



Advanced age increases frequencies of de novo mitochondrial mutations in macaque oocytes and somatic tissues

Barbara Arberthuber^{a,b}, Marzia A. Cremona^{c,d,e}, James Hester^f, Alison Barrett^a, Bonnie Higgins^a, Kate Anthony^a, Francesca Chiaromonte^{e,g,h}, Francisco J. Diaz^f, and Katerina D. Makova^{a,e,1}

Edited by Mark Stoneking, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; received October 19, 2021; accepted February 25, 2022

Mutations in mitochondrial DNA (mtDNA) contribute to multiple diseases. However, how new mtDNA mutations arise and accumulate with age remains understudied because of the high error rates of current sequencing technologies. Duplex sequencing reduces error rates by several orders of magnitude via independently tagging and analyzing each of the two template DNA strands. Here, using duplex sequencing, we obtained high-quality mtDNA sequences for somatic tissues (liver and skeletal muscle) and single oocytes of 30 unrelated rhesus macaques, from 1 to 23 y of age. Sequencing single oocytes minimized effects of natural selection on germline mutations. In total, we identified 17,637 tissue-specific de novo mutations. Their frequency increased ~3.5-fold in liver and ~2.8-fold in muscle over the ~20 y assessed. Mutation frequency in oocytes increased ~2.5-fold until the age of 9 y, but did not increase after that, suggesting that oocytes of older animals maintain the quality of their mtDNA. We found the light-strand origin of replication (OriL) to be a hotspot for mutation accumulation with aging in liver. Indeed, the 33-nucleotide-long OriL harbored 12 variant hotspots, 10 of which likely disrupt its hairpin structure and affect replication efficiency. Moreover, in somatic tissues, protein-coding variants were subject to positive selection (potentially mitigating toxic effects of mitochondrial activity), the strength of which increased with the number of macaques harboring variants. Our work illuminates the origins and accumulation of somatic and germline mtDNA mutations with aging in primates and has implications for delayed reproduction in modern human societies.

mitochondria | oocytes | mutations | duplex sequencing | heteroplasmy

Mitochondria produce energy and are involved in myriad other cellular functions (reviewed in ref. 1). The mammalian mitochondrial DNA (mtDNA) is a small (~16.6 kb in humans), circular, maternally transmitted molecule, which harbors 37 genes encoding 13 proteins (which form oxidative phosphorylation subunits), 22 transfer RNAs (tRNAs), and 2 ribosomal RNAs (rRNAs; reviewed in ref. 2). mtDNA is present in hundreds to thousands of copies per somatic cell and in >100,000 copies in an oocyte (3).

The germline nucleotide substitution rate of mtDNA is an order of magnitude higher than that of nuclear DNA (4, 5). Germline mutations increase in frequency with paternal and maternal age in nuclear DNA of humans (6) and macaques (7); however, whether they accumulate with maternal age in mtDNA of primates has been understudied. Such age-related accumulation was suggested based on the analysis of human pedigrees (4, 8) without the direct examination of germline cells and, thus, might have been influenced by selection. An investigation of mutation accumulation in the oocytes of females of different ages is needed to settle this question unequivocally.

The direct examination of mtDNA mutations in oocytes has been challenging due to methodological limitations. Most studies either focused on a limited number of mtDNA sites (e.g., refs. 9, 10) or used sequencing methods with high error rates (e.g., refs. 11, 12). Recently, an age-related increase of mtDNA mutations in mouse oocytes was demonstrated with duplex sequencing (13). However, we still do not know definitively whether the frequency of mtDNA mutations increases with age in primate oocytes. Answering this question is critical due to the association of mtDNA mutations with human genetic diseases (reviewed in ref. 14) and because of frequently delayed reproduction in modern human societies. Examining mutations in human oocytes presents multiple logistical and ethical challenges, requiring one to turn to a primate model.

The rhesus macaque is an excellent model organism to study mtDNA mutations in relation to aging due to 1) the high similarity between macaque and human mtDNA, innate defenses against oxidative damage (15), and age-related decline in metabolic rate

Significance

Multiple human genetic diseases are caused by mutations in the maternally transmitted DNA of mitochondria, the powerhouses of the cell. It is important to study how these mutations arise and accumulate with age, especially because humans in many societies now choose to have children at an older age. However, this is difficult to accomplish in humans, particularly for female germline cells, oocytes. To overcome this limitation, we studied mitochondrial mutation origins and accumulation with age in a primate model species, rhesus macaque. We found that new mutations accumulate the fastest in metabolically active liver and the slowest in oocytes. Thus, primate oocytes might have developed a mechanism to protect their mitochondrial DNA from excessive mutations, allowing reproduction later in life.

Author contributions: B.A., F.C., F.J.D., and K.D.M. designed research; B.A., M.A.C., J.H., A.B., B.H., K.A., F.C., and F.J.D. performed research; B.A. and M.A.C. analyzed data; and B.A., M.A.C., F.C., F.J.D., and K.D.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: kdm16@psu.edu.

This article contains supporting information online at <http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2118740119/-/DCSupplemental>.

Published April 8, 2022.

(16); and 2) the possibility of collecting oocytes from macaques starting at a young age. For humans, oocyte collection is mainly restricted to the reproductive lifespan, when in vitro fertilization procedures are performed.

Here, we analyzed mutations in single oocytes and somatic tissues of rhesus macaques over an age span of >20 y, including samples from animals who have not reached sexual maturity (occurring at ~3 y; ref. 17), as well as from animals up to the age of 23 y, covering the whole reproductive lifespan (macaques reach menopause at the age of ~25 y; ref. 18). To measure de novo mutations, we used highly accurate duplex sequencing (19), allowing one to distinguish bona fide DNA variants from artifacts (sequencing and PCR errors, or DNA lesions) by bar-coding double-stranded sequencing templates and achieving error rates $<10^{-7}$. With this method, first, single-strand consensus sequences (SSCSs) are formed for reads originating from each of the two template strands separately. Next, a duplex consensus sequence (DCS) is formed from the two SSCSs. True DNA variants are expected to be present in both SSCSs and, thus, in the DCS. Using this method, we directly measured the frequency of de novo germline and somatic mutations across the whole mtDNA in macaques, demonstrating their accumulation with age. We identified variant hotspots, analyzed the effect of selection, and examined the dependence of allele frequencies of inheritable mtDNA heteroplasmies on age.

Results

Duplex Sequencing and Mutation Detection. To study the role of age in somatic and germline mutation accumulation, we generated high-quality full-length mtDNA sequences for 30 unrelated Indian rhesus macaques ranging in age from 1 to 23 y (Fig. 1A). The macaques were collected at five primate research centers (SI Appendix, Table S1). For every animal, we assayed mtDNA from liver, skeletal muscle (henceforth called “muscle”), and from 1 to 14 single oocytes (SI Appendix, Table S1; a total of 152 single oocytes were assayed). We also assayed mtDNA from heart for nine macaques selected based on the presence of inheritable heteroplasmies (SI Appendix, Table S1).

To identify DNA variants, we used duplex sequencing (19), following the published protocol (13). Briefly, the samples were enriched for circular mtDNA with the exonuclease V digestion of linear nuclear DNA (SI Appendix, Table S2 and Fig. S1A). DNA was sheared to ~550 bp and sequenced on the Illumina HiSeq 2500 platform using 250-nt paired-end reads and analyzed with the Du Novo software (20) to form DCSs. We achieved a median mtDNA enrichment of 92.4%, 96.6%, 98.1%, and 73.5% and a median DCS depth of 2,906 \times , 3,729 \times , 2,565 \times , and 260 \times per site per sample for liver, muscle, heart, and single oocytes, respectively (SI Appendix, Table S2 and Fig. S1 A and B), with a uniform DCS depth across the mtDNA length (SI Appendix, Fig. S1C). The DCSs were mapped to the macaque reference genome consisting of nuclear DNA and mtDNA, and only alignments with the best match to the mtDNA were retained. The enzymatic digestion of linear nuclear DNA, the relatively long fragment size and read length during sequencing (4), and the retention of only reads mapping to mtDNA all minimize the contribution from regions of nuclear DNA that are homologous to mtDNA (Numts) to our mutation dataset (see *Materials and Methods*). At 122 different mtDNA positions, we identified homoplasmic differences from the reference sequence (supported by all DCSs in all tissues studied for an animal), with 7 to 16 such sites per animal (SI Appendix, Fig. S2). Additionally, we identified sites with

nucleotides different from the majority of DCSs for a sample as “variants” (see *Materials and Methods*).

