OPEN

Sharing of heteroplasmies between human liver lobes varies across the mtDNA genome

Alexander Hübner [1], Manja Wachsmuth[1], Roland Schröder[1], Mingkun Li[2], Anna Maria Eis-Hübinger[3], Burkhard Madea[4] & Mark Stoneking[1]

Mitochondrial DNA (mtDNA) heteroplasmy (intra-individual variation) varies among different human tissues and increases with age, suggesting that the majority of mtDNA heteroplasmies are acquired, rather than inherited. However, the extent to which heteroplasmic sites are shared across a tissue remains an open question. We therefore investigated heteroplasmy in two liver samples (one from each primary lobe) from 83 Europeans, sampled at autopsy. Minor allele frequencies (MAF) at heteroplasmic sites were significantly correlated between the two liver samples from an individual, with significantly more sharing of heteroplasmic sites in the control region than in the non-control region. We show that this increased sharing for the control region cannot be explained by recent mutations at just a few specific heteroplasmic sites or by the possible presence of 7S DNA. Moreover, we carried out simulations to show that there is significantly more sharing than would be predicted from random genetic drift from a common progenitor cell. We also observe a significant excess of non-synonymous vs. synonymous heteroplasmies in the protein-coding region, but significantly more sharing of synonymous heteroplasmies. These contrasting patterns for the control vs. the non-control region, and for non-synonymous vs. synonymous heteroplasmies, suggest that selection plays a role in heteroplasmy sharing.

The mitochondrial genome is present in many copies in a single cell, and intra-individual variation in the mitochondrial genome of an individual is called mtDNA heteroplasmy[1,2]. In humans, it has been shown that detrimental mtDNA mutations are usually present in a heteroplasmic state at low frequencies, with high frequencies of the deleterious allele leading to functional defects and a disease phenotype[1,3–6]. In addition, heteroplasmy is a general phenomenon in aging individuals, where the minor allele is present at rather low frequencies (often below 4%) and many of the affected sites are part of the control region[4]. The total number of heteroplasmic sites strongly correlates with age and several studies have shown that heteroplasmic sites are tissue specific[7–12], i.e. sites which are frequently heteroplasmic in one tissue are homoplasmic in all other tissues of the same individual.

The tissue specificity of heteroplasmic sites and the association between the number of heteroplasmies and age would suggest that the majority of heteroplasmies are not inherited from the previous generation but are acquired during the lifetime of an individual in a tissue-dependent manner. This raises the following question: do cells from the same tissue share a similar profile of heteroplasmic sites despite being separated from a common progenitor cell for many cell divisions? Studies on mouse embryos containing two different mtDNA haplotypes have shown that mtDNA segregation occurs rapidly between generations and the distribution of mtDNA haplotypes in the F1 generation resembles the pattern expected under random genetic drift[13]. This result is expected when the underlying heteroplasmies are evolving neutrally. However, it has been shown that mtDNA variants do not behave fully neutrally[14,15] and more recent studies on human cells indicated that the observed variance in heteroplasmic levels between cells is less stochastic than expected by random genetic drift[16–18]. This suggests that there might be further population genetic forces, e.g. within-cell mtDNA population structure[19,20], exchange of mtDNA between cells[16] or selection, that shape the distribution of heteroplasmies within as well as between tissues.

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103, Leipzig, Germany. [2]Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, 100101, Beijing, China. [3]Institut für Virologie, Universitätsklinikum Bonn, Sigmund-Freud-Str. 25, D-53105, Bonn, Germany. [4]Institut für Rechtsmedizin, Universitätsklinikum Bonn, Stiftsplatz 12, D-53111, Bonn, Germany. Alexander Hübner and Manja Wachsmuth contributed equally. Correspondence and requests for materials should be addressed to A.H. (email: alexander_huebner@eva.mpg.de) or M.S. (email: stonekg@eva.mpg.de)

While most age-related heteroplasmies occur in the control region, human liver tissue is unusual in showing an excess of heteroplasmies involving non-synonymous mutations in the mtDNA protein-coding genes[8]. This result is remarkable because these protein-coding region mutations are likely to have a functional effect[8], and protein-coding region mutations are strongly selected against during transmission from one generation to another[21]. Thus, it seems that mtDNA in human liver tissue exhibits a relaxation of purifying selection and age-related positive selection for somatic mutations that decrease mitochondrial function[8]. This makes liver a good candidate tissue to analyse the sharing of heteroplasmic sites with respect to different evolutionary forces.

While some studies have investigated the amount of variation in levels of heteroplasmy in cells that arose from a single ancestor cell in cell culture[16,17], to date there has been no such investigation comparing heteroplasmy across a tissue. We therefore obtained one blood sample and two liver samples (one from each primary lobe) from 83 Europeans, sampled at autopsy. We chose liver because it consists of two distinct lobes, which makes sampling different parts of the tissue straightforward and because a high number of heteroplasmies was previously reported in liver samples with an excess of non-synonymous over synonymous heteroplasmies in protein-coding genes[8]. MtDNA heteroplasmy was evaluated by capture-enrichment sequencing[8,22–24], and we analysed sharing of mtDNA heteroplasmy between the liver lobes for different regions of the mitochondrial genome.

Results

Heteroplasmy sharing within the liver. We investigated mtDNA heteroplasmy in liver and blood tissue samples of 83 individuals. For liver, two samples taken from different lobes were analysed in order to compare the heteroplasmic pattern in different parts of the tissue. The samples were capture-enriched for mtDNA and sequenced to an average sequencing depth of 1,175-fold for blood samples and 2,640-fold for liver samples. Applying a threshold of 2.5% MAF, we detected 541 heteroplasmic sites at 343 different positions (Supplementary Table S1). More heteroplasmic sites were observed in the non-control region (sites 577 to 16,023 in the revised Cambridge reference sequence (rCRS)) for liver (280 sites) compared to blood (64 sites), but the most abundant heteroplasmic sites in liver were in the control region (site 72: 67 individuals, site 60: 26 individuals, site 94: 20 individuals), which were only rarely observed to be heteroplasmic in blood (site 72: 1 individual, site 60: 1 individual, site 94: no individuals). These data are in accordance with results from a previous study[8] (Supplementary Fig. S1), indicating that heteroplasmy is tissue specific, with different individuals exhibiting similar heteroplasmic patterns.

