditis elegans*. Our analysis shows that a large fraction of mtDNA breakpoint positions can be explained by the thermodynamics of short ≤ 5-nt misalignments. The significance of short DNA misalignments supports an important role for erroneous non-homologous and micro-homology-dependent DSB repair in mtDNA deletion formation. The consistency of the results of our analysis across species further suggests a shared mode of mtDNA deletion mutagenesis.

INTRODUCTION

Pathogenic deletion mutations of mitochondrial DNA (mtDNA) typically involve the loss of several hundreds to thousands of nucleotides (nts) encoding key proteins involved in the mitochondrial electron transport chain and ATP synthesis. The accumulation of such mutant mtDNA molecules in a cell leads to mitochondrial dysfunction and ultimately cellular energy crisis. In humans, inherited mtDNA deletion mutations cause mitochondrial diseases such as Kearns-Sayre syndrome (KSS) (Holt et al., 1989; Moraes et al., 1989), chronic progressive external ophthalmoplegia (CPEO) (Holt et al., 1989; Moraes et al., 1989), and Pearson syndrome (Cormier et al., 1990; Rotig et al., 1995), and sporadic mtDNA deletions have been implicated in age-related diseases such as sarcopenia (Aiken et al., 2002). mtDNA deletions have been reported in tissues of patients suffering from Parkinson disease (Bender et al., 2006), epilepsy (Volmering et al., 2016), inclusion body myositis (Rygiel et al., 2016), Charcot-Marie-Tooth disease (Vielhaber et al., 2013), diabetes (Ballinger et al., 1992), and cancer (Zhu et al., 2004), but their role in these diseases is not clearly defined.

The mechanism of mtDNA deletion formation has been a subject of great interest and debate (Guo et al., 2010; Krishnan et al., 2008). A commonly reported feature of mtDNA deletions is the presence of direct repeat (DR) sequences precisely flanking the deletion breakpoints (Lakshmanan et al., 2012, 2015; Mita et al., 1990; Samuels et al., 2004). This observation has led to two major models of mtDNA deletion formation, namely the "slip-strand" model and the "double-strand break (DSB) repair error" model. The slip-strand model hypothesizes that mtDNA deletions occur during mtDNA replications, more specifically involving the strand-displacement mode of replication (Shoffner et al., 1989). In the strand-displacement replication, the synthesis of daughter heavy and light mtDNA strands is not synchronized. Here, the parental heavy strand is displaced and remains single stranded until roughly 65% of the daughter heavy strand is synthesized. Such a single-stranded heavy strand is prone to misalign (i.e. to hybridize with DNA sequences at an inappropriate position) with the exposed light strand, an error that involves homologous sequences such as DRs, leading to the formation of a deletion mutation (Shoffner et al., 1989). But, the slip-strand model has been questioned (Krishnan et al., 2008). One of the main objections is that this model requires the presence of "naked" single-stranded mtDNA segments during replication. However, studies on mtDNA replication have demonstrated that the displaced heavy strand is protected by

[1]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[2]Institute for Chemical and Bioengineering, ETH Zurich, Zurich, Switzerland

[3]Department of Biological Sciences, Faculty of Science, National University of Singapore, Singapore, Singapore

[4]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[5]Ageing Research Laboratory, Science Division, Yale-NUS College, Singapore, Singapore

[6]Department of Chemical and Biological Engineering, University at Buffalo, Buffalo, NY, USA

[7]Lead contact

*Correspondence:
rgunawan@buffalo.edu

https://doi.org/10.1016/j.isci.2021.102138

mitochondrial single-strand binding proteins and often bound by RNA molecules and therefore not as open as previously thought (Holt and Reyes, 2012).

Meanwhile, the DSB repair error model, as the name suggests, attributes mtDNA deletions to erroneous repair of DSBs. Direct evidence for DSBs as an initiating step in mtDNA deletion formation comes from transgenic mouse models expressing mitochondria-targeted restriction endonuclease (Bacman et al., 2009; Fukui and Moraes, 2009; Srivastava and Moraes, 2005). Mitochondrial expression of these mtDNA targeting enzymes causes DSBs specifically at the targeted sequence locations (restriction sites) in the mtDNA. In mouse models, induction of DSBs by this method reliably promotes the formation of mtDNA deletions with breakpoints near the restriction sites. Also, mitochondrially targeted transcription activator-like effector nucleases (mito-TALEN) induced strand breaks at specific positions in human mtDNA have been shown to promote the occurrence of human mtDNA common deletion through a replisome-dependent mechanism (Phillips et al., 2016). Endogenous and exogenous DNA damaging agents that induce or promote DSBs, such as oxidative stress and ionizing radiation, have also been shown to increase the occurrence of mtDNA deletions (Ji et al., 2006; Volmering et al., 2016). Furthermore, there exists broad experimental evidence that mtDNA replication fork stalling, the collapse of which results in DSBs, promotes formation of mtDNA deletions (see, for example, Wanrooij et al. (Wanrooij and Falkenberg, 2010) and references therein). In agreement with this observation, mtDNA sequences in the neighborhood of deletion breakpoints have been shown to be enriched with sequence motifs that can promote mtDNA replication stalling, such as homopolymeric repeats (Song et al., 2003; Wanrooij et al., 2004), G-quadruplexes (Bharti et al., 2014; Dong et al., 2014), and stem loops (Damas et al., 2012).

Despite the broad evidence for an involvement of DSBs in initiating mtDNA deletions, the mechanism by which DSBs cause mtDNA deletions has yet to be established. In nuclear DNA, DSBs activate one of several DNA repair pathways, including non-homologous end joining (NHEJ), micro-homology-mediated end joining (MMEJ or alternative NHEJ), single-strand annealing (SSA), homologous recombination (HR), synthesis-dependent strand annealing, and break-induced replication (BIR) (Rodgers and McVey, 2016). *Ex vivo* experiments using mitochondrial extracts have demonstrated to different degrees of confidence the existence of mitochondrial NHEJ (Coffey and Campbell, 2000; Coffey et al., 1999; Lakshmipathy and Campbell, 1999), MMEJ (Tadi et al., 2016), and HR repair pathways (Thyagarajan et al., 1996). Mitochondrial BIR activity has been previously reported for yeast (Ling et al., 2013), but the activity of this repair pathway in mammalian mitochondria has not yet been demonstrated. One of the key features differentiating DSB repair pathways is the length of a homologous sequence (HS) around the DSB sites involved in initiating the strand annealing step of the repair process (Kent et al., 2015; Kostyrko and Mermod, 2016; McVey and Lee, 2008; Sakofsky et al., 2015). The NHEJ pathway relies on very short HSs of 0-5 nts, the MMEJ uses 0–25-nt homology (Kent et al., 2015), and the BIR pathway generally requires short (micro) HSs of at least 5-25 nts (Sakofsky et al., 2015). On the other hand, SSA and HR both utilize and require longer HSs of $\geq$30 nts and >100 nts, respectively (Kent et al., 2015; Kostyrko and Mermod, 2016; McVey and Lee, 2008). In addition, HR requires a template, for example, from a sister chromatid, to repair DSBs. Except for HR, errors in each of the above DSB repair pathways have been shown to cause deletion mutations in nuclear DNA (Rodgers and McVey, 2016). Therefore, in the context of the DSB repair, the length of the HSs associated with mtDNA deletions can provide both discriminating and incriminating information to deduce that a specific DSB repair pathway(s) is involved in the deletion mutagenesis (Hartlerode et al., 2016).

The DRs flanking mtDNA deletion breakpoints of an mtDNA deletion are believed to be the HS involved in the formation of that deletion. mtDNA deletions without such flanking DRs are less common and have been suggested to arise through non-homologous mechanisms (Mita et al., 1990). However, other studies have also proposed 50-100-nt sequences spanning multiple DRs around mtDNA breakpoint positions to be the alternative HS in the mtDNA deletion formation (Guo et al., 2010; Sadikovic et al., 2010). In reality, the size of mtDNA misalignments may vary among deletions, conditions, and species, i.e., there may exist an intrinsic heterogeneity in the lengths of HS involved in the deletion mutagenesis. As the aforementioned studies used only one specific sequence length (i.e. 50 or 100 nts) for studying the HS near deletion breakpoints, they were unable to provide any indication of such heterogeneity.

For this study, we developed an analytical method for identifying the distribution of mtDNA sequence lengths that are associated with a given set of mtDNA deletions. We relied on DNA hybridization thermodynamics to compute the probability of differently sized misalignments between 0 and 100 nt to occur

within a range of mtDNA sequences around each mtDNA deletion breakpoint. We employed a mixture distribution model to identify the optimal distribution of mtDNA misalignment lengths that maximizes the probability of mtDNA misalignments at the observed breakpoint positions. We applied our analysis to 9,921 mtDNA deletions gathered from literature, by far the largest reported compendium, for humans, monkeys, mice, rats, and nematodes. Finally, we carried out a detailed comparative analysis of the misalignment length distributions associated with mtDNA deletions from different species and offered two scenarios by which mtDNA deletions arise from DSB repair errors.

RESULTS

Characterizing the probability of DNA hybridization near breakpoints

Using a human mtDNA deletion breakpoint at the nt position pair (6329:13993) as an example, we illustrate our analytical method (Figure 1). For any given mtDNA deletion, we first extract a 201-nt L-strand sequence centered on the 5′ breakpoint position and a 201-nt H-strand sequence centered on the 3′ breakpoint position. These two 201-nt sequences form the 201 nt × 201 nt, two-dimensional window of analysis centered on the mtDNA deletion breakpoint position. For the example mtDNA deletion, this window comprises L-strand sequence between 6229 and 6429 and H-strand sequence between 13,893 and 14,093, as illustrated in Figure 1. For each nt position pair in the window of analysis (i.e. one nt position from L-strand and one nt position from H-strand), we identify DNA duplexes of length l-nt segments—one l-nt segment from the L-strand and one l-nt segment from the H-strand—that overlap the position pair. As depicted in Figure 1B, there are $l \times l$ or 5 × 5 = 25 duplexes overlapping the nt position pair 6326-C (in L-strand) and 13993-G (in H-strand). For each of the $l \times l$ duplexes, we then compute the partition function of DNA hybridization using nearest-neighbor thermodynamics (see Methods). The partition function reflects the thermodynamic stability of duplex configurations that can form between two DNA sequences. We finally assign a propensity value for each position pair in the window of analysis. This value is equal to the sum of the partition functions over all $l \times l$ duplexes. The propensity value of a position pair is therefore indicative of the probability of the corresponding nt pair to take part in an l-nt long misalignment.

In the next step, we divide the window of analysis into 10 × 10-nt bins. The binning strategy is implemented to address two potential issues: (1) imprecision in the determination of the breakpoint position (e.g. due to presence of DRs) and (2) the deletion breakpoint occurring adjacent to—not within—the DNA misalignment. The misalignment propensity value for each bin is set to the sum of the propensity values of all nt position pairs in the bin. The heatmap in Figure 1C illustrates the distribution of the bin propensity values in the window of analysis for 5-nt misalignment length for the example human mtDNA deletion. Such variation is a consequence of the differences in the thermal stability of l-nt misalignments within the window of analysis (l = 5 in Figure 1C). If a given mtDNA deletion arises from a DNA misalignment of length l-nt, then we expect the bin containing the deletion breakpoint to be more thermally stable and thus to have a higher propensity value relative to the other bins in the window of analysis. Note that the hybridization partition function and thus the propensity value depend on the sequence length l.

Figure 1D depicts the relative bin propensity for the mtDNA deletion breakpoint position, i.e., the bin propensity value containing the deletion breakpoint divided by the sum of all bin propensity values in the window of analysis, as a function of the misalignment sequence length l. For the example mtDNA deletion, the relative bin propensity value peaks at 20 nts. While the example mtDNA deletion has a 5-nt flanking DR (see Figure 1A), our analysis suggest that this deletion could occur due to the favorable energetics of a longer misalignment of 20 nts, which includes the 5-nt DR but is not limited to it.

In this study, we tested nine sequence lengths ranging from 0 nt to 100 nts, specifically 0, 5, 10, 15, 20, 25, 50, 75, and 100 nts. These values were chosen to cover the range of misalignment lengths commonly believed to be involved in mtDNA deletions. The 0-nt length accounts for a hybridization-independent formation of mtDNA deletions, for which the (bin) propensity values are set to the same value (i.e. the uniform distribution). For higher computational efficiency, we pre-computed the partition functions of DNA hybridization for each of the sequence lengths mentioned above and for mtDNA sequences containing the majority of the reported mtDNA deletions (e.g., the entire major arc sequence of human, rhesus monkey, mouse, and rat mtDNA).

We posit that the length of DNA misalignment associated with an mtDNA deletion corresponds to the sequence length that maximizes the misalignment propensity at the mtDNA breakpoint position relative
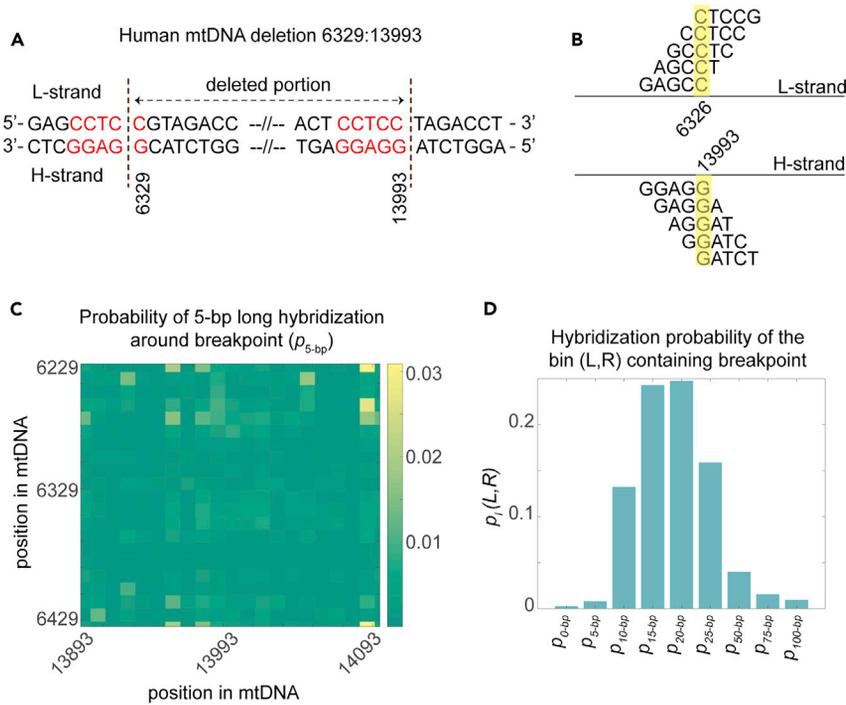
Figure 1. Misalignment analysis of mtDNA deletions

An illustration of misalignment analysis using an example human mtDNA deletion.