Across all samples studied, we identified 18,532 variants. Among them, we detected and analyzed separately 221 inheritable variants (i.e., heteroplasmies present in all somatic tissues and ≥ 1 oocyte of an animal). We filtered out 1) 391 variants with ambiguous timing of occurrence during development (i.e., representing potential inheritable heteroplasmies, early somatic mutations, or early germline mutations; see SI Appendix, *Materials and Methods*); 2) 213 variants present at the beginning or the end of long homopolymer runs (representing potential read mapping errors); and 3) 70 low-confidence variants present in oocytes with a mean DCS depth $<100\times$ (we retained variants from 119 oocytes with a mean DCS depth $\geq 100\times$, i.e., from 1 to 12 oocytes per animal for a total of 29 animals; included oocytes did not significantly differ in diameter between prepubertal and sexually mature females; SI Appendix, Fig. S1D). As a result, we obtained 17,637 high-confidence tissue-specific de novo mutations (SI Appendix, Fig. S3). Among them, we observed 56 mutation sites—harboring 117 variants—shared by oocytes of the same animal in eight animals. This is 37 more sites than expected by random chance (SI Appendix, Note S1), and, thus, a small number of de novo germline mutations might have arisen in lineages prior to divergence of individual oocytes. We nevertheless kept these mutations because we cannot distinguish them from shared de novo mutations arising independently in two oocytes.

To study the frequencies and patterns of de novo mutations and inheritable heteroplasmies depending on age, we assigned macaques to four age groups of approximately equal size (Fig. 1A): 1) young (<5 y; 9 animals; 36 single oocytes), 2) intermediate 1 (5 to <10 y; 7 animals; 13 single oocytes), 3) intermediate 2 (>10 to 15 y; 7 animals; 46 single oocytes), and 4) old (>15 y; 7 animals; 24 single oocytes; SI Appendix, Table S1). The distribution of the 17,637 de novo mutations among the age groups and tissues is shown in SI Appendix, Table S3. The majority of de novo mutations (15,810) were each measured in a single DCS and, thus, had low minor allele frequencies (MAFs). Only 53 de novo mutations had a MAF $\geq 1\%$ (SI Appendix, Fig. S4A), a common cutoff for mtDNA mutation detection in other studies. Because these mutations are tissue-specific, they likely rose to high frequencies due to clonal expansions.

De Novo Mutations.

Increase in mutation frequency with age. To study the effect of age on mutation frequency, we built a generalized mixed-effects linear model (see *Materials and Methods*). The model predicted the probability of having a mutation in a sequenced nucleotide as a function of age (Fig. 1B) by using the mutation frequency as response; the number of sequenced nucleotides per sample as weight; and age, tissue, and their interactions as fixed-effect predictors. The macaque ID was used as a random effect. The model indicated a significant increase in mutation frequency with age in all tissues analyzed (Z -test $P = 2.7 \times 10^{-30}$, $P = 2.4 \times 10^{-20}$, $P = 3.4 \times 10^{-45}$, and $P = 1.7 \times 10^{-12}$ for muscle, heart, liver, and oocytes, respectively). Compared with the average mutation frequency at birth for muscle, such frequency was not significantly different for heart but was higher for liver and lower for oocytes (SI Appendix, Fig. S5A and Table S4). There was a steeper increase in mutation frequency with age for liver and heart, and a more moderate one for oocytes, than for muscle. A moderate increase in oocytes prompted us to build a separate generalized mixed-effects

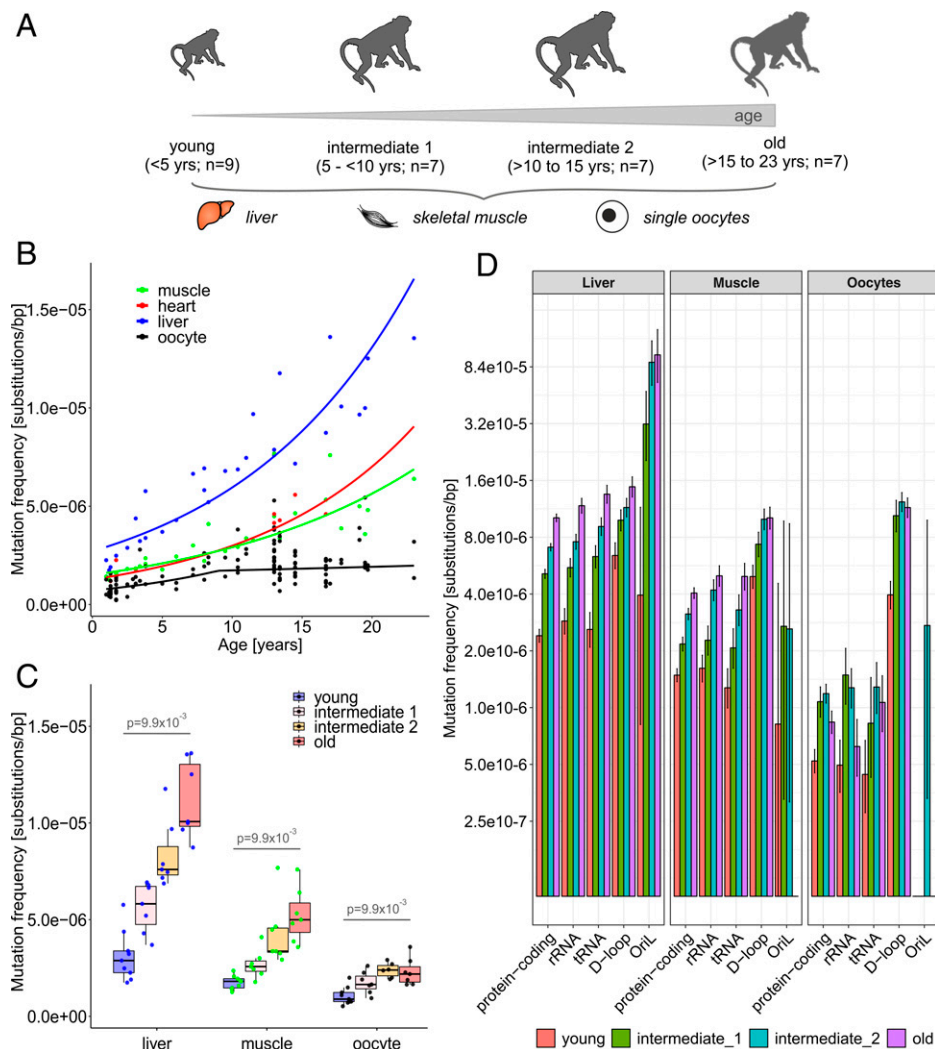


Fig. 1. Age-related increase in mutation frequency. (A) Liver, muscle, and oocyte samples were sequenced from 1- to 23-y-old macaques, classified into four age groups: young, intermediate 1, intermediate 2, and old. (B) Mixed-effects linear model analyzing the effect of age on mutation frequencies (a piecewise model was used for oocytes). Curves show the predicted mutation frequency based on the fixed-effect part of the model. Dots are the observed mutation frequencies per tissue per macaque; predicted frequencies and bootstrap confidence intervals are in *SI Appendix, Fig. S5B*. (C) Mutation frequencies measured in liver, muscle, and single oocytes shown for each macaque and for each age group. The average mutation frequency among oocytes analyzed for an animal was calculated; the results did not change qualitatively when each oocyte was considered individually (*SI Appendix, Table S3 and Fig. S6*). *P* values are for a permutation test comparing young and old age groups (one-sided test based on medians; 10,000 permutations; corrected for multiple testing, see *Materials and Methods*). (D) Tissue-specific mutation frequencies (i.e., total number of tissue-specific mutations divided by the product of mtDNA region length and mean DCS depth) for each age group in protein-coding (11,370 bp), rRNA (2,505 bp), tRNA (1,504 bp), D-loop (1085 bp), and Or1L (33 bp; including a 3-bp overlap with tRNA) sequences. Mutation frequencies for noncoding sequences outside of the D-loop and Or1L (98 bp) are shown in *SI Appendix, Table S6*. Mutation frequency bars are shown with 95% Poisson confidence intervals.

model with the same response, weight, and predictors as the previous model but allowing for a change in slope in the relationship between oocyte mutation frequency and age (Fig. 1B and *SI Appendix, Table S5*). The resulting piecewise model fit the data significantly better than the one not allowing for a change in slope (likelihood ratio test $P = 1.6 \times 10^{-6}$) and identified a break at the age of 9 y, prior to which oocyte mutation frequency increased with age (Z -test $P = 3.6 \times 10^{-15}$), but after which it did not (Z -test $P = 0.32$; odds ratios are in *SI Appendix, Table S5*). Macaque ID had a significant effect (likelihood ratio test $P = 6.4 \times 10^{-71}$), suggesting that some macaques have more highly mutable mtDNA than others (*SI Appendix, Table S5*). The primate research center was not significant (likelihood ratio test $P = 0.78$) and, thus, not added to the model.