Virological tests revealed that three individuals had active hepatitis B virus infection, one had active hepatitis C virus infection and one individual was HIV positive, with low viral load (Supplementary Table S2). Those individuals were kept in all downstream analyses, as the number of positive cases was too low to analyse separately, and they were not outliers with respect to the number or type of heteroplasmies detected. There was no effect of liver fat content on either the total number or the MAFs of heteroplasmic sites (Supplementary Fig. S2, p > 0.05) and the mtDNA copy numbers, estimated for each liver sample as described before[25], were highly correlated between corresponding liver samples of an individual, suggesting no functional differences between the liver lobes (Supplementary Fig. S3, r = 0.81, p < 0.001). Post-mortem interval (i.e., the interval between death of the individual and autopsy) varied from 1–11 days (Supplementary Table S2). To test for an effect of post-mortem interval on heteroplasmy detection, we compared the number of heteroplasmies in individuals who were transported the same day as death (n = 68) to those who were transported more than one day after death (n = 15); there was no significant difference between the two groups in number of heteroplasmies detected (Supplementary Table S2; two-sided Mann Whitney U-Test p > 0.05), in agreement with a previous study[8].

We next asked whether heteroplasmy was correlated between the three different samples from an individual. For blood and each liver sample, there is a low but significant correlation (r = 0.35 and 0.38, p < 0.001, Supplementary Fig. S4) with only 7.5% of the heteroplasmies shared between the tissues. However, the correlation between the two liver samples was higher (Fig. 1a, r = 0.90, p < 0.001). While 355 sites were heteroplasmic in only one of the two liver samples of an individual, 136 sites (28%) were heteroplasmic in both liver samples and these exhibited similar MAFs (Fig. 1a). Moreover, there were more shared heteroplasmies from the control region than from the rest of the genome (Table 1, p < 0.001): 55% of control region heteroplasmies were shared, vs. 7.5% of heteroplasmies outside the control region. The high correlation between the MAFs in the two liver samples was mainly driven by the control region (Fig. 1b, r = 0.90, p < 0.001), while there was a lower but still significant correlation outside the control region (Fig. 1c, r = 0.29, p < 0.001). The difference in correlation coefficients is significant, based on random partitions of all of the heteroplasmic sites into two sets with the same number of sites as observed for the control region and the non-control region (p < 0.001, Supplementary Fig. S5). Thus, not only are more heteroplasmies shared in the control region than outside the control region, but control region heteroplasmies also have more similar MAFs.

Potential mechanisms underlying the more frequent sharing in the control region. *Frequent heteroplasmic sites.* There are three sites that are frequently heteroplasmic in liver (Supplementary Table S1) and all are in the control region (sites 60, 72, and 94). To determine if these three sites are driving the higher correlation in MAF in the control region, we analysed them separately. While these three sites did indeed show a high correlation in MAF between the two liver samples (Fig. 2a, r = 0.91, p < 0.001), the remaining sites in the control region still showed a significant correlation (Fig. 2b, r = 0.82, p < 0.001) that is higher than the correlation for the non-control region (Fig. 1c). To determine if the difference in correlations for the three sites vs. the remaining sites in the control region was statistically significant, we performed a subsampling test. As there were 113 occurrences of heteroplasmy at these three sites, we sampled 113 heteroplasmies at random from the control region, calculated the correlation in MAF between the two liver samples, and repeated this 1000 times. The r-value for the three sites was not significantly higher than the r-values for the random subsamples (Fig. 2c), indicating that
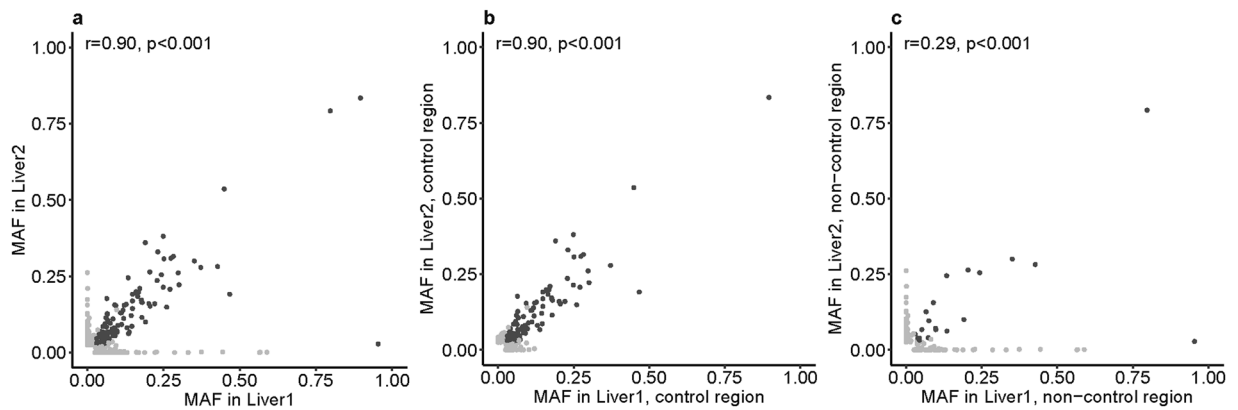
Figure 1. Correlation of MAFs at heteroplasmic sites in liver lobes. Each dot is one heteroplasmic site in one individual. Pearson's correlation coefficient r is given. Heteroplasmic sites are compared in (**a**) liver sample 1 and 2, (**b**) the control region of liver sample 1 and 2, (**c**) the non-control region of liver sample 1 and 2. Heteroplasmic sites shared between liver sample 1 and 2 are in dark grey, non-shared sites in light grey.