(A) The human mtDNA deletion has a breakpoint position pair of (6329:13,993) with a 5-nt long direct repeat flanking the breakpoint (shown in red).

(B) A total of 5 × 5 = 25 distinct pairs of l = 5-nt sequences can form DNA duplexes that overlap the position pair (6326, 13,993) (highlighted in yellow). One of the 5-nt sequences comes from the L-strand sequence and the other from the H-strand sequence.

(C) Distribution of bin propensity values for l = 5-nt within the 201 × 201-nt window of analysis. The color represents the relative propensity values, i.e., the propensity value of each bin divided by the sum of all bin propensity values in the window.

(D) Relative bin propensity value containing the breakpoint position as a function of sequence lengths used in the analysis.

(E) In the mixture model, the overall likelihood for each mtDNA deletion is computed as a weighted sum of relative bin propensity values of the mtDNA deletion breakpoint over the sequence lengths between 0 and 100 nts in the analysis.

(F) For each data set, the unknown weights are calculated by maximizing the total likelihood for a given set of mtDNA deletions.

to those at the surrounding positions in the window of analysis. We employ a mixture distribution model for evaluating the relative bin propensity of a given mtDNA deletion (see Figure 1E). In this model, the relative propensity is computed as a weighted sum of the relative propensities for different misalignment lengths

Table 1. Breakpoint data sets

Data set count	Data set name	Total deletions	Source/description
1	Human MTL epilepsy	208	Mesial temporal lobe epilepsy
2	Human aging	151	Aging-associated deletions
3	Single deletion myopathy	138	Non-autosomal KSS/CPEO
4	Human POLG	85	Autosomal CPEO with POLG gene mutations
5	Inclusion body myositis	49	Inflammation in muscle
6	Pearson syndrome	34	Rare childhood disorder
7	Charcot-Marie-Tooth disease	31	Inherited nervous disorder
8	Rat aging	121	Aging-associated deletions
9	Mouse aging	63	Aging-associated deletions
10	DSB mouse	57	Mice with transgenic Pst1 and Sca1 expression
11	WT mouse (NGS)	707	Wild-type control mice
12	$Sod2^{+/-}$ mouse (NGS)	2,169	Mice heterozygous for superoxide dismutase ($Sod2^{+/-}$)
13	$TWNK^{+}$ (NGS)	3,455	Wild-type mice with TWINKLE overexpression
14	$Sod2^{+/-}$; $TWNK^{+}$ (NGS)	2,355	$Sod2^{+/-}$ mice with TWINKLE overexpression
15	Monkey aging	32	Aging-associated deletions
16	*C. elegans* aging	266	Aging-associated deletions
	Total	9,921	

used in the analysis. The weights, denoted by θ_l (l = 0, 5, 10, 15, 20, 25, 50, 75, and 100 nts), are the unknown model parameters of interest. Here, we formulate a maximum likelihood estimation to obtain the unknown weights of the mixture distribution model for a group of mtDNA deletions (for example, mtDNA deletions in human aging) (see Figure 1F). The likelihood of an mtDNA deletion to occur is set to the logarithm of the relative propensity from the mixture distribution model, given the weights θ_l. We used the interior point algorithm to obtain the optimal weights that maximize the total likelihood (MATLAB, version 2015a; Math-Works, Inc.). For a given group of mtDNA deletions, the optimal weight θ_l^* represents the proportion of mtDNA deletions in the group that are associated with the specific length l. We tested and validated our analysis using *in-silico*-generated mtDNA deletion data sets with known θ_l^* by random sampling (see Methods). Our analytical method accurately recovered the fractions of mtDNA deletions that were used to generate the *in silico* breakpoints (see Figure S1; Supplemental information). Importantly, we were able to confirm that our findings are not sensitive to the analysis window length and bin size values used for hybridization probability calculations (see Figure S2; Supplemental information).

For our analysis, we compiled an exhaustive compendium of 9,921 mtDNA deletion breakpoint positions previously reported in human, rhesus monkey, mouse, rat, and *C. elegans* and categorized these deletions into 16 groups based on the species and the experimental conditions (patient groups used, mutant strains, treatment conditions; for more details, see Table 1, Table S1; Supplemental information, and Data S1). This compendium comprised, to the best of our knowledge, the biggest data set publicly available at the time of our study.

Hybridization signatures near deletion breakpoints in rodents and nematodes

The first class of mtDNA deletions in our analysis was that reported in transgenic mice expressing mitochondria-targeted restriction enzymes PstI and ScaI (see Figure 2A) (Bacman et al., 2009; Fukui and Moraes, 2009; Srivastava and Moraes, 2005). Overexpression of restriction endonucleases introduces DSBs at specific restriction sites in the mtDNA sequence and results in mtDNA deletions typically with one or both breakpoints located in the proximity to these restriction sites. Since this class of deletions derived from DSBs, the mtDNA deletion breakpoints represent the ideal case to investigate the misalignment length signature associated with DSB-induced mtDNA deletions. The outcome of our analysis as shown in Figure 2A shows a significant fraction of short misalignments (0–5 nts: 84% $\pm$ 8% [mean $\pm$ standard deviation (s.d.)]), a minor contribution from medium length (10–25 nts: 16% $\pm$ 7%), and practically no participation by longer misalignments ($\geq$50 nts: 0% $\pm$ 4%). The significance of short misalignments is indicative of a
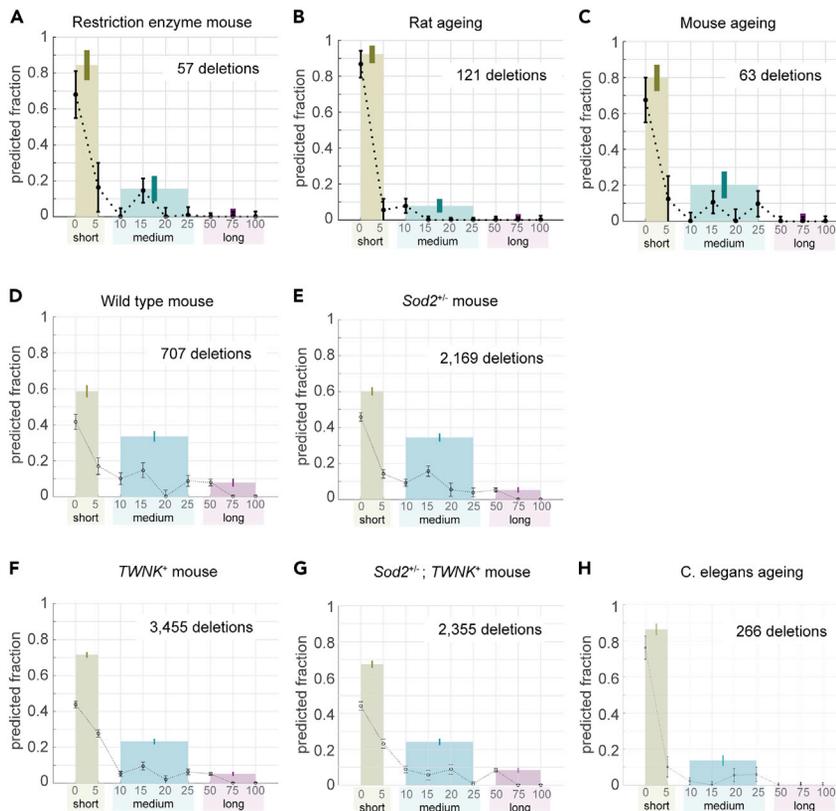
Figure 2. Analysis of mtDNA deletion breakpoints from rodents and nematodes

Mixture model coefficients for (A) transgenic mice with mitochondrial restriction enzymes, (B) rat aging, (C) mouse aging, (D) wild-type mice (14 weeks old), (E) $Sod2^{+/-}$ mice, (F) wild-type mice with TWINKLE overexpression, (G) $Sod2^{+/-}$ mice with TWINKLE overexpression, and (H) *C. elegans* aging. The colored bars show the sum of the coefficients for short (0–5 nts), medium (10–25 nts), and long ($\geq$ 50 nts) duplexes. The error bars indicate the sample standard deviation estimated using 100 *in silico* data sets.

central role of non-homologous and micro-homology-dependent repair mechanisms in the creation of this class of mtDNA deletions. On the other hand, the absence of contribution from long misalignments signifies a lack of involvement from HR repair pathway in this class of deletions.

We subsequently analyzed three different groups of mtDNA deletions found in wild-type (WT) rats and mice without any transgenic expression or diseases. The rat data set (Figure 2B) and one of the mouse data sets (Figure 2C) comprised mtDNA deletion breakpoints gathered from different reports in the literature. The second and larger mouse data set (Figure 2D) included mtDNA deletion breakpoints from a single next-generation sequencing (NGS)-based study (Pohjoismaki et al., 2013). The results of the analysis of these data sets suggest misalignment length compositions that closely resemble those of the DSB-induced mtDNA deletion data set above. Again, the largest fraction of deletions is associated with short misalignments (rat aging: 92% $\pm$ 5%, mouse aging: 80% $\pm$ 7%, WT mouse: 59% $\pm$ 3%), with the medium lengths having a smaller contribution (rat aging: 8% $\pm$ 4%, mouse aging: 20% $\pm$ 7%, WT mouse: 34% $\pm$ 3%) and the long misalignments contributing the least (rat aging: 0% $\pm$ 3%, mouse aging: 0% $\pm$ 4%, WT mouse: 8% $\pm$ 2%). The similarity in the misalignment length signature to DSB-induced mtDNA deletions supports an important role of DSBs in the formation of naturally occurring mtDNA deletions in WT mice and rats.

The aforementioned NGS study also generated mtDNA deletion breakpoints for mice heterozygous for *sod2* gene expression. Mitochondrial superoxide dismutase (Sod2) converts superoxide anions ($O_2^{\cdot-}$) generated by electron transport chain into hydrogen peroxide. $Sod2^{-/-}$ mice experience early postnatal lethality (Li et al., 1995; Loch et al., 2009). Meanwhile, heterozygous $Sod2^{+/-}$ mice suffer increased mitochondrial oxidative damage, higher burden of mtDNA point mutations, more frequent mtDNA replication stalling,

and elevated levels of recombined mtDNA molecules carrying deletions (Pohjoismaki et al., 2013). The breakpoint hotspots in $Sod2^{+/-}$ also coincide with G-rich regions in the mtDNA H-strand, suggesting increased oxidative damage (8-oxoG) as the causative factor of higher levels of mtDNA deletions in the $Sod2^{+/-}$ mice (Pohjoismaki et al., 2013). Our analysis of mtDNA deletion breakpoints in these $Sod2^{+/-}$ mice (Figure 2E) suggests a misalignment length distribution that is similar to that of WT mice (Figures 2C and 2D). A comparison of the deletion breakpoint positions further shows that the mtDNA deletion hotspot regions in $Sod2^{+/-}$ mice coincides with those in WT mice (see Figure S3; Supplemental information). The similarity in the misalignment length compositions and the deletion hotspots between WT and $Sod2^{+/-}$ mice suggests that mtDNA deletions induced by increased oxidative damage and those in WT mice occur by similar mutagenesis mechanisms.

The same NGS study also demonstrated that the overexpression of TWINKLE (TWNK) helicase in $Sod2^{+/-}$ mice ($Sod2^{+/-}$;$TWNK^+$) was able to rescue the oxidative stress phenotypes of $Sod2^{+/-}$ (Pohjoismaki et al., 2013). While the frequency of mtDNA rearrangements, including deletions, in $Sod2^{+/-}$;$TWNK^+$ mice is lower than that in $Sod2^{+/-}$ mice, it remains elevated in comparison to WT. Furthermore, TWINKLE overexpression in WT mice ($TWNK^+$) leads to an increased level of mtDNA rearrangements (see Figure S6 in the original publication [Pohjoismaki et al., 2013]). In addition to its putative function as a helicase, TWINKLE has previously been shown to catalyze DNA recombination (Sen et al., 2012, 2016). The observed increased frequency of mtDNA rearrangements in $TWNK^+$ mice may thus arise from higher DNA recombination activity due to TWINKLE overexpression. The misalignment length compositions of mtDNA deletions in $TWNK^+$ and $Sod2^{+/-}$; $TWNK^+$ mice mirror each other (see Figures 2F and 2G), implying similar mechanisms of mutagenesis. In other words, TWINKLE overexpression appears to lower the frequency of de novo mutations but does not significantly alter the mechanism by which mtDNA deletions form in mice. But, in comparison to WT and $Sod2^{+/-}$ mice, mtDNA deletions of $TWNK^+$ and $Sod2^{+/-}$; $TWNK^+$ mice have higher fractions of the short length misalignments, specifically 5 nts (WT: 59% $\pm$ 3%, $Sod2^{+/-}$: 60% $\pm$ 2%, $TWNK^+$: 72% $\pm$ 1%, $Sod2^{+/-}$; $TWNK^+$: 67% $\pm$ 2%). A recent study demonstrated that the strand exchange activity of TWINKLE requires 3-6-nt sequence homology (Sen et al., 2016). Thus, our analysis is sensitive enough to detect the higher frequency of short 5-nt misalignment that is expected from the overexpression of TWINKLE in the above mice.