We then compared the mutation frequencies among animals from different age groups (here and in the subsequent de novo mutation analyses, we excluded heart due to the limited data). The mutation frequency was significantly higher in old than in

young animals (i.e., between the two extreme age groups, with a 3.5-fold, 2.8-fold, and 2.5-fold increase in liver, muscle, and oocytes, respectively; one-sided permutation test for equality in medians $P = 9.9 \times 10^{-3}$, $P = 9.9 \times 10^{-3}$, and $P = 9.9 \times 10^{-3}$, respectively; Fig. 1C and *SI Appendix, Table S3*). Moreover, the mutation frequency was almost always significantly higher in the older than in the younger animals from any two adjacent age groups, except for insignificant differences between muscle of intermediate 1 and intermediate 2 animals and between oocytes of intermediate 2 and old animals (*SI Appendix, Table S3*). Similar results were obtained when aggregating mutations across animals in each tissue and age group (*SI Appendix, Table S6*) and when considering only mutations measured in >1 DCS (*SI Appendix, Fig. S4B*). To increase statistical power, such aggregated mutations were used for the rest of the analyses in this section.

Variation in mutation frequency along the mtDNA. The mutation frequency was significantly higher in the D-loop than in the

protein-coding, rRNA, and tRNA regions combined for each tissue and each age group studied (Fig. 1D; see *SI Appendix, Table S6* for Fisher's exact test *P* values). For each age group, in the D-loop, the mutation frequency was relatively similar among the tissues analyzed; however, in the protein-coding, rRNA, and tRNA regions, the mutation frequency was highest in liver, intermediate in muscle, and lowest in oocytes. As a result, for the D-loop, fold increases in mutation frequency between old and young animals were rather consistent among tissues: 2.3, 2.0, and 2.9 in liver, muscle, and oocytes, respectively. In contrast, coding regions had the highest fold differences in mutation frequency between old and young animals in liver, intermediate fold differences in muscle, and the lowest fold differences in oocytes: 4.2, 2.7, and 1.6, respectively, for protein-coding regions.

When inspecting mutations along the mtDNA in 80×207 -bp bins, we noticed a peak around 5,700 nt in liver of intermediate 1, intermediate 2, and old animals (*SI Appendix, Fig. S7A*). This peak included the light-strand origin of replication (OriL, 33 bp) and was particularly strong for old animals. Therefore, we next compared mutation frequencies among noncoding regions: the OriL, the D-loop (1,085 bp) containing the heavy-strand origin of replication (Fig. 1D), and the noncoding DNA outside these two regions (98 bp; *SI Appendix, Table S7*). This analysis confirmed that in liver, OriL exhibited high mutation frequencies—particularly in older animals (Fig. 1D and *SI Appendix, Fig. S7*)—and, thus, is a region of preferential mutation accumulation with aging. Indeed, in liver, OriL in old animals had a 5.0-fold-higher mutation frequency compared to the D-loop in old animals (Fisher's exact test, corrected for multiple testing, $P = 4.5 \times 10^{-15}$) and a 19-fold-higher mutation frequency compared to OriL in young animals (Fisher's exact test, corrected for multiple testing, $P = 2.7 \times 10^{-12}$; Fig. 1D and *SI Appendix, Table S7*). Furthermore, a sliding window analysis (33-bp windows shifted by 1 bp at a time) showed that in liver of intermediate 1, intermediate 2, and old animals, only 13, 2, and 13 out of 16,532 windows, respectively, had mutation frequencies equal to or higher than that measured in the window consisting of the OriL sequence (*SI Appendix, Fig. S7C*). Notably, 5, 2, and 13 of these windows, respectively, overlapped with OriL, and the remaining 8 windows were in the D-loop. The mutation frequency in noncoding DNA outside of OriL and the D-loop was low (*SI Appendix, Table S7*).

By analyzing observed versus expected numbers of de novo mutations based on the length of mtDNA functional regions (*SI Appendix, Table S8*), we found the D-loop to harbor a higher number of mutations than expected by chance in each tissue and each age group analyzed. OriL had a higher-than-expected number of mutations in liver of each age group except for young animals (two-sided binomial test $P = 0.47$, $P = 3.1 \times 10^{-11}$, $P = 2.9 \times 10^{-33}$, and $P = 4.2 \times 10^{-20}$ for young, intermediate 1, intermediate 2, and old, respectively). De novo mutations were significantly underrepresented in protein-coding regions in each tissue and age group analyzed (see *SI Appendix, Table S8* for *P* values). In oocytes, the nonsynonymous-to-synonymous rate ratio (21) (hN/hS, equal to 1.56, 0.72, 1.45, and 1.03 for young, intermediate 1, intermediate 2, and old animals, respectively; *SI Appendix, Fig. S8*) was within (or close to) the range of neutral expectations. In somatic tissues, hN/hS ratios were higher than neutral expectations (*SI Appendix, Fig. S8*), with particularly high ratios in old animals (1.62 in liver and 1.41 in muscle), suggesting positive selection. Therefore, our results were largely inconsistent with purifying selection acting against de novo mutations in

protein-coding regions. We did not observe a significant depletion of mutations in regulatory versus nonregulatory regions of the D-loop (*SI Appendix, Table S9* and Fig. S9), providing no evidence of purifying selection in it.

Preferential accumulation of transitions with aging, mutations at CpGs, and strand bias. The majority of de novo mutations were transitions, mostly accumulated with aging (*SI Appendix, Table S10*). Between young and old macaques, there were 4.5-, 3.1-, and 2.8-fold increases (Fisher's exact test $P < 1.0 \times 10^{-250}$, 1.2×10^{-153} , and 5.0×10^{-42}) in the frequency of transitions in liver, muscle, and oocytes, respectively. However, there were only 1.5- and 1.2-fold increases—and, in fact, a 1.1-fold decrease (Fisher's exact test $P = 1.2 \times 10^{-3}$, 0.069, and 0.306; *SI Appendix, Table S11*)—in the frequency of transversions in liver, muscle, and oocytes, respectively. Significant increases in frequencies in young compared to old macaques were observed for both A > G and/or T > C (4.0-, 2.9-, and 3.0-fold in liver, muscle, and oocytes, respectively) and C > T and/or G > A (4.8-, 3.1-, and 2.7-fold in liver, muscle, and oocytes, respectively) transitions (*SI Appendix, Fig. S10*; see *SI Appendix, Table S11* for mutation frequencies and Fisher's exact test *P* values).

The effect of methylation of CpG sites on mtDNA mutations might be tissue specific, depending on mutation frequency. With more mutations generally occurring in the liver of older animals, the effect of CpG sites on mtDNA mutagenesis might be stronger in this tissue because of more rounds of mtDNA replication and more time spent by the mtDNA in the single-stranded state. In liver, in older animals (intermediate 1, intermediate 2, and old), the mutation frequency for G > A substitutions was significantly higher at CpG than non-CpG sites (1.3-, 1.3-, and 1.4-fold; Fisher's exact test $P = 7.0 \times 10^{-3}$, 1.3×10^{-5} , and 8.8×10^{-6} , respectively; *SI Appendix, Fig. S11* and Table S12). In muscle and oocytes, in most age groups, the mutation frequency was slightly higher at CpG sites than at non-CpG sites, but this difference was not significant (*SI Appendix, Fig. S11* and Table S12).