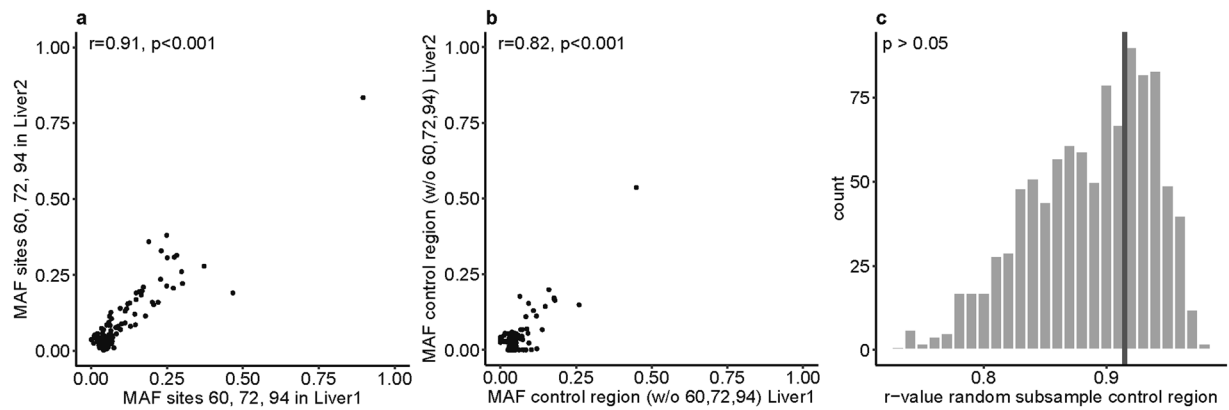| heteroplasmies in liver1 and liver2	control region	non-control region	total
shared	115	21	136
not shared	95	260	355
total	210	281	

Table 1. Heteroplasmic mutations in the control region and the non-control region and shared and non-shared heteroplasmies. $p < 0.001$ (two-sided Fisher's exact test).

the higher correlation between MAF for the control region than for the rest of the genome is not driven solely by these three sites.

7S DNA.   The control region includes the D-loop region, in which a third strand, the 7S DNA, displaces the heavy strand and binds to the light strand. Inferred heteroplasmies in the D-loop region might therefore reflect mutations in the 7S DNA rather than mutations in the mtDNA itself. To see if sequences from 7S DNA were likely to be present in the sequencing libraries, we estimated the relative mtDNA copy number from the capture-enrichment sequencing coverage of the D-loop region and the rest of the mtDNA molecule separately, as described before[25]. As the 7S DNA has several starting and end points, we used the outer limits reported in the literature, namely from site 16,097 to site 191[26,27]. The D-loop did not exhibit a higher copy number than the other parts of the mtDNA genome, indicating that 7S DNA is unlikely to be present in the DNA libraries ($p = 0.30$ and 0.41 for blood and liver, respectively, Supplementary Fig. S6A). Furthermore, the correlation between the MAF for the two liver samples is almost as high for the D-loop region as it for the rest of the control region (Supplementary Fig. S6B,C: $r = 0.90$ vs. $r = 0.89$). Hence, the significant correlation in MAFs in the control region is likely a phenomenon of the entire control region.

Higher mutation rate.   We also tested if heteroplasmic sites showed a higher mutation rate than non- heteroplasmic sites by comparing the number of heteroplasmies to the inferred mutation rate at each site, based on observed polymorphism data[28]. Heteroplasmic sites had significantly higher inferred mutation rates than sites that were not heteroplasmic ($p < 0.001$, Supplementary Fig. S7) with the mutation rate being higher in the control region than outside the control region. However, mutation rates did not differ between shared vs. non-shared heteroplasmies ($p > 0.05$, Supplementary Fig. S7), suggesting that the mutation rate does not increase the probability of heteroplasmies to be shared.

Random genetic drift.   If control region heteroplasmies are more likely to have arisen in a common progenitor cell of the cells sampled in the two liver lobes, then the higher correlation in MAF might reflect this common ancestry. We therefore tested if the correlation in MAF is compatible with the expected amount of random genetic drift from a common progenitor cell by simulating mtDNA transmission as a two-step model (Fig. 3a) following[16]. We started with different MAFs at the initial generation n_0 (5%, 10%, 20%, 50%) and different constant mtDNA copy numbers (1,000, 2,500, and 5,000) and sampled the difference in MAF 10, 20, 50, 100, and 250 generations after the initial split of the cells (Fig. 3b). For all simulations, we observed a higher MAF difference with an increasing number of cell replications. This effect was more pronounced for higher MAF at generation n_0 and lower mtDNA copy number. Based on the observed mtDNA copy number (Fig. 3c; mean = 2,528) and MAF of shared heteroplasmies with the same consensus allele (Fig. 3d; mean = 10.8%), the most similar corresponding simulation is with a copy number of 2,500 and a MAF at generation n_0 of 10%; for this simulation, the variance in observed MAF differences was smaller than would be expected after 50 generations of random genetic drift

Figure 2. Heteroplasmy sharing in liver samples and correlation with single sites. MAFs per site in liver sample 1 and 2 for: (**a**) sites 60, 72 and 94; and (**b**) for all other heteroplasmic sites in the control region. Each dot is one heteroplasmic site in one individual. Pearson's correlation coefficient r is given. (**c**) Distribution of r-values for correlation of liver sample 1 and 2 MAFs per site for random subsamples of heteroplasmies in the control region (same number as the sum of sites 60, 72 and 94). The r-value for sites 60, 72 and 94 only is shown as a dark grey bar.

(one-sided F-test for the equality of two variances, $p < 0.001$). With an increasing number of cell replications, the simulated MAF difference further diverged from the observed values, which suggests that random genetic drift cannot explain the observed sharing.