Besides rodents, we analyzed the age-related mtDNA deletion data sets recently reported for nematode C. elegans (Lakshmanan et al., 2018). Like mammals, mtDNA deletions in C. elegans are also frequently flanked by DR motifs (Lakshmanan et al., 2018; Melov et al., 1994). As depicted in Figure 2H, the misalignment length fractions for mtDNA deletions in nematodes (0–5 nts: 86% $\pm$ 3%, 10–25 nts: 14% $\pm$ 3%, $\geq$50 nts: 0% $\pm$ 2%) closely resemble the findings from the analyses of rodent data sets, indicating a conserved mechanism of mtDNA deletion mutagenesis between rodents and nematodes, two species evolutionarily separated by almost a billion years. While short-lived animal models are commonly used to understand the mitochondrial involvement in aging, significant differences exist between short-lived models and humans. For example, recently we have demonstrated that mtDNA deletion clonal expansion with age, a key process implicated in human sarcopenia, is undetectable in C. elegans (Lakshmanan et al., 2018). It would be interesting to compare whether the heterogeneity in misalignment sequence lengths observed near mtDNA deletions in short-lived animal models is comparable to that of human mtDNA deletions.

Misalignment signatures near breakpoints in primates

Our analysis of mtDNA deletions in aged rhesus monkey and human shows misalignment length compositions that are similar to those of rodent mtDNA deletions, albeit with nominally higher fractions of medium length misalignments (see Figures 3A and 3B; rhesus monkey: 40% $\pm$ 11%, human: 45% $\pm$ 5% vs. rat: 8% $\pm$ 4%, mouse: 20% $\pm$ 7%). In addition to mtDNA deletions associated with aging, inherited mtDNA deletions in human have also been commonly associated with autosomal myopathies and several other neuromuscular and multisystem disorders. We compiled human mtDNA deletions reported for patients with neuromuscular disorders from the literature and categorized them into 6 disease groups based on the patient description in the source articles (see Methods). Except for mtDNA deletions associated with polymerase gamma (polg) gene mutations, our analysis of these groups of mtDNA deletions gives compositions of misalignment lengths that are similar to those of age-related mtDNA deletions in human (see Figures 3C–3G). Thus, human mtDNA deletions during aging and in neuromuscular diseases (except autosomal polg mutations dataset) appear to occur by a shared mechanism.
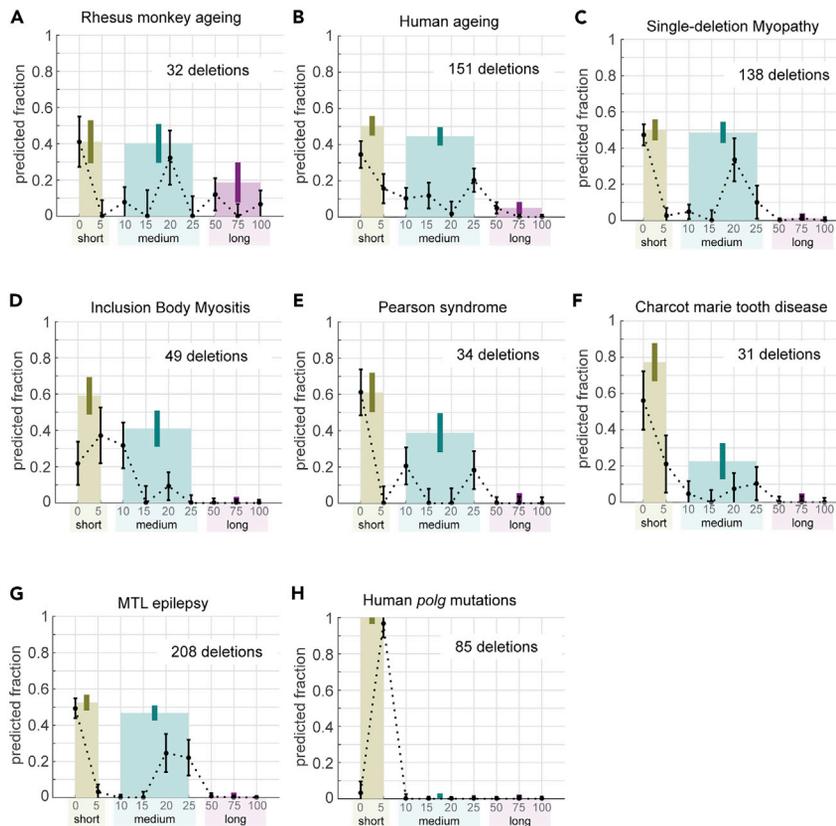
Figure 3. Analysis of mtDNA deletion breakpoints from primates

Mixture model coefficients for (A) rhesus monkey aging, (B) human aging, (C) single deletion myopathy, (D) inclusion body myositis, (E) Pearson syndrome, (F) Charcot-Marie-Tooth disease, (G) mesial temporal lope (MTL) epilepsy, and (H) a patient with myopathy and compound *polg* gene mutations. The colored bars show the sum of the coefficients for short (0–5 nts), medium (10–25 nts), and long ($\geq$ 50 nts) duplexes. The error bars indicate the sample standard deviation estimated using 100 *in silico* data sets (see Methods).

The one exception to this pattern is the autosomal *polg* mutations (see Figure 3H). POLG is the sole DNA polymerase present in mammalian mitochondria. Mutations in *polg* cause defects in the mtDNA replication process, leading to mtDNA depletions and mutations, including deletion and point mutations (Wanrooij et al., 2004, 2007). The associated mtDNA deletions have breakpoints near homopolymeric runs (HPs) and G-quadruplexes (GQs), sequence motifs that are known to induce replication stalling (Bharti et al., 2014; Wanrooij et al., 2004). Our analysis of mtDNA deletion breakpoints from autosomal myopathy patient with compound *polg* mutations indicates that, unlike all other deletion data sets, the mtDNA deletions in these patients are associated exclusively with short HSs (see Figure 3H; 0–5 nts: 100% $\pm$ 3%). The absence of medium and long misalignments in the mutant POLG data set (10–25 nts: 0% $\pm$ 3%, $\geq$ 50 nts: 0% $\pm$ 2%) is in agreement with a previous study demonstrating a role of POLG in misalignment-dependent DSB repair in human (Phillips et al., 2016). More specifically, the study showed that the common 4,977-nt mtDNA deletion in human could be induced artificially by DSBs, through a mechanism that requires functional POLG. In addition, other DNA polymerases, such as Pol-θ, have previously been shown to take part directly in micro-homology (i.e. medium length)-mediated end-joining DSB repair in nuclear DNA (Kent et al., 2015).

In summary, our analyses of mtDNA deletion breakpoints from four mammalian species, including human, and from nematodes (*C. elegans*) point to a shared mtDNA deletion mutagenesis mechanism, involving pre-dominantly DNA misalignments of short and medium HS lengths (0-25 nts). Meanwhile, DNA misalignments of >25 nts appear to have little or no involvement in the formation of mtDNA deletions. Except for PstI and ScaI transgenic mouse data sets, the specific event initiating the formation of mtDNA deletions in these samples remains unknown. However, the similarity of the misalignment length compositions for PstI and ScaI

mice, WT mice and rats, and Sod2$^{+/-}$ mice supports involvement of DSBs and oxidative damage in the deletion mutagenesis in rodents.

DISCUSSION

The formation and clonal expansion of sporadic as well as inherited mtDNA deletions ultimately causes mitochondrial respiratory dysfunction, which in human leads to a class of clinically heterogeneous disorders collectively known as mitochondrial diseases. To date, there is no effective therapeutical treatment for mitochondrial diseases due to the lack of mechanistic understanding of the disease progression. Investigations into how pathogenic mtDNA deletion mutations form and accumulate may lead to the formulation of testable hypotheses and eventually to the identification of disease-modifying targets. Over the past decades, experimental and modeling studies have provided a better understanding of the etiology of mtDNA mutation expansion (Chinnery et al., 2002; Kowald and Kirkwood, 2014; Poovathingal et al., 2009, 2012; Tam et al., 2015). In contrast, the origin and mechanism underlying mtDNA deletions are less well understood. There exists broad empirical evidence from human and model organisms that DSBs are involved in the etiology of mtDNA deletions and that deletions may result from erroneous DSB repair (Krishnan et al., 2008). However, the specific DSB repair pathway(s) involved and the mechanism by which the corresponding errors cause mtDNA deletions have yet to be established. Here, we developed a method that combines DNA hybridization thermodynamics, a mixture distribution model and maximum likelihood estimation, to determine the lengths of DNA misalignments associated with experimentally determined mtDNA deletion breakpoints in diverse species. We used the misalignment lengths as a signature for identifying the specific DSB repair mechanisms involved in mtDNA deletion formation and to examine similarities and differences in the mutagenesis mechanisms of mtDNA deletions from different species and causes.

Our analytical method overcomes some of the limitations in studying mtDNA deletion formation by only looking at DRs in the direct proximity of the reported deletion breakpoints. One limitation is associated with the fact that a large fraction of mtDNA sequences is involved in DRs (Lakshmanan et al., 2012; Samuels et al., 2004). As we have shown in a previous study, the presence of DRs around deletion breakpoints may arise simple by chance and does not unequivocally support the causal role of DRs in mtDNA deletion mutagenesis (Lakshmanan et al., 2012, 2015). In addition, the position of DRs with respect to the breakpoints may be imprecise due to technical errors during polymerase chain reaction (PCR) amplification and sequencing. For the deletion breakpoints considered in this study, many do not have precisely flanking DRs. Further, besides a few well-known deletion breakpoints, such as the 4977-bp mtDNA common deletion in human that is flanked by a 13-bp DR, the majority of the reported deletion breakpoints in the mtDNA major arcs are associated with much shorter DRs. As we reported before, a large majority of short DRs ($\leq$5 bp) in mtDNA have positive free energies for DNA hybridization and hence in principle do not spontaneously form thermally stable duplexes (Lakshmanan et al., 2012). For these reasons, 100-bp imperfect DNA duplexes of sequences around the deletion breakpoints that may include multiple short DRs have previously been postulated to be the causative DNA misalignment involved in mtDNA deletions (Guo et al., 2010). While we also consider such imperfect duplexes, our method accounts for a range of duplex lengths, thus enabling a more precise characterization of the relevant DNA misalignments involved in mtDNA deletion formation.

Our analysis of mtDNA deletions across five species (human, rhesus monkey, rat, mouse, and *C. elegans*) suggests that the deletion breakpoint positions are most consistent with a mutagenesis mechanism that is driven by the thermodynamics of short to medium length DNA misalignments (0-25 nts). The relative contributions from longer misalignments vary slightly among species and conditions but, in general, are low. Assuming that mtDNA deletions indeed arise by errors during DSB repairs, the significance of short and medium DNA misalignments implicates NHEJ and micro-homology-dependent mechanisms (e.g. MMEJ) as key drivers of deletion formation in all species. NHEJ is generally known as the fastest and the most common route of repairing DSBs *in vivo* (Dueva and Iliakis, 2013). Our findings thus point to mitochondrial DSB repair, specifically NHEJ, as an attractive target for therapeutics for preventing or reducing the formation of mtDNA deletions.

At the same time, the small contribution from long misalignments excludes HR as an important source of mtDNA deletions. The lack of HR involvement in mtDNA deletion formation is in good agreement with two well-known aspects of the HR repair in mitochondria. First, HR is generally considered an error-free mechanism as the repair uses a template DNA molecule (Rodgers and McVey, 2016). Second, individual mtDNA

molecules are typically isolated into separate nucleoids, limiting inter-molecular recombination that is necessary for HR repair (Kazak et al., 2012). While our analyses do not exonerate replication slippage as a mechanism of deletion formation, one might expect higher contributions from longer and more thermodynamically stable misalignments if deletions were to occur predominantly by the slip-strand mechanism (Bzymek and Lovett, 2001). Moreover, computational validation experiments show that our findings are not sensitive to the length of the analysis window and the length of the bins used in the analysis (Figure S2; Supplemental information). Also, applying our analysis to mtDNA deletions from individual reports separately produces the same general findings as reported above (see Figure S4; Supplemental information).

In contrast to our finding on the lack of involvement of >25-nt misalignments, a previous study by Guo et al. using mtDNA deletion breakpoints from the human frontal cortex found that the breakpoints are over-represented in regions of mtDNA sequences that can form highly stable 100-nt misalignments (Guo et al., 2010). While the study by Guo et al. and our study were based on DNA hybridization thermodynamics, our analysis differs from the one in the study by Guo et al. in two major aspects: (1) Our analysis accounts for all feasible duplex configurations associated with any given duplex of *l*-nt sequences, whereas Guo et al.'s analysis uses only the duplex configuration corresponding to the minimum free energy value. (2) Importantly, our analysis also takes into consideration DNA misalignments of different lengths, whereas Guo et al.'s analysis uses only pre-selected single length (100-nt) long misalignments. When we repeated the analysis of Guo et al. using the same breakpoints but allowing for shorter misalignment lengths, more specifically 5, 10, 25, 50, and 100 nts, we observed over-representation of deletion breakpoints in regions of mtDNA forming stable duplexes for each of these lengths, suggesting that there is, in fact, no contradiction between the two studies insofar as both methods identify the same sequences (see Figure S5; Supplemental information).