Consistent with previous reports based on conventional (22) and duplex (23) sequencing, several mtDNA mutation types exhibited strand bias in our data. Strand bias is a bias in the occurrence of mutations between the light and heavy strands (L-strands vs. H-strands). Without strand bias, and when correcting for the unequal nucleotide composition of the two strands, we expect similar numbers of mutations of the same type (e.g., C > T) originating on the L-strand versus on the H-strand. Duplex sequencing measures mutations on both DNA strands, but mutations are reported only with respect to the reference L-strand sequence. Using the L-strand as a reference, we expect similar numbers of mutations of one type (e.g., C > T) and of the complementary type (e.g., G > A; this is how C > T mutations originating from the H-strand manifest themselves) under the hypothesis of no strand bias (*SI Appendix, Materials and Methods*). However, this was not what we observed in our data on de novo mutations. Transitions showed a strong strand bias, particularly for C > T and G > A mutations, across all tissues, with similar patterns between younger and older animals (*SI Appendix, Fig. S12* and Table S13). With the L-strand as a reference, there were more G > A than C > T mutations—ranging from 10.8-fold to 16.0-fold more in somatic tissues and from 3.5-fold to 5.6-fold more in oocytes. The G > A over C > T strand bias was previously observed in human mtDNA (23). For transversions, a significant strand bias was observed for C > G and G > C (3.5-fold and 3.2-fold more G > C than C > G in liver of intermediate 2 and old animals, respectively) and A > C and T > G

(6.3-fold more A > C than T > G in liver of intermediate 2 animals). Note that some mutations originating from the L-strand (e.g., C > T mutations) might have been complementary mutations (e.g., G > A mutations) originating from the H-strand; however, this does not affect strand bias estimates (*SI Appendix, Materials and Methods*).

Variant hotspots. To identify variant hotspots, we built a probabilistic model that takes into account tissue-specific estimates of the average mutation frequencies, the higher mutability in the D-loop, and the mean DCS depth in each sample (*SI Appendix, Note S2*). We also considered the number of sequenced oocytes per animal. For each tissue, we computed

the expected number of mutations present in exactly one macaque and shared by several macaques (Fig. 2A). Results suggest that due to random chance, mutations at the same site are expected to occur in two animals and can sometimes occur in three or four animals, but they are rarely expected to occur in liver and muscle, and are not expected to occur at all in the oocytes, in five or more animals (*SI Appendix, Note S2*). Thus, we defined tissue-specific sites mutated in at least five different macaques as variant hotspots (*SI Appendix, Fig. S13 and Dataset S1*). We detected a total of 472 hotspots: 354, 93, and 25 in liver, muscle, and oocytes, respectively (the number of tissue-specific hotspots depends on the total number of

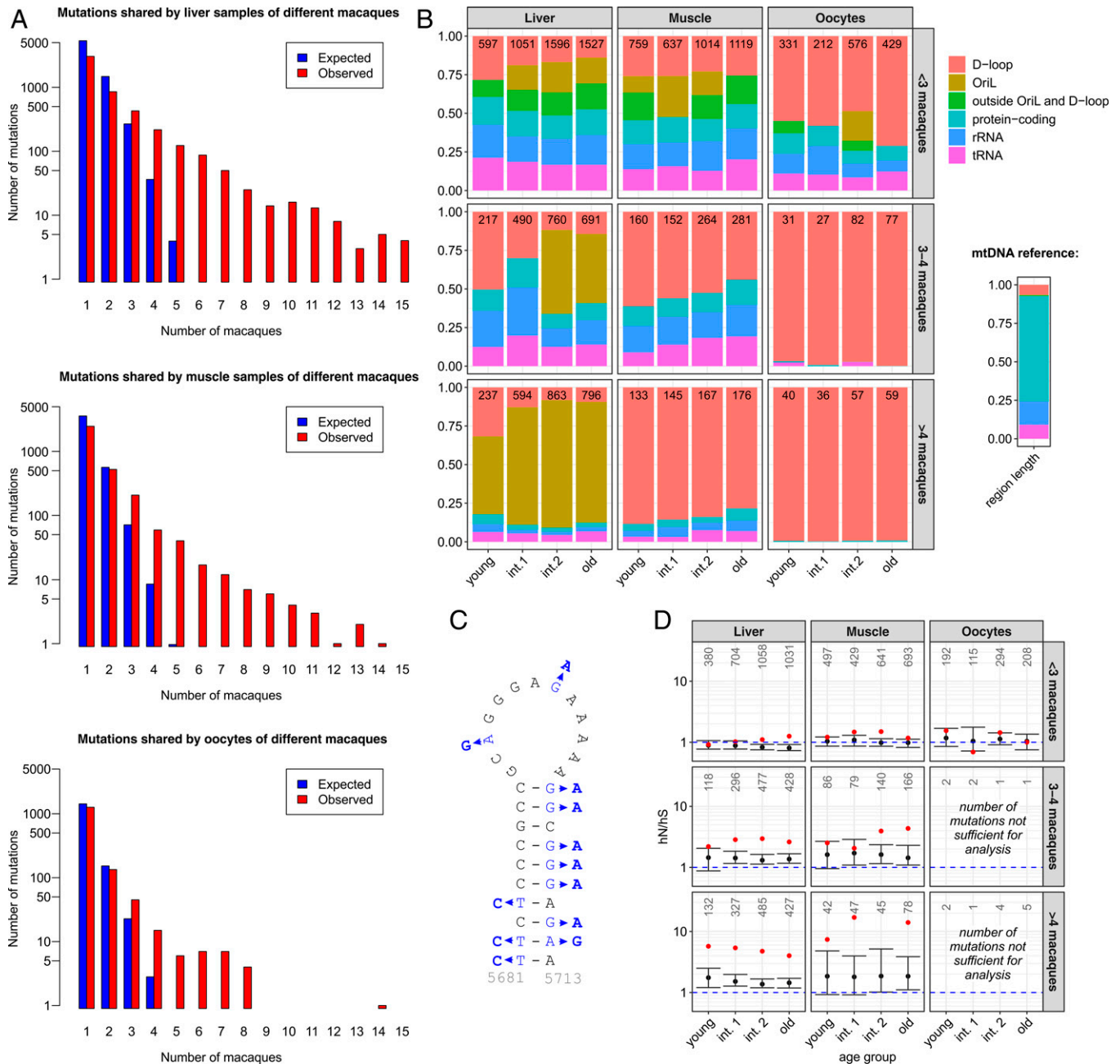


Fig. 2. Analysis of variant hotspots. (A) The observed and expected distributions of the number of macaques with individual variants in liver, muscle, and oocytes (with all oocytes from an animal considered together). (B) Distribution of mutations (normalized by the length of the respective region) among D-loop, OriL, noncoding DNA outside of OriL and D-loop, protein-coding regions, rRNA, and tRNA, shown separately for each tissue, and for mutations observed in 1-2, 3-4, or >4 macaques. Numbers indicate the total number of mutations analyzed (due to overlapping annotation of regions, some mutations were counted twice). The proportion of mtDNA occupied by each region is shown on the right. (C) Hairpin structure of OriL; variant hotspots are in blue. (D) hN/hS ratios for mutations observed in 1-2, 3-4, or >4 macaques. Numbers indicate the total number of mutations analyzed. In muscle of intermediate 2 animals, we did not find any synonymous mutations and, thus, could not compute the hN/hS ratio.

mutations measured in each tissue and, thus, is low in oocytes). The locations of 62 hotspot sites overlapped among tissues (*SI Appendix, Fig. S13*), leading to a total of 401 mtDNA sites affected by hotspots. Overall, 10–26% of all de novo variants (192, 621, and 2,483 out of 1,952, 4,998, and 9,389, for oocytes, muscle, and liver, respectively) were found at hotspot sites occupying only ~2.4% of the mtDNA length (401 out of 16,564).

Among the 401 hotspot sites, 87 were located in the D-loop, 42 in rRNA, 43 in tRNA, 218 in protein-coding regions, and 12 in OriL (one OriL site is also annotated as part of tRNA). This is different from what is expected based on the length of the different mtDNA regions (*SI Appendix, Table S5E*) as well as from the distribution of mutations occurring in one or two animals (*Fig. 2B*). While 3% of all hotspot sites were located in OriL, this region covers only 0.2% of mtDNA. While in muscle and oocytes we observed the highest hotspot frequency in the D-loop, all of the OriL hotspots were liver specific. Therefore OriL is the region with the highest hotspot frequency in liver (*SI Appendix, Fig. S13C*). Ten of 12 OriL hotspots disrupt proper pairing in the hairpin stem (*Fig. 2C* and *SI Appendix, Fig. S14*), potentially decreasing replication efficiency.

For de novo mutations in protein-coding regions, the hN/hS depended on the number of macaques in which a variant was found (*Fig. 2D* and *SI Appendix, Table S14*). The hN/hS was within, or close to, the range of neutral expectations (≤ 1.5 for all tissues and age groups) for de novo mutations present in one or two macaques. For mutations present in several macaques (three or more), in somatic tissues of most age groups, the hN/hS was higher than expected under neutrality. Therefore, our results suggest that protein-coding variants present in somatic tissues of multiple macaques evolve under positive selection. For oocytes, we could not compute the hN/hS ratio for sites mutated in multiple macaques due to the paucity of such sites.