Differences in MAF between corresponding liver samples are not correlated with age. We next investigated the influence of age on heteroplasmy sharing between different liver regions of an individual. Overall, the total number of different heteroplasmic sites in an individual increased with age for both the control region and the non-control region (Supplementary Fig. S8A, $r = 0.42$ and adjusted $p < 0.001$ both within and outside of the control region). However, the MAF at heteroplasmic sites did not increase with age (Supplementary Fig. S8B, $r = 0.09$ for the control region, -0.01 for outside of the control region, adjusted $p > 0.05$). Although some heteroplasmies exhibit high MAFs only at ages above 50, many sites remain at low frequencies even at older ages (Supplementary Fig. S8B). The hypothesis that random genetic drift has an effect on the difference in MAF of two corresponding liver samples would suggest that with increasing age the difference in MAF would increase, too. Yet, when testing for a correlation between the difference in MAF between two corresponding liver samples and age of the individual, we did not observe a significant correlation either within or outside the control region (Fig. 4, $r = 0.06$ within control region, $r = -0.01$ outside the control region, both $p > 0.05$). In order to test whether, in contrast, specific sites have a significant correlation between the MAF difference and age, we separately calculated the correlation between MAF difference and age for the three most frequent heteroplasmy sites (sites 60, 72, and 94), for which at least 20 individuals were heteroplasmic (Supplementary Fig. S9). The correlation between MAF and age was not significant for any of these three sites (adjusted $p > 0.05$), which further supports the hypothesis that the difference in MAF does not increase with age, contrary to what would be expected from random genetic drift.

Synonymous heteroplasmies are more often shared than non-synonymous ones. Previous studies showed that liver has a significant excess of non-synonymous heteroplasmies which are predicted to have an impact on function[8]. We investigated this by calculating the ratio of non-synonymous heteroplasmies per non-synonymous site vs. synonymous heteroplasmies per synonymous site (hN/hS) across all protein-coding genes. In the absence of any selection this ratio has an expected value of 1; purifying selection results in values less than 1 and positive selection results in values greater than 1. As found previously[8], the hN/hS ratio is significantly greater than 1 (adjusted $p < 0.05$, Supplementary Fig. S10A), indicating positive selection for non-synonymous heteroplasmies rather than relaxation of functional constraints[8]. Additionally, we observed a strong increase in the number of heteroplasmies occurring in the non-control region compared to the control region in individuals older than 60 years (Supplementary Fig. S8A). We therefore investigated whether the excess in non-synonymous heteroplasmies is age-dependent (Supplementary Fig. S10B). While both synonymous and non-synonymous heteroplasmies increased in number with increasing age, only non-synonymous heteroplasmies showed a significant difference between the age groups below 60 years and above 60 years (two-sided Mann-Whitney U test, adjusted $p < 0.05$). We then asked if either synonymous or non-synonymous heteroplasmic sites were more likely to be shared between different liver samples. Although there were more than twice as many non-synonymous than synonymous heteroplasmies in the data set (163 vs. 62), there were significantly more synonymous sites shared than non-synonymous sites (11 vs. 6, Table 2). Accordingly, the median MAF difference between corresponding liver samples was on average higher for non-synonymous heteroplasmies than synonymous ones (median MAF difference of 4.4% to 3.8%), albeit not significantly so ($p > 0.05$, Supplementary Fig. S11A). These results suggest that non-synonymous heteroplasmies more often arise independently in different cells. However, neither non-synonymous nor synonymous heteroplasmies showed a significant correlation for the difference in MAF between corresponding liver samples with age ($p > 0.5$, Supplementary Fig. S11B).
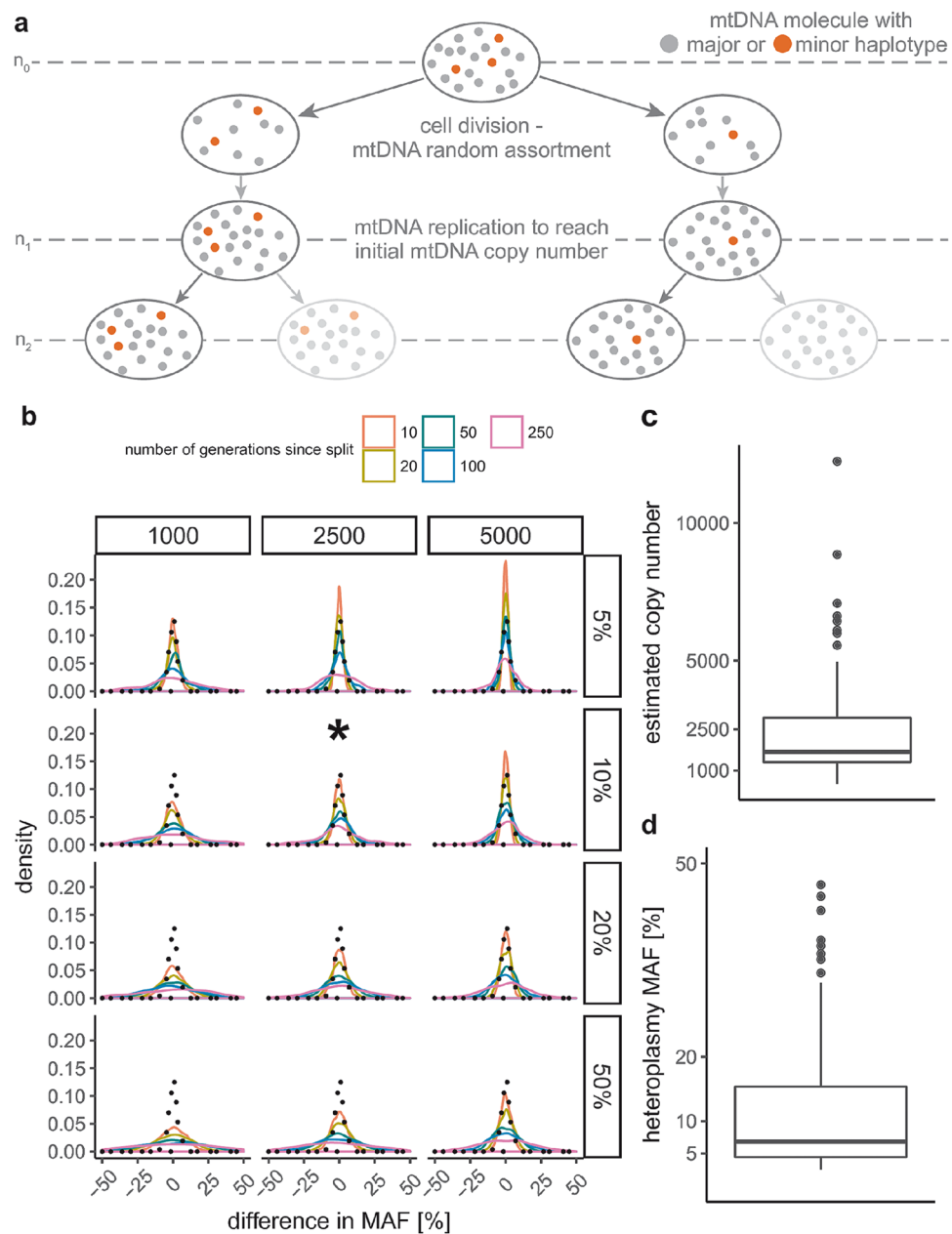
Figure 3. Expected difference in minor allele frequency (MAF) assuming random genetic drift. (**a**) Schematic scheme of the two-step mtDNA replication model used for the simulations: first, all mtDNA molecules are segregated equally into two daughter cells followed by replication step to re-gain the same number of mtDNA copies as in the progenitor cell. After each generation n_i, the MAF frequency of one random cell on each site of the pedigree is determined and the MAF difference between the two cells calculated. (**b**) Expected difference in MAF after 10, 20, 50, 100, and 250 generations since the split of the two cells. Different mtDNA copy numbers (1,000, 2,500, and 5,000; kept constant throughout simulation) and initial MAFs (5%, 10%, 20%, and 50%) were used to simulate random genetic drift between two cells. For each combination of mtDNA copy number and initial MAF, 1000 replicates were simulated. The black, dotted line indicates the observed MAF difference between the shared liver heteroplasmies in the data set. The asterisk highlights the simulation with the parameters closest to the ones observed in the data set. (**c**) The mtDNA copy number distribution estimated from the capture-enriched sequencing data of the liver samples. The values were corrected using a correction "ratio" of 1/150[25] to convert the relative to absolute mtDNA copy numbers. (**d**) The distribution of MAF in the liver heteroplasmies.