There still exists an obvious inconsistency between the 2–25-bp short deletions commonly caused by NHEJ/MMEJ errors and the 1000s of base-pair-long deletions observed in mtDNA. Despite the rarity of large deletions (>1000bp) resulting from erroneous NHEJ/MMEJ repair, previous studies of nuclear DNA have shown a scenario involving multiple DSBs by which erroneous DSB repairs could cause such deletions (Schwer et al., 2016; Xiong et al., 2015). We therefore posit that multiple DSBs (at least two) on a single mtDNA molecule could initiate the formation of a large deletion, where the DNA ends from two distinct DSBs are misjoined by NHEJ/MMEJ repair. Any two distantly located DSBs on an mtDNA molecule might come into close proximity to each other because of the folding and packing of mtDNA by mitochondrial transcription factor A (TFAM) (Gustafsson et al., 2016). Alternatively, misjoining of distant DSB ends could also occur if the mtDNA sequence in between these ends is degraded. The recombination between two mtDNA molecules, each with a DSB, may also result in a deletion mutation (Bacman et al., 2009). But such events have been shown to take place only very rarely *in vivo* (Bacman et al., 2009) and therefore are unlikely to contribute significantly to mtDNA deletion formation under physiological conditions.

An alternative scenario for lengthy mtDNA deletion mutagenesis is a single DSB invading a distant but accessible open region in mtDNA, similar to the explanation for D-loop hotspot deletions proposed in the study by Srivastava et al. (Srivastava and Moraes, 2005). The immediate question here is whether there exists a possible mitochondrial protein that is able to assist the strand invasion. Recent studies have demonstrated that TWINKLE helicase possesses several DNA modifying capacities including strand invasion (or strand exchange) (Sen et al., 2012, 2016). Consistent with the importance of short to medium HSs in our analyses, the efficiency of the strand exchange assisted by TWINKLE depends strongly on the sequence complementarity of the first 3-6 nts of the invading strand (Sen et al., 2012, 2016). Finally, there could be other, yet uncharacterized, micro-homology-dependent repair and mutagenesis mechanisms in mitochondria. For instance, a recent study reported that a strand break adjacent to the 5' 13-nt DR flanking the common deletion in human mtDNA can induce the formation of the common deletion through a yet-to-be-characterized process that is independent of the canonical NHEJ and MMEJ enzymes (Phillips et al., 2016). Based on *in vitro* replication studies using single-stranded DNA templates with 13-nt DR flanking the common deletion in human mtDNA, another recent study has suggested that mtDNA deletions could occur during mtDNA replication using copy choice recombination process (Persson et al., 2019). Both these recent studies focused primarily on human mtDNA common deletion flanked by 13-nt DR. However, it is not immediately clear whether the remaining vast majority of mtDNA deletions with either short or no DRs could be explained by the mechanisms suggested in these recent studies (Persson et al., 2019; Phillips et al., 2016).

Limitations of the study

The probabilistic mixture model approach proposed in this work is a fundamentally rigorous approach to identify the length of thermodynamically stable misalignments surrounding deletion breakpoints. But the current modeling framework has 2 limitations: (1) it cannot differentiate between two or more deletion mutagenesis mechanisms involving similar misalignment lengths and (2) because of the binning, this approach cannot pinpoint the orientation of misalignments relative to the breakpoints.

Resource availability

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Dr. Rudiyanto Gunawan (rgunawan@buffalo.edu)

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data sets used for this study are provided as Data S1. Codes used for generating result figures can be downloaded from the GitHub webpage: https://github.com/CABSEL/Mitochondrial-DNA-Deletion-Breakpoints-Analysis.

METHODS

All methods can be found in the accompanying Transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102138.

AUTHOR CONTRIBUTIONS

L.N., B.H., J.G., and R.G. conceived and designed the experiments. L.N. and Z.L. performed the experiments. L.N., R.G., and J.G. analyzed the data. L.N., Z.L., R.G., and J.G. contributed reagents/materials/analysis tools. L.N., R.G., J.G., and B.H. wrote the paper. L.N. wrote software used in analysis.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Aiken, J., Bua, E., Cao, Z., Lopez, M., Wanagat, J., Mckenzie, D., and Mckiernan, S. (2002). Mitochondrial DNA deletion mutations and sarcopenia. Ann. N Y Acad. Sci. 959, 412–423.

Bacman, S.R., Williams, S.L., and Moraes, C.T. (2009). Intra- and inter-molecular recombination of mitochondrial DNA after in vivo induction of multiple double-strand breaks. Nucleic Acids Res. 37, 4218–4226.

Ballinger, S.W., Shoffner, J.M., Hedaya, E.V., Trounce, I., Polak, M.A., Koontz, D.A., and Wallace, D.C. (1992). Maternally transmitted diabetes and deafness associated with a 10.4 kb mitochondrial DNA deletion. Nat. Genet. 1, 11–15.

Bender, A., Krishnan, K.J., Morris, C.M., Taylor, G.A., Reeve, A.K., Perry, R.H., Jaros, E., Hersheson, J.S., Betts, J., Klopstock, T., et al. (2006). High levels of mitochondrial DNA

deletions in substantia nigra neurons in aging and Parkinson disease. Nat. Genet. 38, 515–517.

Bharti, S.K., Sommers, J.A., Zhou, J., Kaplan, D.L., Spelbrink, J.N., Mergny, J.-L., and Brosh, R.M. (2014). DNA sequences proximal to human mitochondrial DNA deletion breakpoints prevalent in human disease form G-quadruplexes, a class of DNA Structures Inefficiently Unwound by the mitochondrial

replicative Twinkle helicase. J. Biol. Chem. *289*, 29975–29993.

Bzymek, M., and Lovett, S.T. (2001). Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. Proc. Natl. Acad. Sci. U S A *98*, 8319–8325.

Chinnery, P.F., Samuels, D.C., Elson, J., and Turnbull, D.M. (2002). Accumulation of mitochondrial DNA mutations in ageing, cancer, and mitochondrial disease: is there a common mechanism? Lancet *360*, 1323–1325.

Coffey, G., and Campbell, C. (2000). An alternate form of Ku80 is required for DNA end-binding activity in mammalian mitochondria. Nucleic Acids Res. *28*, 3793–3800.

Coffey, G., Lakshmipathy, U., and Campbell, C. (1999). Mammalian mitochondrial extracts possess DNA end-binding activity. Nucleic Acids Res. *27*, 3348–3354.

Cormier, V., Rötig, A., Quartino, A.R., Forni, G.L., Cerone, R., Maier, M., Saudubray, J.M., and Munnich, A. (1990). Widespread multi-tissue deletions of the mitochondrial genome in the Pearson marrow-pancreas syndrome. J. Pediatr. *117*, 599–602.

Damas, J., Carneiro, J., Gonçalves, J., Stewart, J.B., Samuels, D.C., Amorim, A., and Pereira, F. (2012). Mitochondrial DNA deletions are associated with non-B DNA conformations. Nucleic Acids Res. *40*, 7606–7621.

Dong, D.W., Pereira, F., Barrett, S.P., Kolesar, J.E., Cao, K., Damas, J., Yatsunyk, L.A., Johnson, F.B., and Kaufman, B.A. (2014). Association of G-quadruplex forming sequences with human mtDNA deletion breakpoints. BMC Genomics *15*, 1–15.

Dueva, R., and Iliakis, G. (2013). Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer. Transl. Cancer Res. *2*, 163–177.

Fukui, H., and Moraes, C.T. (2009). Mechanisms of formation and accumulation of mitochondrial DNA deletions in aging neurons. Hum. Mol. Genet. *18*, 1028–1036.

Guo, X., Popadin, K.Y., Markuzon, N., Orlov, Y.L., Kraytsberg, Y., Krishnan, K.J., Zsurka, G., Turnbull, D.M., Kunz, W.S., and Khrapko, K. (2010). Repeats, longevity and the sources of mtDNA deletions: evidence from 'deletional spectra'. Trends Genet. *26*, 340–343.

Gustafsson, C.M., Falkenberg, M., and Larsson, N.G. (2016). Maintenance and expression of mammalian mitochondrial DNA. Annu. Rev. Biochem. *85*, 133–160.

Hartlerode, A.J., Willis, N.A., Rajendran, A., Manis, J.P., and Scully, R. (2016). Complex breakpoints and template Switching associated with non-canonical termination of homologous recombination in mammalian cells. Plos Genet. *12*, e1006410.

Holt, I.J., Harding, A.E., Cooper, J.M., Schapira, A.H., Toscano, A., Clark, J.B., and Morgan-Hughes, J.A. (1989). Mitochondrial myopathies: clinical and biochemical features of 30 patients with major deletions of muscle mitochondrial DNA. Ann. Neurol. *26*, 699–708.

Holt, I.J., and Reyes, A. (2012). Human mitochondrial DNA replication. Cold Spring Harb. Perspect. Biol. *4*, a012971.

Ji, F., Pierre, Z., Liu, S., Hwang, B.-J., Hill, H.Z., Hubbard, K., and Steinberg, M. (2006). Novel mitochondrial deletions in human epithelial cells irradiated with an FS20 ultraviolet light source in vitro. J. Photochem. Photobiol. A Chem. *184*, 340–346.

Kazak, L., Reyes, A., and Holt, I.J. (2012). Minimizing the damage: repair pathways keep mitochondrial DNA intact. Nat. Rev. Mol. Cell Biol. *13*, 659–671.

Kent, T., Chandramouly, G., Mcdevitt, S.M., Ozdemir, A.Y., and Pomerantz, R.T. (2015). Mechanism of microhomology-mediated end-joining promoted by human DNA polymerase theta. Nat. Struct. Mol. Biol. *22*, 230–237.

Kostyrko, K., and Mermod, N. (2016). Assays for DNA double-strand break repair by microhomology-based end-joining repair mechanisms. Nucleic Acids Res. *44*, e56.

Kowald, A., and Kirkwood, T.B.L. (2014). Transcription could be the key to the selection advantage of mitochondrial deletion mutants in aging. Proc. Natl. Acad. Sci. U S A *111*, 2972–2977.

Krishnan, K.J., Reeve, A.K., Samuels, D.C., Chinnery, P.F., Blackwood, J.K., Taylor, R.W., Wanrooij, S., Spelbrink, J.N., Lightowlers, R.N., and Turnbull, D.M. (2008). What causes mitochondrial DNA deletions in human cells? Nat. Genet. *40*, 275–279.

Lakshmanan, L.N., Gruber, J., Halliwell, B., and Gunawan, R. (2012). Role of direct repeat and stem-loop motifs in mtDNA deletions: cause or coincidence? PLoS One *7*, e35271.

Lakshmanan, L.N., Gruber, J., Halliwell, B., and Gunawan, R. (2015). Are mutagenic non D-loop direct repeat motifs in mitochondrial DNA under a negative selection pressure? Nucleic Acids Res. *43*, 4098–4108.

Lakshmanan, L.N., Yee, Z., Ng, L.F., Gunawan, R., Halliwell, B., and Gruber, J. (2018). Clonal expansion of mitochondrial DNA deletions is a private mechanism of aging in long-lived animals. Aging Cell *17*, e12814.

Lakshmipathy, U., and Campbell, C. (1999). Double strand break rejoining by mammalian mitochondrial extracts. Nucleic Acids Res. *27*, 1198–1204.

Li, Y., Huang, T.T., Carlson, E.J., Melov, S., Ursell, P.C., Olson, J.L., Noble, L.J., Yoshimura, M.P., Berger, C., Chan, P.H., et al. (1995). Dilated cardiomyopathy and neonatal lethality in mutant mice lacking manganese superoxide dismutase. Nat. Genet. *11*, 376–381.

Ling, F., Hori, A., Yoshitani, A., Niu, R., Yoshida, M., and Shibata, T. (2013). Din7 and Mhr1 expression levels regulate double-strand-break-induced replication and recombination of mtDNA at ori5 in yeast. Nucleic Acids Res. *41*, 5799–5816.

Loch, T., Vakhrusheva, O., Piotrowska, I., Ziolkowski, W., Ebelt, H., Braun, T., and Bober, E. (2009). Different extent of cardiac malfunction

and resistance to oxidative stress in heterozygous and homozygous manganese-dependent superoxide dismutase-mutant mice. Cardiovasc. Res. *82*, 448–457.

McVey, M., and Lee, S.E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. *24*, 529–538.

Melov, S., Hertz, G.Z., Stormo, G.D., and Johnson, T.E. (1994). Detection of deletions in the mitochondrial genome of Caenorhabditis elegans. Nucleic Acids Res. *22*, 1075–1078.

Mita, S., Rizzuto, R., Moraes, C.T., Shanske, S., Arnaudo, E., Fabrizi, G.M., Koga, Y., Dimauro, S., and Schon, E.A. (1990). Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. Nucleic Acids Res. *18*, 561–567.

Moraes, C.T., Dimauro, S., Zeviani, M., Lombes, A., Shanske, S., Miranda, A.F., Nakase, H., Bonilla, E., Werneck, L.C., Servidei, S., et al. (1989). Mitochondrial DNA deletions in progressive external ophthalmoplegia and Kearns-Sayre syndrome. N. Engl. J. Med. *320*, 1293–1299.

Persson, O., Muthukumar, Y., Basu, S., Jenninger, L., Uhler, J.P., Berglund, A.K., Mcfarland, R., Taylor, R.W., Gustafsson, C.M., Larsson, E., et al. (2019). Copy-choice recombination during mitochondrial L-strand synthesis causes DNA deletions. Nat. Commun. *10*, 759.

Phillips, A.F., Millet, A.R., Tigano, M., Dubois, S.M., Crimmins, H., Babin, L., Charpentier, M., Piganeau, M., Brunet, E., and Sfeir, A. (2016). Single-molecule analysis of mtDNA replication uncovers the basis of the common deletion. Mol. Cell *65*, 527–538.e6.

Pohjoismaki, J.L.O., Williams, S.L., Boettger, T., Goffart, S., Kim, J., Suomalainen, A., Moraes, C.T., and Braun, T. (2013). Overexpression of Twinkle-helicase protects cardiomyocytes from genotoxic stress caused by reactive oxygen species. Proc. Natl. Acad. Sci. U S A *110*, 19408–19413.