Inheritable Heteroplasmies. In addition to de novo mutations, we investigated heteroplasmies present in at least one oocyte and all somatic tissues analyzed for the same animal (*SI Appendix, Table S15*) and, thus, very likely already inherited from the mother. Because of their usually higher MAFs, such heteroplasmies have a higher potential to be inherited by the offspring (“inheritable heteroplasmies”). To boost our statistical power, in addition to the samples used to study de novo mutations, here we considered heart and 33 oocytes with average DCS depth below 100×. In our data set, 10 animals harbored 17 sites with inheritable heteroplasmies (10, 0, 6, and 1 site(s) in young, intermediate 1, intermediate 2, and old animals, respectively; *SI Appendix, Table S15*), which were all transitions. Nine of the 17 sites were located in the D-loop, five in protein-coding regions, and three in tRNA.

Stronger genetic drift between oocytes and somatic tissues than between somatic tissues. MAFs of inheritable heteroplasmies correlated tightly between the somatic tissues studied, with a higher correlation between muscle and heart ($r = 0.970$), which are closely related ontogenetically, than between each of these tissues and liver ($r = 0.925$ and $r = 0.936$, respectively; *Fig. 3 A–C*). The MAF correlation between somatic tissues and oocytes was lower ($r = 0.910$; *Fig. 3D*). These results suggest a stronger genetic drift between oocytes and somatic tissues, an intermediate one between liver and muscle/heart, and a weaker one between muscle and heart.

Genetic drift increases with age in somatic tissues. We next tested whether the normalized variance in MAF (variance divided by $p(1-p)$, where p is the average MAF of the allele among single oocytes or across somatic tissues; refs. 24, 25) increases with age. The normalized variance in MAF across somatic tissues was positively correlated with age (Pearson $r = 0.556$, $P = 0.021$; *Fig. 3I*). The normalized variance in MAF for oocytes also increased with age, although this was not

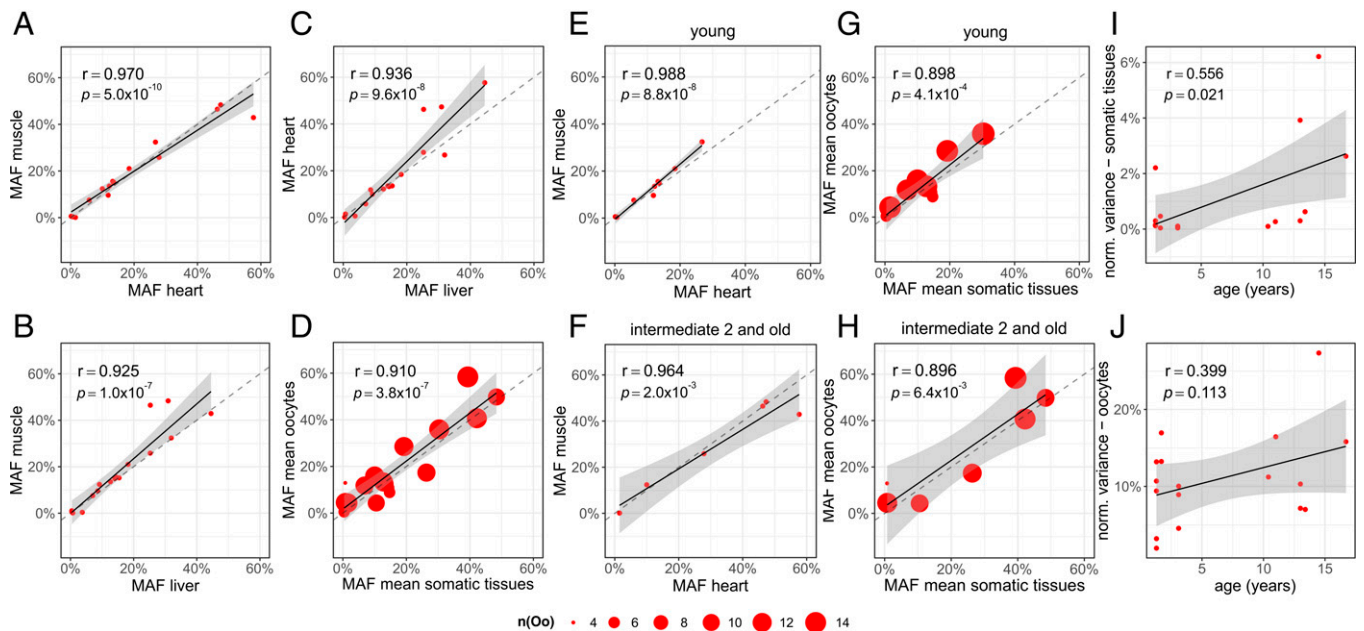


Fig. 3. Correlation of MAFs for inheritable heteroplasmies between different tissues and correlation of the normalized variance with age. (A) MAF of heteroplasmies in heart vs. muscle. (B) MAF of heteroplasmies in liver vs. muscle. (C) MAF of heteroplasmies in heart vs. liver. (D) MAF of heteroplasmies in somatic tissues (averaged among liver, muscle, and heart) vs. oocytes (averaged among all oocytes studied per animal). (E and F) Correlations of MAFs in muscle and heart, separated into young (E) and intermediate 2 plus old (F) age groups. Additional correlations between somatic tissues are in *SI Appendix, Fig. S15*. (G and H) Correlations of mean MAFs in somatic tissues and mean MAFs of all oocytes analyzed per animal, separately for young (G) and intermediate 2 plus old (H) age groups. (I and J) Correlation of the normalized variance of heteroplasmity MAF for somatic tissues (I) or single oocytes (J) with age. The gray bands are confidence intervals around the regression line, and the dashed lines are the 1:1 relationship. Dot size for oocyte plots reflects the number of sampled oocytes for each animal and heteroplasmic site.

significant (Pearson $r = 0.399$, $P = 0.113$; Fig. 3J). These observations suggest that genetic drift increases with age for somatic tissues and perhaps less so for oocytes.

Furthermore, we observed that the MAFs of inheritable heteroplasmies were more similar (i.e., more tightly correlated) between somatic tissues for young than for older (intermediate 2 and old combined) animals (Fig. 3 E and F and *SI Appendix*, Fig. S15), though the differences in correlations were not significant ($P = 0.413$, $P = 0.762$, and $P = 0.079$ for comparisons of muscle with heart, liver with heart, and muscle with liver, respectively; Fisher Z -test for difference between correlation coefficients). This was not observed between oocytes and somatic tissues, where correlation coefficients for MAFs of inheritable heteroplasmies were analogous in young and older animals (Fig. 3 G and H). This suggests that genetic drift increases with age for somatic tissues but not for oocytes. However, because of the small numbers of heteroplasmies analyzed per age group, we had only limited power to assess differences in genetic drift associated with age.

Does selection affect allele frequency of inheritable heteroplasmies?

We next analyzed whether selection influences the changes in MAFs for inheritable heteroplasmies. We did not observe a significant difference between the number of heteroplasmic variants with an increase ($n = 86$) and a decrease ($n = 84$) in MAFs in oocytes and somatic tissues ($P = 0.939$, binomial test; *SI Appendix*, Fig. S16A). This remained true when we accounted for a potential bias that heteroplasmy frequency introduces when analyzing the magnitude of shifts in MAFs (*SI Appendix*, Fig. S16B).

Tight effective germline bottleneck. To estimate the size of the effective germline bottleneck—the size required to explain observed genetic drift—for macaque mtDNA, we applied the population genetics approach described in Barrett et al. and Hendy et al. (25, 26) to the data on MAF shifts at 17 inheritable heteroplasmies. Namely, we compared allele frequencies between somatic tissues (averaged among three tissues per animal) and single oocytes (with 4 to 14 oocytes per animal) for each site separately (170 transmissions to oocytes in total). Because of the relatively small number of sites and samples examined, we could not rigorously test for linkage of variants (*SI Appendix*, Fig. S17). The effective bottleneck size was estimated to be 9.30 segregating mtDNA units (95% bootstrap confidence interval: 0.26–27.5) for all animals, 11.6 units (95% bootstrap confidence interval: 3.60–27.5) for young macaques, and 5.95 units (95% bootstrap confidence interval: 0.26–14.5) for older macaques (intermediate 2 and old combined; *SI Appendix*, Table S16). Thus, the bottleneck size was not significantly different between young and older macaques.