Discussion

In this study, we compared mtDNA heteroplasmy patterns in samples collected from the two primary liver lobes of 83 individuals sampled at autopsy and found a significant correlation in heteroplasmy MAF between the two liver samples from an individual. Moreover, if sites are heteroplasmic in both liver samples of an individual,
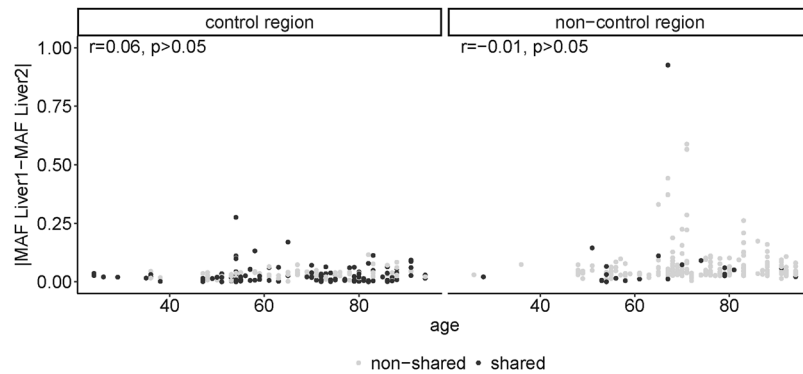
Figure 4. Minor allele frequency difference and age**.** The difference in minor allele frequency between liver sample 1 and 2 versus age plotted separately for sites within and outside the control region. Dark grey dots indicate sites that were shared (i.e. heteroplasmic in both liver samples from an individual), light grey dots indicate sites that were not shared.

heteroplasmies in liver1 and liver2	non-synonymous	synonymous	total
shared	6	11	17
not shared	157	51	208
total	163	62	

Table 2. Non-synonymous and synonymous heteroplasmic mutations in the protein-coding genes and shared and non-shared heteroplasmies. $p < 0.001$ (two-sided Fisher's exact test).

the MAFs at those sites tend to be similar independent of age. Previous studies found similar correlations in MAF between cell colonies derived from single cells grown in culture[16,17], so our results extend these findings to different parts of the same organ within living individuals. In addition, we analysed patterns of heteroplasmy sharing and MAF correlation across different parts of the mtDNA genome, which provide further information concerning the potential mechanism(s) behind these observations. In particular, we found significantly more sharing of heteroplasmic sites and higher MAF correlations for the control region than for the rest of the genome. We showed that these differences are not due to the presence of 7S DNA as it rather unlikely that we observed 7S DNA in our post-mortem sampled tissues due to the short halftime of 7S DNA of about 45 minutes in liver cells[29] and its single-stranded nature[30], which we are not able to capture using paired-end library preparation. These differences are neither driven by just a few sites. Moreover, within the protein-coding genes, we found significantly more sharing of heteroplasmies and higher MAF correlations at synonymous sites than at nonsynonymous sites.

These observations allow us to evaluate several potential explanations for this heteroplasmy sharing within liver. First, a trivial potential explanation is that a heteroplasmy could appear to be shared between liver lobes when it is actually present in blood. About 1 litre of blood per minute flows through the blood vessels of the liver[31], so the DNA from the liver samples also contains DNA arising from blood cells. Heteroplasmies present in the blood cells could therefore be detected as heteroplasmies in the liver samples and hence seem to be shared across lobes. There are ~10^{10} liver cells in 10 g of liver tissue and 5×10^{10} blood cells in 10 ml of blood[32], and in liver autopsy samples there is approximately 10 ml blood per 100 g of tissue[33]. The ratio of liver to blood cells in a liver autopsy sample is therefore about 1.8. As we additionally observed for our samples that the average mitochondrial copy number is five times higher in liver than in blood samples (Supplementary Fig. S6a), the overall ratio of mitochondrial genomes from liver cells to those from blood cells is approximately 9 to 1. Therefore, a heteroplasmy in blood would need a MAF of at least 22.5% to be called a heteroplasmy in liver (minimum MAF = 2.5%) when it was actually totally absent in the liver samples. Only 8 out of the 541 heteroplasmies in blood have a MAF of $\geq 22.5\%$, of which four are shared across blood and both liver lobes (Supplementary Table S1). Moreover, the heteroplasmies that are commonly observed in liver (at positions 60, 72, and 94) are practically absent from blood; these results rule out experimental contamination with blood DNA as the primary driver for heteroplasmy sharing across liver lobes.