Poovathingal, S.K., Gruber, J., Halliwell, B., and Gunawan, R. (2009). Stochastic drift in mitochondrial DNA point mutations: a novel perspective ex silico. PLoS Comput. Biol. *5*, e1000572.

Poovathingal, S.K., Gruber, J., Lakshmanan, L., Halliwell, B., and Gunawan, R. (2012). Is mitochondrial DNA turnover slower than commonly assumed? Biogerontology *13*, 557–564.

Rodgers, K., and McVey, M. (2016). Error-prone repair of DNA double-strand breaks. J. Cell Physiol. *231*, 15–24.

Rotig, A., Bourgeron, T., Chretien, D., Rustin, P., and Munnich, A. (1995). Spectrum of mitochondrial DNA rearrangements in the Pearson marrow-pancreas syndrome. Hum. Mol. Genet. *4*, 1327–1330.

Rygiel, K.A., Tuppen, H.A., Grady, J.P., Vincent, A., Blakely, E.L., Reeve, A.K., Taylor, R.W., Picard, M., Miller, J., and Turnbull, D.M. (2016). Complex mitochondrial DNA rearrangements in individual cells from patients with sporadic inclusion body myositis. Nucleic Acids Res. *44*, 5313–5329.

Sadikovic, B., Wang, J., El-Hattab, A., Landsverk, M., Douglas, G., Brundage, E.K., Craigen, W.J., Schmitt, E.S., and Wong, L.-J.C. (2010). Sequence homology at the breakpoint and clinical phenotype of mitochondrial DNA deletion syndromes. PLoS One 5, e15687.

Sakofsky, C.J., Ayyar, S., Deem, A.K., Chung, W.H., Ira, G., and Malkova, A. (2015). Translesion polymerases drive microhomology-mediated break-induced replication leading to complex chromosomal rearrangements. Mol. Cell 60, 860–872.

Samuels, D.C., Schon, E.A., and Chinnery, P.F. (2004). Two direct repeats cause most human mtDNA deletions. Trends Genet. 20, 393–398.

Schwer, B., Wei, P.C., Chang, A.N., Kao, J., Du, Z., Meyers, R.M., and Alt, F.W. (2016). Transcription-associated processes cause DNA double-strand breaks and translocations in neural stem/progenitor cells. Proc. Natl. Acad. Sci. U S A 113, 2258–2263.

Sen, D., Nandakumar, D., Tang, G.Q., and Patel, S.S. (2012). Human mitochondrial DNA helicase TWINKLE is both an Unwinding and annealing helicase. J. Biol. Chem. 287, 14545–14556.

Sen, D., Patel, G., and Patel, S.S. (2016). Homologous DNA strand exchange activity of the human mitochondrial DNA helicase TWINKLE. Nucleic Acids Res. 44, 4200–4210.

Shoffner, J.M., Lott, M.T., Voljavec, A.S., Soueidan, S.A., Costigan, D.A., and Wallace, D.C. (1989). Spontaneous Kearns-Sayre/chronic external ophthalmoplegia plus syndrome associated with a mitochondrial DNA deletion: a slip-replication model and metabolic therapy. Proc. Natl. Acad. Sci. U S A 86, 7952–7956.

Song, S., Wheeler, L.J., and Mathews, C.K. (2003). Deoxyribonucleotide pool imbalance stimulates deletions in HeLa cell mitochondrial DNA. J. Biol. Chem. 278, 43893–43896.

Srivastava, S., and Moraes, C.T. (2005). Double-strand breaks of mouse muscle mtDNA promote large deletions similar to multiple mtDNA deletions in humans. Hum. Mol. Genet. 14, 893–902.

Tadi, S.K., Sebastian, R., Dahal, S., Babu, R.K., Choudhary, B., and Raghavan, S.C. (2016). Microhomology-mediated end joining is the principal mediator of double-strand break repair during mitochondrial DNA lesions. Mol. Biol. Cell 27, 223–235.

Tam, Z.Y., Gruber, J., Halliwell, B., and Gunawan, R. (2015). Context-dependent role of mitochondrial fusion-fission in clonal expansion of mtDNA mutations. PLoS Comput. Biol. 11, e1004183.

Thyagarajan, B., Padua, R.A., and Campbell, C. (1996). Mammalian mitochondria possess homologous DNA recombination activity. J. Biol. Chem. 271, 27536–27543.

Vielhaber, S., Debska-Vielhaber, G., Peeva, V., Schoeler, S., Kudin, A.P., Minin, I., Schreiber, S., Dengler, R., Kollewe, K., Zuschratter, W., et al. (2013). Mitofusin 2 mutations affect mitochondrial function by mitochondrial DNA depletion. Acta Neuropathol. 125, 245–256.

Volmering, E., Niehusmann, P., Peeva, V., Grote, A., Zsurka, G., Altmuller, J., Nurnberg, P., Becker, A.J., Schoch, S., Elger, C.E., et al. (2016). Neuropathological signs of inflammation correlate with mitochondrial DNA deletions in mesial temporal lobe epilepsy. Acta Neuropathol. 132, 277–288.

Wanrooij, S., and Falkenberg, M. (2010). The human mitochondrial replication fork in health and disease. Biochim. Biophys. Acta 1797, 1378–1388.

Wanrooij, S., Goffart, S., Pohjoismaki, J.L., Yasukawa, T., and Spelbrink, J.N. (2007). Expression of catalytic mutants of the mtDNA helicase Twinkle and polymerase POLG causes distinct replication stalling phenotypes. Nucleic Acids Res. 35, 3238–3251.

Wanrooij, S., Luoma, P., Van Goethem, G., Van Broeckhoven, C., Suomalainen, A., and Spelbrink, J.N. (2004). Twinkle and POLG defects enhance age-dependent accumulation of mutations in the control region of mtDNA. Nucleic Acids Res. 32, 3053–3064.

Xiong, X., Du, Z., Wang, Y., Feng, Z., Fan, P., Yan, C., Willers, H., and Zhang, J. (2015). 53BP1 promotes microhomology-mediated end-joining in G1-phase cells. Nucleic Acids Res. 43, 1659–1670.

Zhu, W., Qin, W., and Sauter, E.R. (2004). Large-scale mitochondrial DNA deletion mutations and nuclear genome instability in human breast cancer. Cancer Detect Prev. 28, 119–126.

Supplemental Information

Thermodynamic analysis

of DNA hybridization signatures

near mitochondrial DNA deletion breakpoints

Lakshmi Narayanan Lakshmanan, Zhuangli Yee, Barry Halliwell, Jan Gruber, and Rudiyanto Gunawan

Supplemental Information

Figure S1. Related to Figure 1
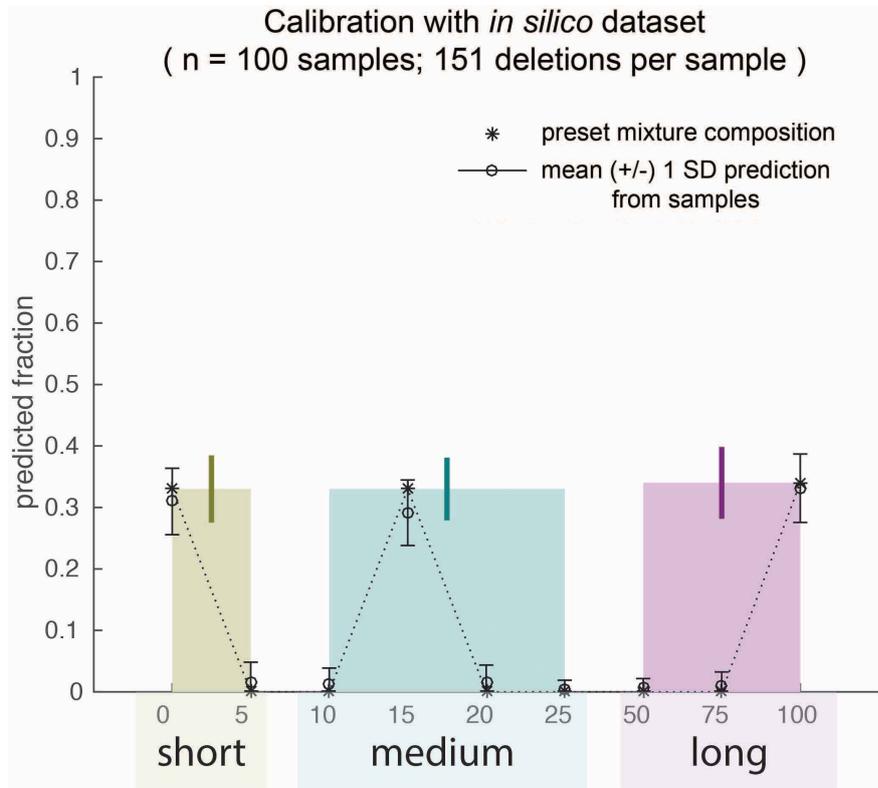
Figure S1. Mixture model analysis of *in silico* breakpoints datasets.

In silico random breakpoint datasets were generated using 3 misalignment length components with the following length composition: 33% 0-nt, 33% 15-nt and 34% 100-nt, based on human mtDNA length-specific likelihood values. Each *in silico* dataset (out of 100 datasets) comprises 151 deletions, the sample size of human aging dataset. The mixture model and the maximum likelihood estimation accurately recover the composition used to generate the *in silico* datasets. The true fractions (asterisk) and the mean ± standard deviation of the maximum likelihood estimates (circle and error bar) of the 100 *in silico* datasets are shown.

Figure S2. Related to Figure 1
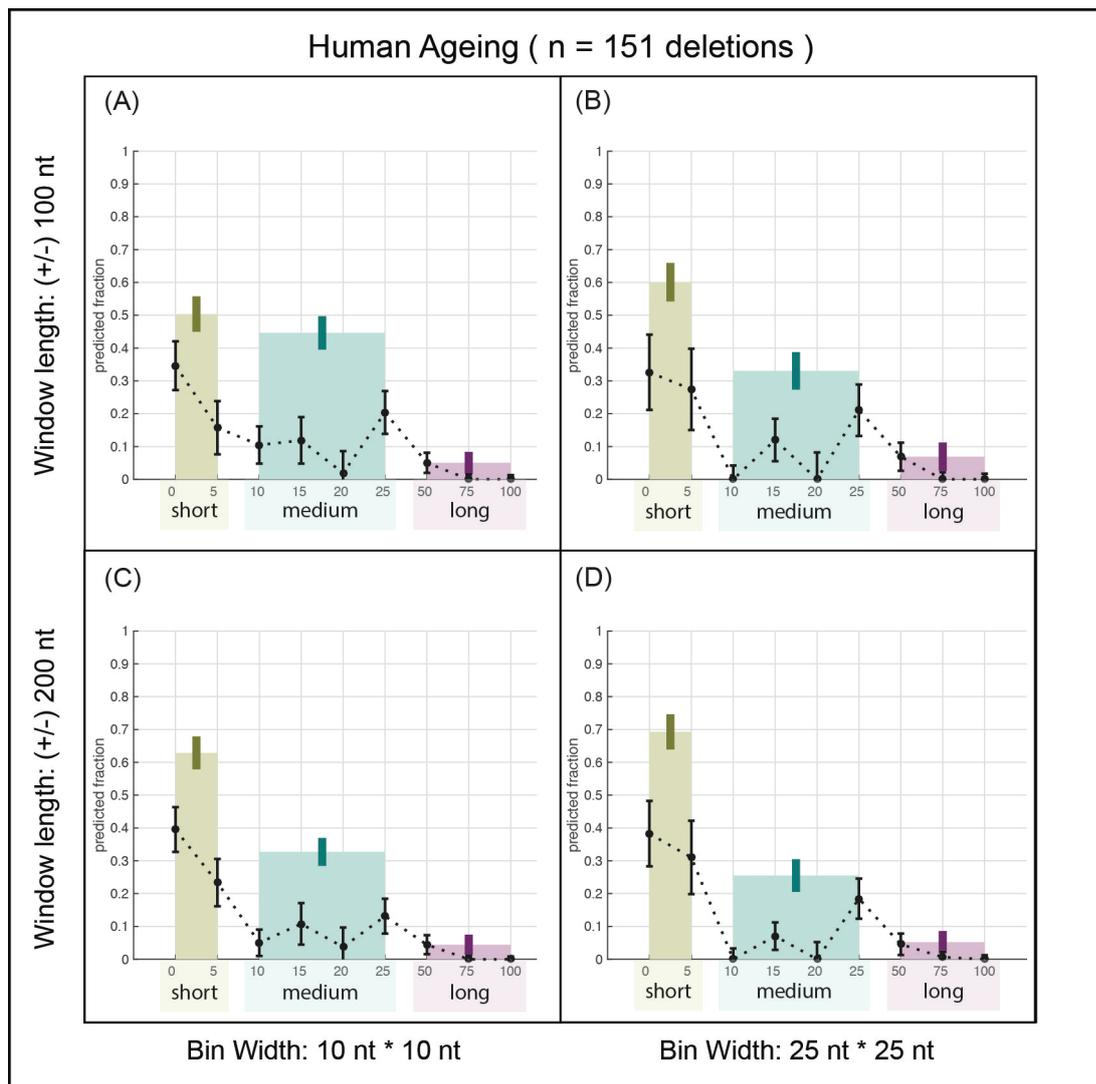
Figure S2. Mixture model analysis using differently sized window of analysis and bins.

We compared the maximum likelihood fraction θ_i^* of the mixture distribution model analysis using (A) ±100-nt window, 10-nt × 10-nt bin (default in main text), (B) ±100-nt window, 25-nt × 25-nt bin, (C) ±200-nt window, 10-nt × 10-nt bin and (D) ±200-nt window, 25-nt × 25-nt bin. The optimal fractions θ_i^* vary only slightly with the analysis window lengths and the bin sizes.