Discussion

Using the rhesus macaque as a model organism and utilizing duplex sequencing, we studied the age-related accumulation of germline and somatic de novo mtDNA mutations. We analyzed oocytes from animals who have not reached sexual maturity (occurring at ~ 3 y) and those almost up to the age of menopause (occurring at 25 y; ref. 27), thus sampling the whole reproductive lifespan of macaques. With duplex sequencing, we reliably detected de novo mutations occurring at MAFs $< 1\%$, a cutoff commonly used in conventional sequencing experiments (e.g., refs. 4, 8). Sequencing several somatic tissues and multiple single oocytes per animal allowed us to separate somatic and germline mutations and to analyze mutations arising at different stages of development. We took several experimental and computational

steps to minimize the contribution of Numts to our data set of de novo mutations (*SI Appendix*, Note S3). While we cannot completely exclude the Numts' contribution, our analysis suggests that they do not influence our results in any measurable way.

Age-Related Accumulation of Germline Mutations. We showed an increase in de novo mtDNA mutations with age directly in the primate germline. A recent study (13) demonstrated such an increase in mouse oocytes; however, because mutations were measured at only two time points, the shape of the dependence of mutation frequency on age could not be investigated. Here, by analyzing mutations in oocytes from macaques with ages ranging from 1 to 23 y, we could evaluate how mutation frequency depends on age. We observed that the germline mutation frequency increased for macaques younger than 9 y, with no further increase afterward. While we currently lack an explanation for the change in slope occurring at this particular age, in older animals, mitochondria with a high number of deleterious mutations might be removed by mitophagy (reviewed in ref. 28), or oocytes with high mtDNA mutational load might be eliminated by follicular atresia (reviewed in ref. 29).

The age-related accumulation of mtDNA germline mutations in macaque oocytes found here is consistent with indirect observations in humans, reporting an increased number of de novo mutations in children of older women (8). Another study, however, did not find an increase in the number of mutations in human oocytes with ovarian aging (12). Direct sequencing of oocytes from women of different ages is required to answer how mtDNA germline mutations accumulate with age in humans and whether such an accumulation exhibits a change in slope.

Based on the median mutation frequencies measured in young and old macaques (with an average age difference of 16.6 y), we computed the germline mutation rate of 8.7×10^{-7} mutations per site per generation (using a generation time of 11 y; ref. 30) and 7.9×10^{-8} mutations per site per year. The mutation rate per site per year is higher than that reported for humans [9.3×10^{-9} (ref. 4) and 1.6×10^{-8} (ref. 8)], using a generation time of 29 y], but lower than that reported in mice (7.6×10^{-7} ; ref. 13). These observations echo higher nuclear substitution rates in rodents than in primates (31) and in Old World monkeys than in hominoids (32).

Age-Related Somatic Mutation Accumulation Is Tissue Dependent.

Consistent with studies in humans and mice (e.g., refs. 4, 5, 8, 13, 23), we found an age-related increase in mtDNA mutations in the somatic tissues of macaques (*SI Appendix*, Fig. 1 B and C). Liver displayed the steepest increase in mutation frequency, followed by heart, and then by muscle; all three had a steeper increase compared to oocytes. Tissue-specific trends might be due to differences in the proliferation and regeneration speed (33) and/or in the rates of mitochondrial turnover (34, 35). Among the analyzed tissues, liver is the most proliferative; its cells are estimated to be replaced approximately once a year (36). Cells in the human heart are renewed at a rate of 4–17% per year (37). Skeletal muscle experiences low turnover and is largely postmitotic—its cells were shown to have an average age of 15 y (38). The analysis of a highly proliferative tissue such as the intestinal crypt, which is replaced every 5 d (36), would further aid in elucidating the role of cell proliferation in age-related mtDNA mutagenesis. Tissue-specific differences in human mtDNA mutation accumulation were reported previously (8, 21).

Molecular Mechanisms of mtDNA Mutagenesis. Several lines of evidence suggest that the mutational patterns observed here

are consistent with replication-associated errors as the primary source of mtDNA mutations, in agreement with other studies (e.g., refs. 4, 13, 23, 39, 40). First, we detected significant age-related increases in the frequencies of transitions, as well as in transition-to-transversion ratios. This signature is consistent with the propensity of DNA polymerase gamma, the main enzyme for mtDNA replication, for transition mutations (39). Second, the observation that liver, the most proliferative tissue analyzed, exhibited the highest transition-to-transversion ratios also points toward the contribution of mechanisms associated with mtDNA replication.

Third, spontaneous deamination of cytosine (C > T) and adenine (A > G), which leads to transitions (41), is another potential mechanism indirectly associated with mtDNA replication. The strong bias for G > A over C > T and for T > C over A > G mutations on the L-strand—also previously observed in humans and mice (e.g., refs. 13, 22, 23)—is consistent with a high incidence of C > T and A > G mutations on the H-strand and might be explained by its single-stranded status during the initial stages of replication (reviewed in ref. 42), facilitating spontaneous deamination. Alternatively, uncoupling of the leading and lagging strands during mtDNA synthesis (43) can increase the probability of oxidative DNA damage. In fact, mutational signatures of redox stress in yeast single-strand DNA and of aging in human mtDNA share common features (44).

We observed less-pronounced replication-related mtDNA mutagenesis with aging in oocytes than in somatic tissues. This suggests limited replication of mtDNA in aging oocytes. Whereas replication of mtDNA and nuclear DNA is not necessarily linked (45), mammalian oocytes do not undergo mitotic cell divisions after birth. Indeed, compared with somatic tissues, oocytes exhibited lower fold differences in transition rates between old and young macaques, had slower increase in the transition-to-transversion rate ratio with age, and had a weaker C > T over G > A strand bias.

Higher rates of C > T/G > A transitions at CpG than at non-CpG sites, particularly for older animals, point toward a role of spontaneous deamination of methylated cytosines (41) in mtDNA mutagenesis, despite the controversial reports regarding CpG methylation in mtDNA (e.g., refs. 46–49). Active cytosine deamination facilitated by Apolipoprotein B mRNA editing enzyme (APOBEC), previously shown to induce mutations in single-stranded DNA (50), does not seem to contribute to mtDNA mutagenesis in macaques. APOBEC targets Cs in a TC nucleotide context, with the highest specificity for the TCW nucleotide context (51). The analysis of the trinucleotide context in our data (*SI Appendix, Fig. S18A*) did not indicate an overrepresentation of C > T/G > A mutations within this context (*SI Appendix, Fig. S18B*).

OriL Is a Variant Hotspot in Liver of Aged Macaques. We found OriL to be a variant hotspot in liver of older macaques. OriL is essential for lagging-strand mtDNA replication: mtDNA-directed RNA polymerase initiates primer synthesis from the polyA stretch of the OriL's hairpin loop and is replaced by DNA polymerase gamma after ~25 nt of synthesis (5260606060606060–545). A previous analysis of 1,802 vertebrate species indicated a high conservation of OriL, particularly of the hairpin stem (55). Despite this, 10 out of 12 OriL variant hotspots we identified likely destabilize the hairpin stem (*SI Appendix, Fig. S14*); mutations that lower the stability of the OriL hairpin structure can decrease lagging-strand replication (52).

In our study, the high mutation frequencies in OriL were unique to liver. In proofreading-deficient mice, mutational load

was lower in OriL than in other mtDNA regions (55), similar to our results for muscle and oocytes. OriL might be a mutation hotspot because it assumes a non-B structure, and such structures were recently shown to increase mutation rates (56). Liver in particular might be affected because it is highly proliferative, and replication errors were suggested to be the primary driver behind increased mutagenesis at non-B DNA (56). Alternatively (or additionally), mutations in OriL, which likely decrease mtDNA replication efficiency, might provide some advantage to aged liver in macaques and are therefore selected for. Tissue-specific positive selection of mtDNA variants in the D-loop, also potentially affecting mtDNA replication efficiency, was reported previously in human liver, muscle, and kidney (21, 57).