A second reason for heteroplasmy sharing could be a high mutational pressure for some sites, with de novo mutations at the same site occurring independently throughout the tissue. If this was the case, one would expect to see a higher correlation of MAFs for common heteroplasmic sites, as those would be under high mutational pressure. We showed that this is not the case for the most common heteroplasmic sites 60, 72, and 94 in the data set (Fig. 2), which also are highly recurrent mutations throughout the mitochondrial phylogeny[34], and hence this explanation is unlikely. While we further observed a higher mutation rate for heteroplasmic sites than for non-heteroplasmic sites (Supplementary Fig. S7), there was no difference in mutation rate between shared and non-shared heteroplasmies, suggesting that the mutation rate has little impact on heteroplasmy sharing.

Third, it has been shown that colonic stem cells with non-synonymous mtDNA mutations can expand clonally from a few cells and spread throughout a tissue by crypt fission[35]. During this process they retain the same level

of mutant DNA on a cellular level. While a similar mechanism in liver could explain the results presented here, this process would require a cellular turnover on a huge scale throughout the entire liver, starting from just a few stem cells. This does not seem very likely, as clonal expansion was shown for single intestinal crypts only and is supposed to be rather slow[35]. Moreover, it is not clear to what extent liver regeneration is driven by stem cells vs. mature hepatocytes[36].

Fourth, heteroplasmy sharing could also derive from a pre-existing, inherited heteroplasmy[3,8,37–39] that remains at similar frequencies across the tissue because drift (random changes in MAF) is limited. We tested this possibility, using a simulation scheme from a previous study[16], with different initial heteroplasmy frequencies, and mtDNA copy numbers that were kept constant[40]. As expected, we observed a larger variance in MAF between daughter cells with increasing starting MAF and decreasing mtDNA copy number (Fig. 3b). For the simulation results closest to our observed average mtDNA copy number and average heteroplasmy MAF (2,500 and 10%, Fig. 3c,d), the observed difference in MAF between shared heteroplasmies was smaller than would be expected after random genetic drift acting on 50 mtDNA replication steps. However, more than these 50 mtDNA replication steps are expected to have taken place during the life of an average individual in our study. Assuming 3.61×10^{11} liver cells in an adult liver[41], a total of 39 hepatocyte cell replications (including mtDNA replication) are needed to obtain a full-size, adult liver from a single hepatocyte cell. After the development stage, the post-mitotic cells continue to replicate their mtDNA independently of cell replication ("relaxed" replication[42]). The estimated half-life of mtDNA in post-mitotic cells ranges between 2–10 days[43] and 30–300 days[44]. Assuming an mtDNA half-life of 30 days, there would be complete replacement of all mtDNA molecules of a cell within a year, given a mtDNA copy number of 2,500 per cell, so approximately one "relaxed" replication cycle occurs within the liver of an adult for every year of age. Thus, for the age range in our data set of 24 to 94 years, the liver samples would have gone through about 62–132 cell replications prior to sampling, and so the observed difference in MAF in the liver samples is significantly smaller than expected. Moreover, liver samples accumulated significantly more heteroplasmies with age (Supplementary Fig. S8A), further arguing that shared heteroplasmies do not reflect pre-existing, inherited heteroplasmies. Also, there was no significant correlation of MAF difference with age, as would be expected with random genetic drift. In sum, our results extend to liver tissues the previous observations[16,17] that random genetic drift alone cannot explain heteroplasmy sharing between cells in culture.

Finally, an equilibrium of heteroplasmy across an entire tissue could be explained by an exchange of genetic material from mitochondria between cells. Cells can donate whole mitochondria to adjacent cells through nanotubes, but this has been suggested for distances up to 100 μM only and the exchange is often triggered by functional impairments in the acceptor cell[45]. An additional way for cells to exchange DNA material could be the uptake of extracellular DNA material that is either secreted by healthy cells or is present as the remains of apoptotic cells[46]. While the uptake and integration of cell-free nuclear DNA material has been shown[47], it is unclear whether cells would also accept mitochondrial DNA. However, studies of heteroplasmy at the single cell level (reviewed in[48]) do suggest the possibility of transfer between cells. Experiments with cell culture mixes of fluorescently labelled cell lines suggested the exchange of mtDNA between co-cultured partner cell lines, although the specific mechanism, either transfer of mitochondrial organelles or transfer of free mtDNA, could not be identified[16]. Overall, such intercellular DNA exchange, followed by incorporation of mtDNA fragments into the mtDNA of the recipient cells, could account for the significant correlation we observe in MAF between liver lobes.

However, other aspects of our data are incompatible with the hypothesis of intercellular DNA exchange. In particular, intercellular exchange cannot explain the significantly higher number of shared heteroplasmies and correlation in MAF for the control region vs. the rest of the genome, unless one postulates that mtDNA fragments arising from the control region are either exchanged or incorporated between cells more frequently than mtDNA fragments arising from the rest of the genome. But even then, intercellular exchange cannot explain the significantly higher number of shared heteroplasmies and correlation in MAF for synonymous vs. non-synonymous heteroplasmies, as both should be exchanged at the same rate between cells.

Instead, our data suggest that even if intercellular exchange is occurring, selection must be involved in the sharing of heteroplasmies and correlation in MAF between liver lobes. Several aspects of the data suggest that selection influences heteroplasmies. In the protein-coding genes of the mtDNA, we observed a significant excess of non-synonymous vs. synonymous heteroplasmies (Table 2), more so than can be explained by relaxation of functional constraints on non-synonymous mutations. Moreover, the number of non-synonymous heteroplasmies increased significantly in individuals above 60 years. Overall, these results strongly suggest positive selection for non-synonymous heteroplasmies in liver, as found previously[8], and possibly reflecting the hypothesis of the "survival of the slowest"[49], which postulates that mitochondria with reduced respiratory function due to increasing mutations suffer less degradation from the production of reactive oxygen species (ROS). Hence, mitochondria that lack these mutations suffer ROS-related damage and are removed from cells, thereby resulting in an increase in frequency in mtDNAs with non-synonymous mutations that decrease mitochondrial function.