Figure S3. Related to Figure 2
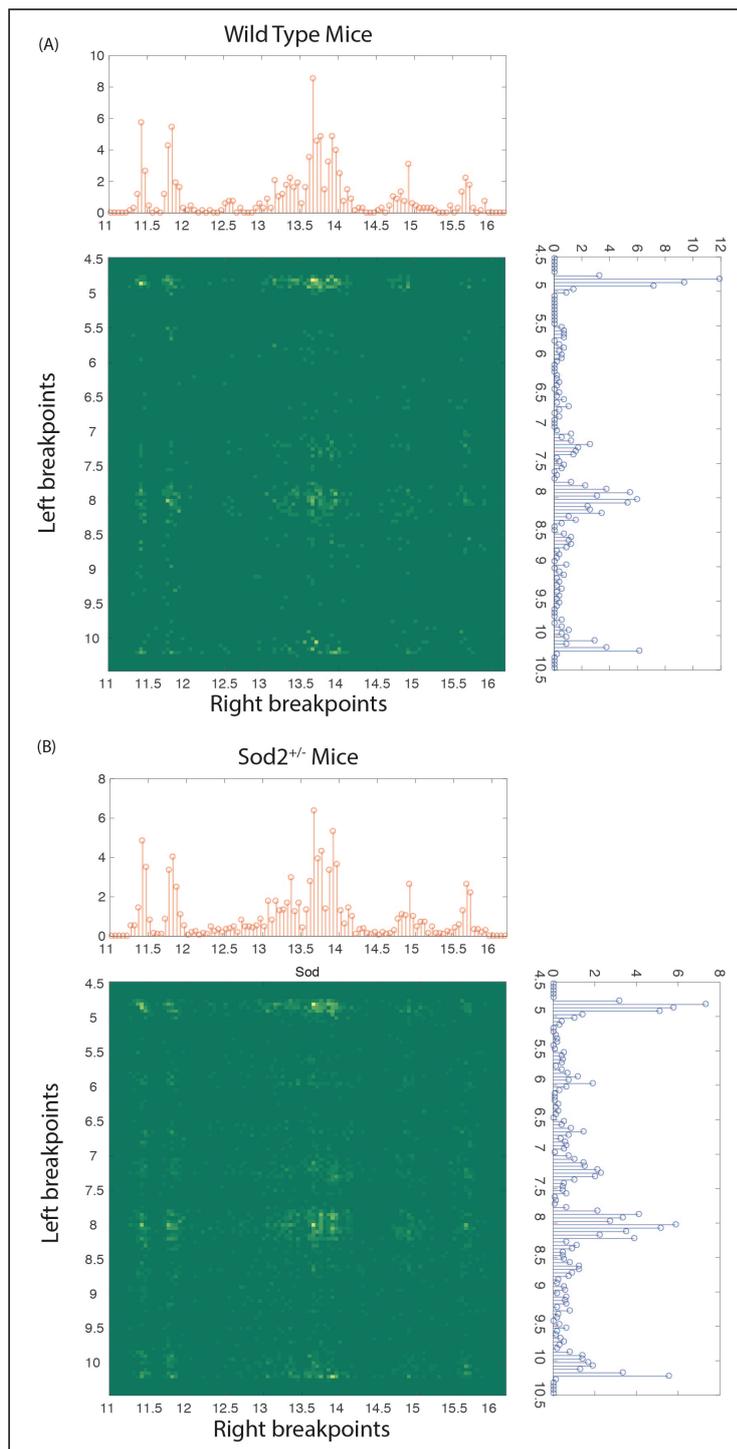
Figure S3. **Distributions of breakpoint positions**

(A) wild type mice and (B) *Sod2*⁺/⁻ mice. The location of hotspots (i.e. peak positions of breakpoint distributions) are similar in both wild type and *Sod2*⁺/⁻ mice, supporting the key role of oxidative damage in the formation of mtDNA deletions in wild type mice. The color intensity in the heat maps corresponds to the frequency of mtDNA deletions.

Figure S4. Related to Figures 2 and 3
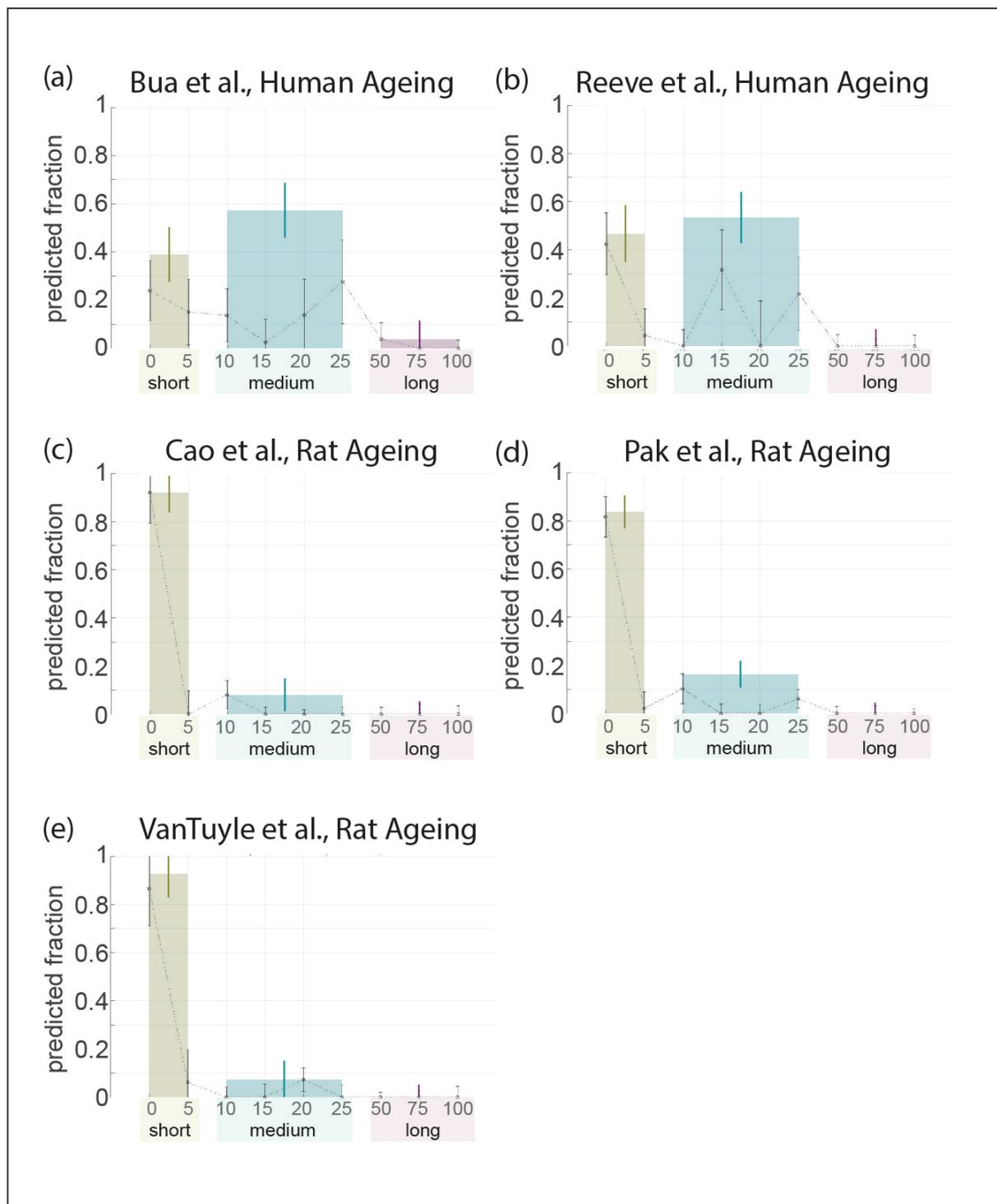
Figure S4. Mixture model analysis using datasets from individual articles.

We performed mixture model analysis on mtDNA deletion datasets reported in individual studies with dataset size above 25 deletions. Example results for individual article datasets from human ageing (a) Bua et al. (Bua et al., 2002), (b) Reeve et al. (Reeve et al., 2008) and rat ageing (c) Cao et al. (Cao et al., 2001), (d) Pak et al. (Pak et al., 2005) and (e) VanTuyle et al. (Van Tuyle et al., 1996) are shown in the figure.

Figure S5. Related to Figures 2 and 3 and Discussion
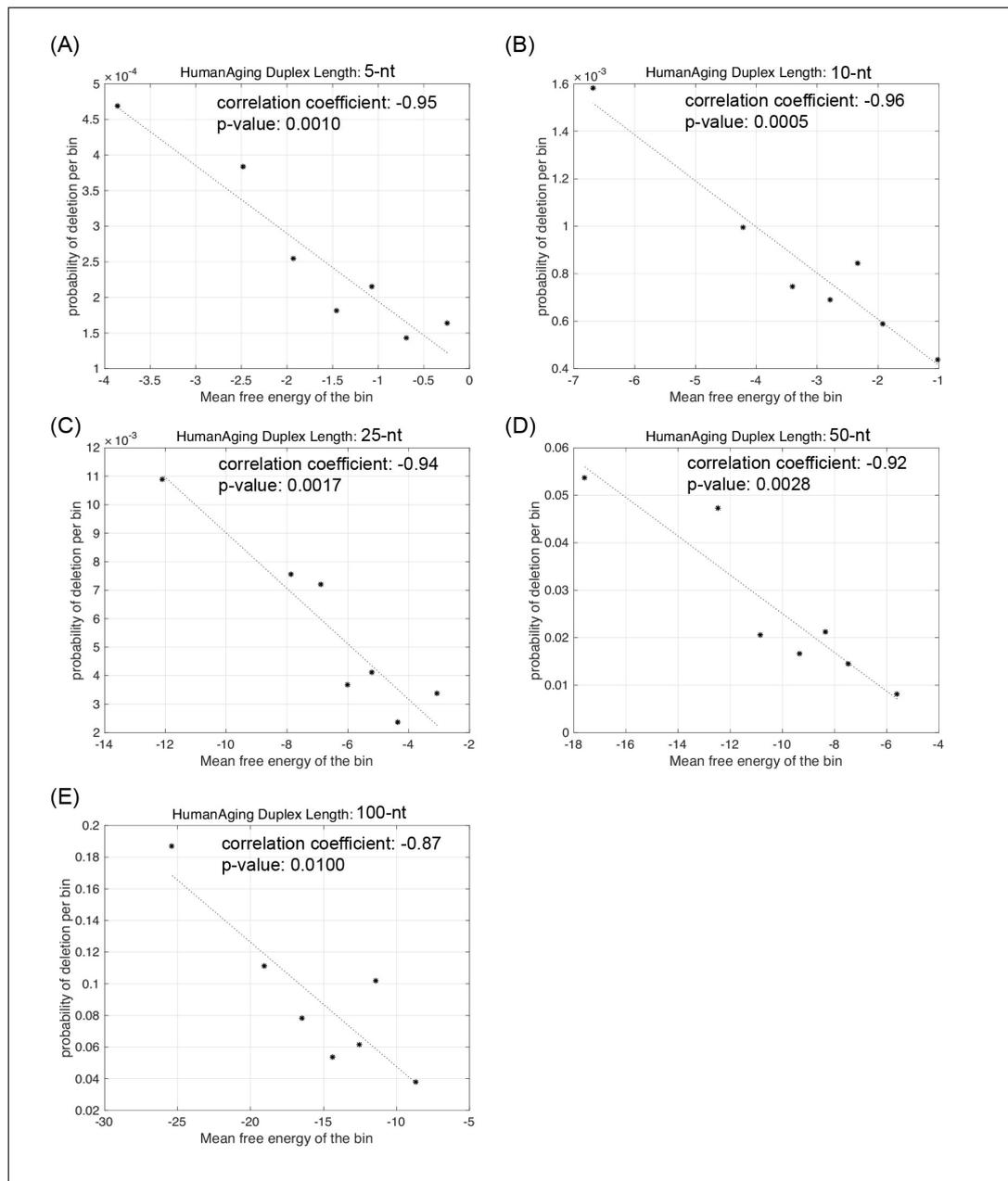
Figure S5. Enrichment analysis of human ageing mtDNA deletion breakpoints.

Guo *et al.* reported higher deletion breakpoint frequencies in regions of human mtDNA that could misalign with higher stability (i.e. more negative Gibbs free energy). The subplots show the frequencies of mtDNA deletion breakpoints in human ageing dataset associated with misalignments of mtDNA of length (A) 5-nt, (B) 10-nt, (C) 25-nt, (D) 50-nt and (E) 100-nt, as a function of the DNA hybridization free energies. Regardless of the segment length, the frequencies of mtDNA deletion breakpoints are higher for mtDNA segments forming more stable hybridization (i.e. more negative Gibbs free energy).