Selection at Protein-Coding Variants. In both liver and skeletal muscle, the hN/hS ratios at protein-coding regions were higher than neutral expectations, with particularly high ratios in older animals, suggesting positive selection. This pattern, though already observed for mutations found in one or two macaques, was mainly driven by mutations at tissue-specific variant hotspots, reinforcing its selective nature. Positive selection for protein-coding variants was previously observed in human liver and was suggested to reduce mitochondrial function to decrease damaging byproducts of mitochondrial metabolism (21). Our study suggests that a similar phenomenon might be operating in primate skeletal muscle, a less proliferative tissue. Furthermore, the oxidative phosphorylation complexes were previously suggested as a potential determinant of selective pressure for mutations in cancer (58). In our data, we also observed differences in mutation frequencies among the mitochondrial complexes, with a higher mutation frequency in complex III in liver of old macaques (*SI Appendix, Fig. S20*).

In agreement with findings for mouse oocytes (13), we did not observe strong evidence of selection acting on de novo mutations in macaque oocytes. Interestingly, in the intermediate 1 age group, we observed a decrease in the hN/hS ratio compared to all other age groups. Indeed, the hN/hS for this group was 0.72, suggestive of purifying selection. Purifying selection might play a role in age-related mutation accumulation for this age group. However, this is unlikely to be the case because purifying selection is not observed in the intermediate 2 and old age groups and, thus, does not appear to be the force keeping mutation frequency relatively constant for these groups. The hN/hS ratio measured for oocytes for any age group was higher than the average pN/pS ratio (the nonsynonymous-to-synonymous rate ratio for homoplasmic polymorphisms among animals, treating polymorphisms at the same position in different animals as separate events), which was equal to 0.19 and was consistent with purifying selection on polymorphisms. Purifying selection was also suggested to act on transmitted variants in the human germline (5, 8). Thus, purifying selection might be acting on polymorphisms and transmitted variants, but not (or much less) on de novo mtDNA mutations in the germline. Alternatively, we might lack power to detect selection in oocytes due to the relatively small number of mutations detected.

Inheritable Heteroplasmies. We estimated the effective bottleneck in the macaque germline to be severe (i.e., 9.30 segregating units; 95% confidence interval: 0.26–27.48) and similar to 7–10 mtDNA segregating units recently reported in humans (8). No significant differences in mtDNA germline bottleneck size were observed between young and older macaques. We observed similar correlations of heteroplasmy MAFs between oocytes and somatic tissues in young and older macaques, again

pointing to no differences in the bottleneck size. The correlation of the normalized variance of MAFs with age was not significant, albeit positive, for oocytes. These results contradict recent observations in humans suggesting that the size of the germline bottleneck decreases, and mtDNA divergence in MAFs between mother and offspring increases, with the mother's age at childbirth (8, 59). The discrepancies between our results and these human studies might be due to a relatively small number of inheritable heteroplasmies examined in our study or to a greater reproductive lifespan investigated for humans.

We detected an increase of random genetic drift with age in somatic tissues. The pairwise correlations of heteroplasmy MAFs between any two somatic tissues compared were higher in young than in older macaques, similar to findings in human children versus their mothers (4). Additionally, we found a significant positive correlation between age and the normalized variance in MAFs of liver and of skeletal muscle in macaques, echoing an observation for human hair (25). In contrast, no significant differences in the normalized variance in MAFs were observed between mouse mothers and pups (13), but they were separated by only ~10 mo, likely explaining the difference with our results.

Materials and Methods

Sample Preparation and Duplex Sequencing. Single oocytes were isolated and lysed. Total DNA was extracted from somatic tissues. The enrichment for circular mtDNA was performed with Exonuclease V digestion of linear nuclear DNA and estimated with real-time PCR. Duplex sequencing was performed as described previously (13). SSCS and DCS consensus formation was performed with Du Novo (20), and DCSs were mapped to the macaque reference sequence. Variants were called and filtered, and de novo mutations were identified (see *SI Appendix, Materials and Methods* for details). The procedures to minimize and analyze the potential bias from Numts are summarized in *SI Appendix, Note S3*.

Analysis of Age Effects, Selection, Variant Hotspots, and Inheritable Heteroplasmies. To analyze the effects of age on mutation frequencies, we used generalized mixed-effects linear models (with binomial family, logit link,

and a breakpoint for oocytes). To analyze selection, we computed the hN/hS ratio, as described in previous studies (refs. 8, 13, 21). Variant hotspots were identified as described in *SI Appendix, Note S2*. The MAFs for inheritable heteroplasmies were used to estimate the effective size of the germline bottleneck as described in Arbeituber et al. (13). All statistical tests were corrected for multiple testing (see *SI Appendix, Materials and Methods* for details).

Data Availability. Sequencing data have been deposited to the Sequence Read Archive (BioProject ID [PRJNA777828](https://doi.org/10.1038/s41588-021-0778-2)) (60). Our code, as well as tables with all samples and identified mutations, are available on GitHub: <https://github.com/makovalab-psu/macaque-duplexSeq>. All other study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We are grateful to Nicholas Stoler for developing the duplex sequencing instance on Galaxy and for advice on the analysis, to Irene Tiemann-Boege for helpful discussions regarding the development of the library preparation protocols, and to Kristin Eckert and Suzanne Hile for their help in establishing the duplex sequencing method in the Makova lab. We are grateful to Southwest (Jera Pecotte), Oregon (Wendy Price), Wisconsin (Heather Simmons), and Washington (Chris English) National Primate Research Centers for collecting samples for this study. Sequencing was performed by the Penn State Genomics Core Facility, University Park, PA. This project was supported by a grant from NIH (R01GM116044) to K.D.M. and a Schrödinger Fellowship from the Austrian Science Fund (FWF) to B.A. (J-4096). Additional funding was provided by the Office of Science Engagement, Eberly College of Sciences, and The Huck Institute of Life Sciences and the Institute for Computational and Data Sciences at Penn State. This research was supported in part by the National Institutes of Health Grant P51OD011092 to the Oregon National Primate Research Center.

Author affiliations: ^aDepartment of Biology, The Pennsylvania State University, University Park, PA 16802; ^bExperimental Gynaecology, Obstetrics and Gynaecological Endocrinology, Kepler University Hospital Linz, Johannes Kepler University Linz, 4020 Linz, Austria; ^cDepartment of Operations and Decision Systems, Université Laval, Québec, QC G1V0A6, Canada; ^dPopulation Health and Optimal Health Practices, CHU de Québec - Université Laval Research Center, Québec, QC G1V4G2, Canada; ^eCenter for Medical Genomics, The Pennsylvania State University, University Park, PA 16802; ^fDepartment of Animal Science, The Pennsylvania State University, University Park, PA 16802; ^gDepartment of Statistics, The Pennsylvania State University, University Park, PA 16802; and ^hInstitute of Economics and EMbeDS, Sant'Anna School of Advanced Studies, Pisa 56127, Italy