However, our results indicate a more complex role for selection in the different patterns of heteroplasmy sharing and MAF correlation across different regions of the mtDNA genome, in keeping with evidence from other studies. In mice and humans, significantly more synonymous than non-synonymous heteroplasmies are transmitted to the next generation[3,21,50], suggesting selection against non-synonymous heteroplasmies during transmission. The notable tissue-specificity and allele-specificity of particular heteroplasmic sites in the control region also suggests a role for positive selection on heteroplasmies during aging[8,11]. The increasing evidence for both purifying and positive selection acting on heteroplasmic variants warrants further investigation, particularly into the potential health-related consequences.

Material and Methods

Tissue collection and DNA extraction.   Blood and liver were sampled at autopsy from 94 individuals (57 males, 37 females, age range: 24–94, mean: 63, median: 63). Two samples were taken from each liver, one from the right lobe and one from the left lobe. However, the labelling of the liver samples (liver1 and liver2) does not correspond to either the right or the left lobe, respectively. DNA was extracted as previously described[8]. The collection of samples, the experimental procedures and the informed consent procedure were approved by the Ethics Commissions of the Rheinische Friedrich Wilhelms University Medical Faculty (Lfd. Nr. 097/15) and the University of Leipzig Medical Faculty (Az. 305-15-24082015). All experiment were performed following the approved procedures and in accordance with the relevant guidelines and regulations.

Virological assays and histological investigation.   Human immunodeficiency virus (HIV) RNA and Hepatitis C virus (HCV) RNA concentration in blood was determined by using the Abbott RealTi*me*® HIV-1 and HCV systems and the m2000sp/m2000rt instruments according to the instructions of the manufacturer. For detection of Hepatitis B virus (HBV) DNA the Abbott RealTi*me*® HBV system was used. The 95% limit of detection (LOD_{95}) of the HIV, HCV and HBV assay was 40 copies/mL, 12 IU/mL, and 10 IU/mL, respectively. If inhibitory effects on enzymatic reactions were present (detected via co-amplification of control RNA or DNA sequences), blood samples were re-tested at dilutions of 1/5, 1/10, and 1/15 (13 samples (13%) for HIV, 91 (93%) for HCV, and 8 (7%) for HBV load). For dilution, a plasma donation from a blood donor negative for HIV, HCV, and HBV was used. Of the diluted samples tested for HIV, one sample was diluted 1/5, nine had to be diluted 1/10, and two had to be diluted 1/15, lowering the LOD_{95} of the assay to 200, 400, 600, respectively. In one sample, inhibition could not be eliminated by sample dilution. Of the diluted samples tested for HCV, one, 78, and two samples were diluted 1/5, 1/10, and 1/15, respectively, lowering the LOD_{95} of the assay to 60, 120, 180 IU/ml, respectively. In 10 additional samples no result could be achieved, even after dilution. Of the diluted samples tested for HBV load, 3, 2, and 2 samples were diluted 1/5, 1/10, and 1/15, respectively, lowering the LOD_{95} of the assay to 50, 100, and 150 IU/mL, respectively. All other samples were tested without any dilution.

Fat content of the liver was determined by histological investigations and Sudan staining[51]. Tissues with <10% hepatocytes including fat droplets were considered low, 10–30% were medium, 31–50% were high fat and >50% were considered adipohepatic.

Illumina library preparation and sequencing.   Double-barcoded DNA libraries for sequencing were prepared and capture-enriched for mtDNA as previously described[8]. DNA was sequenced on the Illumina HiSeq platform in rapid mode with 95 bp paired-end reads. Bases were called with FreeIbis[52] and reads were subsequently trimmed and merged using leeHom[53].

Heteroplasmy detection.   Heteroplasmy was detected according to the DREEP pipeline[22,23]; heteroplasmies called by this pipeline were validated independently in a previous study[8]. First, heteroplasmies were called if: the minor allele frequency (MAF) for the most frequent minor allele was at least 2.5% on both the forward and reverse strand; the sequencing depth was at least 500-fold at a candidate site; and there were at least 10 reads supporting the minor allele on each strand. Additionally, we required a minimum heteroplasmic quality score of 10 on each strand. In order to discriminate a true heteroplasmy from sequencing error, the DREEP pipeline compares the minor allele pattern of any inferred heteroplasmic site to a database that comprises the minor allele patterns at this site from all other individuals in the study. When the majority of the samples are from a single tissue like liver in this study (two liver samples and one blood sample per individual), DREEP is prone to consider commonly heteroplasmic sites as sequencing error and therefore under-estimates the heteroplasmic quality score for these. Thus, we used the information about commonly heteroplasmic sites from[8] to flag these for both blood and liver tissue in this study, respectively. A site was considered commonly heteroplasmic in a tissue when at least five individuals were heteroplasmic for it. Based on this criterion, we ignored the heteroplasmic quality score for sites 60, 72, 94, 185, 189, 203, 11,126, 16,093, and 16,126 for liver and 12,705 for blood and considered a site heteroplasmic at any of these sites to be heteroplasmic in an individual if all other criteria were fulfilled. The following regions were excluded for heteroplasmy analysis because of stretches of repetitive DNA that are prone to incorrect sequence-alignment: 302–316, 513–526, 566–573, and 16,181–16,194. We confirmed that all heteroplasmies were within a coverage between 20% and 200% of the average coverage of the sample. In addition, all samples that could have been contaminated with other samples during library preparation/extraction were removed. To detect such contamination, pairwise comparisons of all liver samples with each other as well as all blood samples with each other were performed. All three samples of an individual were removed if all of the following criteria were fulfilled for any pairwise comparison between individuals across all sites of the mitochondrial genome: 1) for at least 80% of the sites, for which two samples had different consensus alleles, the minor allele in the recipient sample was identical to the major allele in the donor sample; 2) the average MAF across these sites was at least 1%; and 3) at least 60% of all sites in the recipient sample, for which a minor allele was observed, were identified as heteroplasmies by the DREEP pipeline[22,23]. Furthermore, an additional filter for potential contamination was applied, in which the heteroplasmic sites for each sample were checked to see if five or more sites could be explained by contamination from another haplogroup. In total, ten individuals were removed in these contamination filter steps. Finally, we removed a single individual because for all three tissue samples more than 2% of the mtDNA genome had a sequencing depth below 500-fold, the cut-off for being considered a heteroplasmy, and thus would have had a higher false negative rate than the other samples. Overall, we retained 83 individuals for the subsequent analyses. Minor allele frequencies were calculated with respect to the major allele in blood.