Table S1. Related to Table 1

Dataset count	Dataset Name	Sources
1	Human MTL epilepsy	Volmering et al., 2016
2	Human ageing	Bua et al., 2006, Dimberg et al., 2014, Eshaghian et al., 2006, Fayet et al., 2002, Kraytsberg et al., 2006, Nicholas et al., 2009, Reeve et al., 2008, Zhang et al., 1992
3	Single deletion myopathy	Campbell et al., 2014, Degoul et al., 1991, Ferlin et al., 1997, Grady et al., 2014, Kiyomoto et al., 1997, López-Gallardo et al., 2009, Mita et al., 1990, Ota et al., 1991, Sadikovic et al., 2010, Solano et al., 2003, Wilichowski et al., 1997
4	Human POLG	Wanrooij et al., 2004
5	Inclusion body myositis	Jansson et al., 2000, Moslemi et al., 1997, Rygiel et al., 2016
6	Pearson Syndrome	Ayed et al., 2011, Broomfield et al., 2015, de Vries et al., 1992, Ester Almeida, 2007, Gagne et al., 2014, Knerr et al., 2003, Krauch et al., 2002, Lee et al., 2007, Lohi et al., 2005, Morikawa et al., 1993, Muraki et al., 1997, Park et al., 2015, Rotig et al., 1995, Rotig et al., 1990, Sadikovic et al., 2010, Sato et al., 2015, Shanske et al., 2002, Superti-Furga et al., 1993, Tumino et al., 2011, Williams et al., 2012, Wong, 2001, Yanagihara et al., 2001
7	Charcot Marie Tooth Disease	Vielhaber et al., 2013
8	Rat ageing	Cao et al., 2001, Pak et al., 2005, Van Tuyle et al., 1996, Wanagat et al., 2001
9	Mouse ageing	Chung et al., 1996, Inoue et al., 1997, Inoue et al., 2000, Kim, 1997, Nelson et al., 1993, Vermulst et al., 2008, Wang et al., 1997

10	DSB mouse	Bacman et al., 2009, Fukui and Moraes, 2009, Srivastava and Moraes, 2005
11	WT mouse (NGS)	Pohjoismaki et al., 2013
12	Sod2$^{+/-}$ mouse (NGS)	Pohjoismaki et al., 2013
13	*TWNK*$^+$ (NGS)	Pohjoismaki et al., 2013
14	Sod2$^{+/-}$; *TWNK*$^+$ (NGS)	Pohjoismaki et al., 2013
15	Monkey ageing	Gokey et al., 2004, Lee et al., 1994, McKiernan et al., 2009
16	*C. elegans* ageing	Lakshmanan et al., 2018

Transparent Methods Section

Deletion breakpoint positions and mitochondrial DNA sequences

We compiled 9,921 breakpoint positions of mtDNA deletions from published literature and grouped them into species- and clinical condition-specific categories (see Supplemental Data). Breakpoints are categorized into distinct groups based on clinical conditions reported in source articles. The complete list of data sources from which the mtDNA breakpoints data were gathered is given in Supplemental Table 1 (Supplemental Information). Deletions reported in normal aged tissues from human, rhesus monkey, mouse, rat and *C. elegans* were grouped under age-related deletions. Disease-associated human mtDNA deletions were further grouped under single deletion myopathy (including Kearns-Sayre syndrome (KSS) and chronic progressive external ophthalmoplegia (CPEO)), myopathy with compound mutations in *polg* gene, Pearson syndrome (PS), Inclusion body myositis, Mesial temporal lobe epilepsy and Charcot Marie tooth disease based on the clinical description of the patients. Our datasets also included mtDNA deletions from mice expressing mitochondrial Pst1/Sca1 restriction enzymes, *Sod2*$^{+/-}$ heterozygous mice, wild type control mice and Twinkle overexpression mice. For mammals, we have considered only the deletions occurring within the mtDNA major arc region, as deletions involving the minor arc were much less frequent. Within each group of mtDNA deletions, identical breakpoint positions were accounted only once. Partially duplicated mtDNA (pd-mtDNA) molecules containing one wild type and one deleted mtDNA portion have been reported to be present together with mtDNA deletions. Rearrangements reported as duplications have been excluded from the datasets in our analysis. Similar to mtDNA deletions, breakpoint junctions of pd-mtDNA molecules have also been reported to have flanking short DR motifs (Bodyak et al., 2001). Hence, pd-mtDNA molecules, if present, could exhibit misalignments signatures similar to that of mtDNA deletions.

In the analyses, we used the complete, annotated mtDNA sequences for human (NC_012920), rhesus monkey (NC_005943), mouse (NC_005089), rat (X14848) and *C. elegans* (NC_001328) from NCBI database. The breakpoint positions were reported based on the L-strand nucleotide sequence.

DNA hybridization partition function calculation

DNA-DNA hybridization partition function (Z) is defined by

$$Z = \sum_i exp\left(\frac{-\Delta G_i^o}{RT}\right)$$

where, ΔG_i^o is the Gibb's free energy of hybridization of the ith conformation (Dimitrov and Zuker, 2004). Multiple conformations could exist between a given pair of hybridizing duplexes and the partition function sums up the probabilities of occurrence of all possible duplex conformations that could form between the given pair of DNA segments. We employed the *hybrid* subroutine of the UNAFold package, setting the temperature to 37° C for mammalian datasets and 25.5° C for *C. elegans* dataset, while keeping the default values for all other parameters (Markham and Zuker, 2008).

Generation of *in silico* breakpoints

In silico dataset is a set of computationally generated breakpoint pair positions with specific hybridization signatures at their breakpoint positions. To generate an *in silico* breakpoint dataset we first set (1) the composition of hybridization lengths θ_l^*, (2) the sample size (i.e., number of deletions), and (3) the window(s) of analysis (i.e. location in mtDNA from where the breakpoints should be drawn). From θ_l^* and the sample size, we determined the number of breakpoint pairs to be sampled from each misalignment length. For each position pair in the window of analysis, we calculated the length specific propensities of

DNA hybridization, as described in the main text. *In silico* breakpoint pair positions were randomly sampled from within the window of analysis using an inverse sampling procedure based on the propensity for each misalignment length *l*-nt in the analysis (Hammersley and Handscomb, 1964). For the validation of our analytical method, we used the window of analysis from human ageing dataset for the generation of *in silico* datasets.

For computing the variance of θ_l^* in the analysis of reported mtDNA deletions, we generated 100 *in silico* datasets using the windows of analysis and the optimal composition θ_l^* in the mtDNA deletion dataset. We analyzed each of the *in silico* datasets to generate 100 *in silico* θ_l^*'s. The variance of θ_l^* was computed as the variance of the 100 *in silico* θ_l^*'s.

Maximum likelihood analysis and statistical analysis

The optimization for likelihood maximization was performed using *fmincon* function in MATLAB (version 2015a; MathWorks, Inc.) with a multi-start strategy using 5 different starting points to ensure that we obtained a globally optimal solution. Statistical analyses were performed using in built functions in MATLAB. Two-sided z-test was used for statistical comparison of log-likelihood values between dataset and *in silico* generated samples. Two-sided *t*-test and Mann Whitney U (MWU) test were used for statistical comparison of mean, median values of breakpoint DR distance values, respectively. Unless stated otherwise, a statistical significance is set at p-value < 0.05.

Implementation of analysis from Guo et al. (2010)

The analysis was performed following the procedure described in the original study by Guo et al (Guo et al., 2010). We performed the analysis for 5, 10, 25, 50 and 100 nt long duplexes. While analyzing short sequence lengths, 35.8% of 820,000 5-nt duplexes and 1.7% of 205,000 10-nt long duplexes had positive values for minimum free energy of hybridization. These

duplexes with positive free energy values were excluded from the analysis. In Guo et al(Guo et al., 2010)., free energy of hybridization assigned to each deletion mutation corresponded to the free energy of the duplex that contained the breakpoint positions. 22.5% (for 5-nt) and 0.6% (for 10-nt) of the 151 human ageing mtDNA deletions fell within duplexes with positive free energy of hybridization and hence were excluded from the analysis. Pearson correlation coefficient and the statistical significance of the linear correlation were calculated using standard subroutines in MATLAB.

Supplemental References

Ayed, I. B., Chamkha, I., Mkaouar-Rebai, E., Kammoun, T., Mezghani, N., Chabchoub, I., Aloulou, H., Hachicha, M. & Fakhfakh, F. 2011. A Tunisian patient with Pearson syndrome harboring the 4.977kb common deletion associated to two novel large-scale mitochondrial deletions. *Biochem Biophys Res Commun,* 411**,** 381-6.

Bacman, S. R., Williams, S. L. & Moraes, C. T. 2009. Intra- and inter-molecular recombination of mitochondrial DNA after in vivo induction of multiple double-strand breaks. *Nucl. Acids Res.,* 37**,** 4218-4226.

Bodyak, N. D., Nekhaeva, E., Wei, J. Y. & Khrapko, K. 2001. Quantification and sequencing of somatic deleted mtDNA in single cells: Evidence for partially duplicated mtDNA in aged human tissues. *Human Molecular Genetics,* 10**,** 17-24.

Broomfield, A., Sweeney, M. G., Woodward, C. E., Fratter, C., Morris, A. M., Leonard, J. V., Abulhoul, L., Grunewald, S., Clayton, P. T., Hanna, M. G., et al. 2015. Paediatric single mitochondrial DNA deletion disorders: an overlapping spectrum of disease. *J Inherit Metab Dis,* 38**,** 445-57.

Bua, E., Johnson, J., Herbst, A., Delong, B., Mckenzie, D., Salamat, S. & Aiken, J. M. 2006. Mitochondrial DNA-deletion mutations accumulate intracellularly to detrimental levels in aged human skeletal muscle fibers. *Am J Hum Genet,* 79**,** 469-80.

Bua, E. A., Mckiernan, S. H., Wanagat, J., Mckenzie, D. & Aiken, J. M. 2002. Mitochondrial abnormalities are more frequent in muscles undergoing sarcopenia. *Journal of Applied Physiology,* 92**,** 2617-2624.

Campbell, G., Krishnan, K. J., Deschauer, M., Taylor, R. W. & Turnbull, D. M. 2014. Dissecting the mechanisms underlying the accumulation of mitochondrial DNA deletions in human skeletal muscle. *Hum Mol Genet,* 23**,** 4612-20.

Cao, Z., Wanagat, J., Mckiernan, S. H. & Aiken, J. M. 2001. Mitochondrial DNA deletion mutations are concomitant with ragged red regions of individual, aged muscle fibers: analysis by laser-capture microdissection. *Nucleic Acids Research,* 29**,** 4502-8.

Chung, S. S., Eimon, P. M., Weindruch, R. & Aiken, J. M. 1996. Analysis of age-associated mitochondrial DNA deletion breakpoint regions from mice suggests a novel model of deletion formation. *AGE,* 19**,** 117-128.

De Vries, D. D., Buzing, C. J., Ruitenbeek, W., Van Der Wouw, M. P., Sperl, W., Sengers, R. C., Trijbels, J. M. & Van Oost, B. A. 1992. Myopathology and a mitochondrial DNA deletion in the Pearson marrow and pancreas syndrome. *Neuromuscul Disord,* 2**,** 185-95.

Degoul, F., Nelson, I., Amselem, S., Romero, N., Obermaier-Kusser, B., Ponsot, G., Marsac, C. & Lestienne, P. 1991. Different mechanisms inferred from sequences of human mitochondrial DNA deletions in ocular myopathies. *Nucleic Acids Res,* 19**,** 493-6.

Dimberg, J., Hong, T. T., Skarstedt, M., Löfgren, S., Zar, N. & Matussek, A. 2014. Novel and differential accumulation of mitochondrial DNA deletions in Swedish and vietnamese patients with colorectal cancer. *Anticancer Res,* 34**,** 147-52.

Dimitrov, R. A. & Zuker, M. 2004. Prediction of Hybridization and Melting for Double-Stranded Nucleic Acids. *Biophysical Journal,* 87**,** 215-226.

Eshaghian, A., Vleugels, R. A., Canter, J. A., Mcdonald, M. A., Stasko, T. & Sligh, J. E. 2006. Mitochondrial DNA deletions serve as biomarkers of aging in the skin, but are typically absent in nonmelanoma skin cancers. *J Invest Dermatol,* 126**,** 336-44.

Ester Almeida, H. L., Helena Almeida, Maria Do Céu Machado, Aguinaldo Cabral, Laura Vilarinho 2007. Síndrome de Pearson. Caso clínico. *Acta Pediatrica Portuguesa,* 38**,** 79-81.

Fayet, G., Jansson, M., Sternberg, D., Moslemi, A. R., Blondy, P., Lombès, A., Fardeau, M. & Oldfors, A. 2002. Ageing muscle: clonal expansions of mitochondrial DNA point mutations and deletions cause focal impairment of mitochondrial function. *Neuromuscul Disord,* 12**,** 484-93.

Ferlin, T., Guironnet, G., Barnoux, M. C., Dumoulin, R., Stepien, G. & Mousson, B. 1997. Detection of mitochondrial DNA deletions by a screening procedure using the polymerase chain reaction. *Mol Cell Biochem,* 174**,** 221-5.

Fukui, H. & Moraes, C. T. 2009. Mechanisms of formation and accumulation of mitochondrial DNA deletions in aging neurons. *Hum Mol Genet,* 18**,** 1028-1036.

Gagne, K. E., Ghazvinian, R., Yuan, D., Zon, R. L., Storm, K., Mazur-Popinska, M., Andolina, L., Bubala, H., Golebiowska, S., Higman, M. A., et al. 2014. Pearson marrow pancreas syndrome in patients suspected to have Diamond-Blackfan anemia. *Blood,* 124**,** 437-40.

Gokey, N. G., Cao, Z., Pak, J. W., Lee, D., Mckiernan, S. H., Mckenzie, D., Weindruch, R. & Aiken, J. M. 2004. Molecular analyses of mtDNA deletion mutations in microdissected skeletal muscle fibers from aged rhesus monkeys. *Aging Cell,* 3**,** 319-26.

Grady, J. P., Campbell, G., Ratnaike, T., Blakely, E. L., Falkous, G., Nesbitt, V., Schaefer, A. M., Mcnally, R. J., Gorman, G. S., Taylor, R. W., et al. 2014. Disease progression in patients with single, large-scale mitochondrial DNA deletions. *Brain,* 137**,** 323-34.

Guo, X., Popadin, K. Y., Markuzon, N., Orlov, Y. L., Kraytsberg, Y., Krishnan, K. J., Zsurka, G., Turnbull, D. M., Kunz, W. S. & Khrapko, K. 2010. Repeats, longevity and the sources of mtDNA deletions: evidence from 'deletional spectra'. *Trends Genet,* 26**,** 340-343.

Hammersley, J. M. & Handscomb, D. C. 1964. *Monte Carlo methods,* London, New York, Methuen; Wiley.