- N. Pfanner, B. Warscheid, N. Wiedemann, Mitochondrial proteins: From biogenesis to functional networks. *Nat. Rev. Mol. Cell Biol.* **20**, 267–284 (2019).
- B. Pakendorf, M. Stoneking, Mitochondrial DNA and human evolution. *Annu. Rev. Genomics Hum. Genet.* **6**, 165–183 (2005).
- E. A. Shoubridge, T. Wai, "Mitochondrial DNA and the mammalian oocyte" in *The Mitochondrion in the Germline and Early Development*, J. C. St. John, Ed. (Elsevier, 2007), pp. 87–111.
- B. Rebollo-Jaramillo et al., Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15474–15479 (2014).
- W. Wei et al.; NIH Rare Disease Program; 100,000 Genomes Project-Rare Diseases Pilot, Germline selection shapes human mitochondrial DNA diversity. *Science* **364**, eaa6520 (2019).
- R. Rahbari et al.; UK10K Consortium, Timing, rates and spectra of human germline mutation. *Nat. Genet.* **48**, 126–133 (2016).
- R. J. Wang et al., Paternal age in rhesus macaques is positively associated with germline mutation accumulation but not with measures of offspring sociability. *Genome Res.* **30**, 826–834 (2020).
- A. A. Zaidi 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).
- D. T. Brown, D. C. Samuels, E. M. Michael, D. M. Turnbull, P. F. Chinnery, Random genetic drift determines the level of mutant mtDNA in human primary oocytes. *Am. J. Hum. Genet.* **68**, 533–536 (2001).
- N. Gigarel et al., Poor correlations in the levels of pathogenic mitochondrial DNA mutations in polar bodies versus oocytes and blastomeres in humans. *Am. J. Hum. Genet.* **88**, 494–498 (2011).
- M. Ancora et al., Complete sequence of human mitochondrial DNA obtained by combining multiple displacement amplification and next-generation sequencing on a single oocyte. *Mitochondrial DNA A. DNA Mapp. Seq. Anal.* **28**, 180–181 (2017).
- L. Boucrot et al., Deep sequencing shows that oocytes are not prone to accumulate mtDNA heteroplasmic mutations during ovarian ageing. *Hum. Reprod.* **32**, 2101–2109 (2017).
- B. Arbeituber et al., Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues. *PLoS Biol.* **18**, e3000745 (2020).
- L. Craven, C. L. Alston, R. W. Taylor, D. M. Turnbull, Recent advances in mitochondrial disease. *Annu. Rev. Genomics Hum. Genet.* **18**, 257–275 (2017).
- P. Kapahi, M. E. Boulton, T. B. Kirkwood, Positive correlation between mammalian life span and cellular resistance to stress. *Free Radic. Biol. Med.* **26**, 495–500 (1999).
- G. S. Roth et al., Aging in rhesus monkeys: Relevance to human health interventions. *Science* **305**, 1423–1426 (2004).
- R. G. Rawlins, M. J. Kessler, Climate and seasonal reproduction in the Cayo Santiago macaques. *Am. J. Primatol.* **9**, 87–99 (1985).
- E. P. Walker, R. M. Nowak, J. L. Paradiso, *Walker's Mammals of the World* (John Hopkins University Press, 1983).
- M. W. Schmitt et al., Detection of ultra-rare mutations by next-generation sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14508–14513 (2012).
- N. Stoler et al., Family reunion via error correction: An efficient analysis of duplex sequencing data. *BMC Bioinformatics* **21**, 96 (2020).
- M. Li, R. Schröder, S. Ni, B. Madea, M. Stoneking, Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2491–2496 (2015).
- S. L. Williams, D. C. Mash, S. Züchner, C. T. Moraes, Somatic mtDNA mutation spectra in the aging human putamen. *PLoS Genet.* **9**, e1003990 (2013).
- S. R. Kennedy, J. J. Salk, M. W. Schmitt, L. A. Loeb, Ultra-sensitive sequencing reveals an age-related increase in somatic mitochondrial mutations that are inconsistent with oxidative damage. *PLoS Genet.* **9**, e1003794 (2013).
- C. D. Millar et al., Mutation and evolutionary rates in adélie penguins from the antarctic. *PLoS Genet.* **4**, e1000209 (2008).
- A. Barrett et al., Pronounced somatic bottleneck in mitochondrial DNA of human hair. *Phil. Trans. R. Soc. B* **375**, 20190175 (2019).
- M. D. Hendy, M. D. Woodhams, A. Dodd, Modelling mitochondrial site polymorphisms to infer the number of segregating units and mutation rate. *Biol. Lett.* **5**, 397–400 (2009).
- S. Tardif, A. Carville, D. Elmore, L. E. Williams, K. Rice, "Reproduction and breeding of nonhuman primates" in *Nonhuman Primates in Biomedical Research*, C. R. Abbe, K. Mansfield, S. Tardif, T. Morris, Eds. (Elsevier, 2012), pp. 197–249.
- Q. Shen, Y. Liu, H. Li, L. Zhang, Effect of mitophagy in oocytes and granulosa cells on oocyte quality. *Biol. Reprod.* **104**, 294–304 (2021).
- P. May-Panloup et al., Ovarian ageing: The role of mitochondria in oocytes and follicles. *Hum. Reprod. Update* **22**, 725–743 (2016).
- C. Xue et al., The population genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome sequences. *Genome Res.* **26**, 1651–1662 (2016).
- S. J. Lindsay, R. Rahbari, J. Kaplanis, T. Keane, M. E. Hurles, Similarities and differences in patterns of germline mutation between mice and humans. *Nat. Commun.* **10**, 4053 (2019).
- P. Moorjani, C. E. G. Amorim, P. F. Arndt, M. Przeworski, Variation in the molecular clock of primates. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10607–10612 (2016).
- S. E. Iismaa et al., Comparative regenerative mechanisms across different mammalian tissues. *NPJ Regen. Med.* **3**, 6 (2018).
- A. R. Menzies, P. H. Gold, The turnover of mitochondria in a variety of tissues of young adult and aged rats. *J. Biol. Chem.* **246**, 2425–2429 (1971).

35. S. Miwa, C. Lawless, T. von Zglinicki, Mitochondrial turnover in liver is fast in vivo and is accelerated by dietary restriction: Application of a simple dynamic model. *Aging Cell* **7**, 920–923 (2008).
36. J. L. Kopp, M. Grompe, M. Sander, Stem cells versus plasticity in liver and pancreas regeneration. *Nat. Cell Biol.* **18**, 238–245 (2016).
37. O. Bergmann *et al.*, Dynamics of Cell Generation and Turnover in the Human Heart. *Cell* **161**, 1566–1575 (2015).
38. K. L. Spalding, R. D. Bhardwaj, B. A. Buchholz, H. Druid, J. Frisén, Retrospective birth dating of cells in humans. *Cell* **122**, 133–143 (2005).
39. W. Zheng, K. Khrapko, H. A. Collier, W. G. Thilly, W. C. Copeland, Origins of human mitochondrial point mutations as DNA polymerase gamma-mediated errors. *Mutat. Res.* **599**, 11–20 (2006).
40. J. H. K. Kauppila *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).
41. D. A. Kreuzer, J. M. Essigmann, Oxidized, deaminated cytosines are a source of C → T transitions in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 3578–3582 (1998).
42. M. Falkenberg, Mitochondrial DNA replication in mammalian cells: Overview of the pathway. *Essays Biochem.* **62**, 287–296 (2018).
43. I. J. Holt, H. E. Lorimer, H. T. Jacobs, Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* **100**, 515–524 (2000).
44. N. P. Degtyareva *et al.*, Mutational signatures of redox stress in yeast single-strand DNA and of aging in human mitochondrial DNA share a common feature. *PLoS Biol.* **17**, e3000263 (2019).
45. A. Kirillova, J. E. J. Smits, G. T. Sukhikh, I. Mazunin, The role of mitochondria in oocyte maturation. *Cells* **10**, 2484 (2021).
46. M. Mechtal, L. R. Ingerslev, O. Fabre, M. Picard, R. Barrès, Evidence suggesting absence of mitochondrial DNA methylation. *Front. Genet.* **8**, 166 (2017).
47. B. Liu *et al.*, CpG methylation patterns of human mitochondrial DNA. *Sci. Rep.* **6**, 23421 (2016).
48. L.-H. Fan *et al.*, Absence of mitochondrial DNA methylation in mouse oocyte maturation, aging and early embryo development. *Biochem. Biophys. Res. Commun.* **513**, 912–918 (2019).
49. V. Patil *et al.*, Human mitochondrial DNA is extensively methylated in a non-CpG context. *Nucleic Acids Res.* **47**, 10072–10085 (2019).
50. R. Suspène *et al.*, Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4858–4863 (2011).
51. S. A. Roberts *et al.*, An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* **45**, 970–976 (2013).
52. J. M. Fusté *et al.*, Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol. Cell* **37**, 67–78 (2010).
53. D. P. Tapper, D. A. Clayton, Mechanism of replication of human mitochondrial DNA. Localization of the 5' ends of nascent daughter strands. *J. Biol. Chem.* **256**, 5109–5115 (1981).
54. S. Wanrooij *et al.*, Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11122–11127 (2008).
55. S. Wanrooij *et al.*, In vivo mutagenesis reveals that Oril is essential for mitochondrial DNA replication. *EMBO Rep.* **13**, 1130–1137 (2012).
56. W. M. Guiblet *et al.*, Non-B DNA: A major contributor to small- and large-scale variation in nucleotide substitution frequencies across the genome. *Nucleic Acids Res.* **49**, 1497–1516 (2021).
57. D. C. Samuels *et al.*, Recurrent tissue-specific mtDNA mutations are common in humans. *PLoS Genet.* **9**, e1003929 (2013).
58. A. N. Gorelick *et al.*, Respiratory complex and tissue lineage drive recurrent mutations in tumour mtDNA. *Nat. Metab.* **3**, 558–570 (2021).
59. P. R. Wilton, A. Zaidi, K. Makova, R. Nielsen, A population phylogenetic view of mitochondrial heteroplasmy. *Genetics* **208**, 1261–1274 (2018).
60. B. Arbeithuber, Raw sequencing reads. Sequence Read Archive. <https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA77828>. Deposited 8 October 2021.