Inoue, K., Ito, S., Takai, D., Soejima, A., Shisa, H., Lepecq, J. B., Segal-Bendirdjian, E., Kagawa, Y. & Hayashi, J. I. 1997. Isolation of mitochondrial DNA-less mouse cell lines and their application for trapping mouse synaptosomal mitochondrial DNA with deletion mutations. *J Biol Chem,* 272**,** 15510-5.

Inoue, K., Nakada, K., Ogura, A., Isobe, K., Goto, Y., Nonaka, I. & Hayashi, J. I. 2000. Generation of mice with mitochondrial dysfunction by introducing mouse mtDNA carrying a deletion into zygotes. *Nat Genet,* 26**,** 176-81.

Jansson, M., Darin, N., Kyllerman, M., Martinsson, T., Wahlström, J. & Oldfors, A. 2000. Multiple mitochondrial DNA deletions in hereditary inclusion body myopathy. *Acta Neuropathol,* 100**,** 23-8.

Kim, J.-S. K., Min-Jung; Kwon, in-Sook; Song, Eun-Sook 1997. Multiple Age-Associated Mitochondrial DNA Deletions in Mouse Brain. *BMB Reports,* 30**,** 33-36.

Kiyomoto, B. H., Tengan, C. H., Moraes, C. T., Oliveira, A. S. & Gabbai, A. A. 1997. Mitochondrial DNA defects in Brazilian patients with chronic progressive external ophthalmoplegia. *J Neurol Sci,* 152**,** 160-5.

Knerr, I., Metzler, M., Niemeyer, C. M., Holter, W., Gerecke, A., Baumann, I., Trollmann, R. & Repp, R. 2003. Hematologic features and clinical course of an infant with Pearson syndrome caused by a novel deletion of mitochondrial DNA. *J Pediatr Hematol Oncol,* 25**,** 948-51.

Krauch, G., Wilichowski, E., Schmidt, K. G. & Mayatepek, E. 2002. Pearson marrow-pancreas syndrome with worsening cardiac function caused by pleiotropic rearrangement of mitochondrial DNA. *Am J Med Genet,* 110**,** 57-61.

Kraytsberg, Y., Kudryavtseva, E., Mckee, A. C., Geula, C., Kowall, N. W. & Khrapko, K. 2006. Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons. *Nat Genet,* 38**,** 518-20.

Lakshmanan, L. N., Yee, Z., Ng, L. F., Gunawan, R., Halliwell, B. & Gruber, J. 2018. Clonal expansion of mitochondrial DNA deletions is a private mechanism of aging in long-lived animals. *Aging Cell***,** e12814.

Lee, C. M., Eimon, P., Weindruch, R. & Aiken, J. M. 1994. Direct repeat sequences are not required at the breakpoints of age-associated mitochondrial DNA deletions in rhesus monkeys. *Mech Ageing Dev,* 75**,** 69-79.

Lee, H. F., Lee, H. J., Chi, C. S., Tsai, C. R., Chang, T. K. & Wang, C. J. 2007. The neurological evolution of Pearson syndrome: case report and literature review. *Eur J Paediatr Neurol,* 11**,** 208-14.

Lohi, O., Kuusela, A. L. & Arola, M. 2005. A novel deletion in a Pearson syndrome infant with hypospadias and cleft lip and palate. *J Inherit Metab Dis,* 28**,** 1165-6.

López-Gallardo, E., López-Pérez, M. J., Montoya, J. & Ruiz-Pesini, E. 2009. CPEO and KSS differ in the percentage and location of the mtDNA deletion. *Mitochondrion,* 9**,** 314-7.

Markham, N. R. & Zuker, M. 2008. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol,* 453**,** 3-31.

Mckiernan, S. H., Colman, R., Lopez, M., Mark Beasley, T., Weindruch, R. & Aiken, J. M. 2009. Longitudinal analysis of early stage sarcopenia in aging rhesus monkeys. *Experimental Gerontology,* 44**,** 170-176.

Mita, S., Rizzuto, R., Moraes, C. T., Shanske, S., Arnaudo, E., Fabrizi, G. M., Koga, Y., Dimauro, S. & Schon, E. A. 1990. Recombination via flanking direct repeats is a major cause of large-scale deletions of human mitochondrial DNA. *Nucl. Acids Res.,* 18**,** 561-7.

Morikawa, Y., Matsuura, N., Kakudo, K., Higuchi, R., Koike, M. & Kobayashi, Y. 1993. Pearson's marrow/pancreas syndrome: a histological and genetic study. *Virchows Arch A Pathol Anat Histopathol,* 423**,** 227-31.

Moslemi, A. R., Lindberg, C. & Oldfors, A. 1997. Analysis of multiple mitochondrial DNA deletions in inclusion body myositis. *Hum Mutat,* 10**,** 381-6.

Muraki, K., Goto, Y., Nishino, I., Hayashidani, M., Takeuchi, S., Horai, S., Sakura, N. & Ueda, K. 1997. Severe lactic acidosis and neonatal death in Pearson syndrome. *J Inherit Metab Dis,* 20**,** 43-8.

Nelson, I., Gerasimov, S., Marsac, C., Lestienne, P. & Boursot, P. 1993. Sequence analysis of a deleted mitochondrial DNA molecule in heteroplasmic mice. *Mamm Genome,* 4**,** 680-3.

Nicholas, A., Kraytsberg, Y., Guo, X. & Khrapko, K. 2009. On the timing and the extent of clonal expansion of mtDNA deletions: evidence from single-molecule PCR. *Exp Neurol,* 218**,** 316-9.

Ota, Y., Tanaka, M., Sato, W., Ohno, K., Yamamoto, T., Maehara, M., Negoro, T., Watanabe, K., Awaya, S. & Ozawa, T. 1991. Detection of platelet mitochondrial DNA deletions in Kearns-Sayre syndrome. *Invest Ophthalmol Vis Sci,* 32**,** 2667-75.

Pak, J., Vang, F., Johnson, C., Mckenzie, D. & Aiken, J. 2005. MtDNA point mutations are associated with deletion mutations in aged rat. *Exp Gerontol,* 40**,** 209-218.

Park, J., Ryu, H., Jang, W., Chae, H., Kim, M., Kim, Y., Kim, J., Lee, J. W., Chung, N. G., Cho, B., et al. 2015. Novel 5.712 kb mitochondrial DNA deletion in a patient with Pearson syndrome: a case report. *Mol Med Rep,* 11**,** 3741-5.

Pohjoismaki, J. L. O., Williams, S. L., Boettger, T., Goffart, S., Kim, J., Suomalainen, A., Moraes, C. T. & Braun, T. 2013. Overexpression of Twinkle-helicase protects

cardiomyocytes from genotoxic stress caused by reactive oxygen species. *Proceedings of the National Academy of Sciences,* 110**,** 19408-19413.

Reeve, A. K., Krishnan, K. J., Elson, J. L., Morris, C. M., Bender, A., Lightowlers, R. N. & Turnbull, D. M. 2008. Nature of mitochondrial DNA deletions in substantia nigra neurons. *Am J Hum Genet,* 82**,** 228-35.

Rotig, A., Bourgeron, T., Chretien, D., Rustin, P. & Munnich, A. 1995. Spectrum of mitochondrial DNA rearrangements in the Pearson marrow-pancreas syndrome. *Human Molecular Genetics,* 4**,** 1327-1330.

Rotig, A., Cormier, V., Blanche, S., Bonnefont, J. P., Ledeist, F., Romero, N., Schmitz, J., Rustin, P., Fischer, A., Saudubray, J. M., et al. 1990. Pearson's marrow-pancreas syndrome. A multisystem mitochondrial disorder in infancy. *The Journal of Clinical Investigation,* 86**,** 1601-8.

Rygiel, K. A., Tuppen, H. A., Grady, J. P., Vincent, A., Blakely, E. L., Reeve, A. K., Taylor, R. W., Picard, M., Miller, J. & Turnbull, D. M. 2016. Complex mitochondrial DNA rearrangements in individual cells from patients with sporadic inclusion body myositis. *Nucl. Acids Res.,* 44**,** 5313-29.

Sadikovic, B., Wang, J., El-Hattab, A., Landsverk, M., Douglas, G., Brundage, E. K., Craigen, W. J., Schmitt, E. S. & Wong, L.-J. C. 2010. Sequence homology at the breakpoint and clinical phenotype of mitochondrial DNA deletion syndromes. *PLoS One,* 5**,** e15687.

Sato, T., Muroya, K., Hanakawa, J., Iwano, R., Asakura, Y., Tanaka, Y., Murayama, K., Ohtake, A., Hasegawa, T. & Adachi, M. 2015. Clinical manifestations and enzymatic activities of mitochondrial respiratory chain complexes in Pearson marrow-pancreas syndrome with 3-methylglutaconic aciduria: a case report and literature review. *Eur J Pediatr,* 174**,** 1593-602.

Shanske, S., Tang, Y., Hirano, M., Nishigaki, Y., Tanji, K., Bonilla, E., Sue, C., Krishna, S., Carlo, J. R., Willner, J., et al. 2002. Identical mitochondrial DNA deletion in a woman with ocular myopathy and in her son with pearson syndrome. *Am J Hum Genet,* 71**,** 679-83.

Solano, A., Gámez, J., Carod, F. J., Pineda, M., Playán, A., López-Gallardo, E., Andreu, A. L. & Montoya, J. 2003. Characterisation of repeat and palindrome elements in patients harbouring single deletions of mitochondrial DNA. *J Med Genet,* 40**,** e86.

Srivastava, S. & Moraes, C. T. 2005. Double-strand breaks of mouse muscle mtDNA promote large deletions similar to multiple mtDNA deletions in humans. *Hum Mol Genet,* 14**,** 893-902.

Superti-Furga, A., Schoenle, E., Tuchschmid, P., Caduff, R., Sabato, V., Demattia, D., Gitzelmann, R. & Steinmann, B. 1993. Pearson bone marrow-pancreas syndrome with insulin-dependent diabetes, progressive renal tubulopathy, organic aciduria and elevated fetal haemoglobin caused by deletion and duplication of mitochondrial DNA. *Eur J Pediatr,* 152**,** 44-50.

Tumino, M., Meli, C., Farruggia, P., La Spina, M., Faraci, M., Castana, C., Di Raimondo, V., Alfano, M., Pittalà, A., Lo Nigro, L., et al. 2011. Clinical manifestations and management of four children with Pearson syndrome. *Am J Med Genet A,* 155a**,** 3063-6.

Van Tuyle, G. C., Gudikote, J. P., Hurt, V. R., Miller, B. B. & Moore, C. A. 1996. Multiple, large deletions in rat mitochondrial DNA: evidence for a major hot spot. *Mutat Res,* 349**,** 95-107.

Vermulst, M., Wanagat, J., Kujoth, G. C., Bielas, J. H., Rabinovitch, P. S., Prolla, T. A. & Loeb, L. A. 2008. DNA deletions and clonal mutations drive premature aging in mitochondrial mutator mice. *Nat Genet,* 40**,** 392-4.

Vielhaber, S., Debska-Vielhaber, G., Peeva, V., Schoeler, S., Kudin, A. P., Minin, I., Schreiber, S., Dengler, R., Kollewe, K., Zuschratter, W., et al. 2013. Mitofusin 2 mutations affect mitochondrial function by mitochondrial DNA depletion. *Acta Neuropathol,* 125**,** 245-56.

Volmering, E., Niehusmann, P., Peeva, V., Grote, A., Zsurka, G., Altmuller, J., Nurnberg, P., Becker, A. J., Schoch, S., Elger, C. E., et al. 2016. Neuropathological signs of inflammation correlate with mitochondrial DNA deletions in mesial temporal lobe epilepsy. *Acta Neuropathol,* 132**,** 277-288.

Wanagat, J., Cao, Z., Pathare, P. & Aiken, J. M. 2001. Mitochondrial DNA deletion mutations colocalize with segmental electron transport system abnormalities, muscle fiber atrophy, fiber splitting, and oxidative damage in sarcopenia. *Faseb j,* 15**,** 322-32.

Wang, E., Wong, A. & Cortopassi, G. 1997. The rate of mitochondrial mutagenesis is faster in mice than humans. *Mutat Res,* 377**,** 157-66.

Wanrooij, S., Luoma, P., Van Goethem, G., Van Broeckhoven, C., Suomalainen, A. & Spelbrink, J. N. **2004**. Twinkle and POLG defects enhance age-dependent accumulation of mutations in the control region of mtDNA. *Nucl. Acids Res.,* 32**,** 3053-64.

Wilichowski, E., Grüters, A., Kruse, K., Rating, D., Beetz, R., Korenke, G. C., Ernst, B. P., Christen, H. J. & Hanefeld, F. 1997. Hypoparathyroidism and deafness associated with pleioplasmic large scale rearrangements of the mitochondrial DNA: a clinical and molecular genetic study of four children with Kearns-Sayre syndrome. *Pediatr Res,* 41**,** 193-200.

Williams, T. B., Daniels, M., Puthenveetil, G., Chang, R., Wang, R. Y. & Abdenur, J. E. 2012. Pearson syndrome: unique endocrine manifestations including neonatal diabetes and adrenal insufficiency. *Mol Genet Metab,* 106**,** 104-7.

Wong, L. J. 2001. Recognition of mitochondrial DNA deletion syndrome with non-neuromuscular multisystemic manifestation. *Genet Med,* 3**,** 399-404.

Yanagihara, I., Inui, K., Yanagihara, K., Park, Y. D., Tanaka, J., Ozono, K., Okada, S. & Kurahashi, H. 2001. Fluorescence in situ hybridization analysis of peripheral blood cells in Pearson marrow-pancreas syndrome. *J Pediatr,* 139**,** 452-5.

Zhang, C., Baumer, A., Maxwell, R. J., Linnane, A. W. & Nagley, P. 1992. Multiple mitochondrial DNA deletions in an elderly human individual. *FEBS Lett,* 297**,** 34-8.