Correlation analysis.   Statistical analysis was performed using R (https://www.R-project.org), with analyses performed for the entire mtDNA genome and separately for the non-control region (577–16,023), the control

region (16,024–576) and the D-loop region (16,097–191). For correlation of MAFs, we selected only sites that were identified as heteroplasmies and passed our quality filters in at least one of the tissues. We then compared the MAFs of these heteroplasmies to the MAF in the other tissues of an individual, even if the site was not called as a heteroplasmy in the other tissues. Pearson correlation coefficients were calculated for correlations between MAFs among samples as well as for correlations with age assuming a two-sided alternative hypothesis; the significance of the correlation was tested by randomly permuting the data. Permutation tests were also used to assess the association between specific sites or regions and minor allele sharing between liver lobes; all permutations were carried out 1000 times. Whenever categorical data were compared (e.g. synonymous vs. non-synonymous sites), two-sided Mann-Whitney U tests were used to test for significant differences. Fisher's exact test was used to test for an association of sharing of heteroplasmic sites between liver lobes for the control region vs. non-control region and for synonymous vs. non-synonymous heteroplasmies using a two-sided alternative hypothesis. P-values were adjusted for multiple testing using Benjamini-Hochberg correction[54].

Coverage across the mitochondrial genome.   We used per-site coverage determined by the *filter_and_summary.pl* script of the DREEP pipeline[23] for each sample and calculated the average coverage across the D-loop region and across the rest of the mtDNA genome.

Non-synonymous heteroplasmies.   The hN/hS ratio[8] was obtained by calculating the Ka/Ks ratio using *KaKs_Calculator 2.0*[55] between the revised Cambridge Reference Sequence (rCRS; Andrews, 1999 https://doi.org/10.1038/13779) and a mock sequence created by introducing all minor alleles of heteroplasmies into the coding region of the rCRS. A significance test was performed by randomly introducing the same substitutions as observed for heteroplasmies at any site of the coding region of rCRS and calculating the hN/hS ratio in comparison to the non-altered rCRS.

The potential functional impact of non-synonymous heteroplasmies was analysed by overlapping the position of the heteroplasmy and its minor allele with the MitImpact database[56] and comparing the *MutationAssessor*[57] score.

Data access.   The raw sequencing data have been deposited with the European Nucleotide Archive under accession number PRJEB27731. The scripts for heteroplasmy detection and filtering can be found at http://dmcrop.sourceforge.net/ and https://github.com/alexhbnr/liverlobes _heteroplasmy.

References

1. Larsson, N. G. Somatic mitochondrial DNA mutations in mammalian aging. *Annual review of biochemistry* **79**, 683–706, https://doi.org/10.1146/annurev-biochem-060408-093701 (2010).
2. Bogenhagen, D. & Clayton, D. A. Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. *Cell* **11**, 719–727 (1977).
3. Rebolledo-Jaramillo, B. *et al.* Maternal age effect and severe germ-line bottleneck in the inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15474–15479, https://doi.org/10.1073/pnas.1409328111 (2014).
4. Stewart, J. B. & Chinnery, P. F. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet* **16**, 530–542, https://doi.org/10.1038/nrg3966 (2015).
5. Wallace, D. C., Chalkia, D. & Mitochondrial, D. N. A. Genetics and the Heteroplasmy Conundrum in Evolution and Disease. *Cold Spring Harbor Perspectives in Biology* **5**, a021220, https://doi.org/10.1101/cshperspect.a021220 (2013).
6. Ye, K., Lu, J., Ma, F., Keinan, A. & Gu, Z. Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proceedings of the National Academy of Sciences* **111**, 10654–10659 (2014).
7. Calloway, C. D., Reynolds, R. L., Herrin, G. L. & Anderson, W. W. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *American Journal of Human Genetics* **66**, 1384–1397, https://doi.org/10.1086/302844 (2000).
8. Li, M., Schroeder, R., Ni, S., Madea, B. & Stoneking, M. Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 2491–2496, https://doi.org/10.1073/pnas.1419651112 (2015).
9. Michikawa, Y., Mazzucchelli, F., Bresolin, N., Scarlato, G. & Attardi, G. Aging-dependent large accumulation of point mutations in the human mtDNA control region for replication. *Science* **286**, 774–779, https://doi.org/10.1126/science.286.5440.774 (1999).
10. Naue, J. *et al.* Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion* **20**, 82–94, https://doi.org/10.1016/j.mito.2014.12.002 (2015).
11. Samuels, D. C. *et al.* Recurrent Tissue-Specific mtDNA Mutations Are Common in Humans. *Plos Genetics* **9**, e1003929, https://doi.org/10.1371/journal.pgen.1003929 (2013).
12. Wang, Y. *et al.* Muscle-specific mutations accumulate with aging in critical human mtDNA control sites for replication. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4022–4027, https://doi.org/10.1073/pnas.061013598 (2001).
13. Jenuth, J. P., Peterson, A. C., Fu, K. & Shoubridge, E. A. Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. *Nature genetics* **14**, 146–146 (1996).
14. Jenuth, J. P., Peterson, A. C. & Shoubridge, E. A. Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice. *Nature genetics* **16**, 93–93 (1997).
15. Nachman, M. W., Brown, W. M., Stoneking, M. & Aquadro, C. F. Nonneutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* **142**, 953–963 (1996).
16. Jayaprakash, A. D. *et al.* Stable heteroplasmy at the single-cell level is facilitated by intercellular exchange of mtDNA. *Nucleic Acids Research* **43**, 2177–2187, https://doi.org/10.1093/nar/gkv052 (2015).
17. Raap, A. K. *et al.* Non-random mtDNA segregation patterns indicate a metastable heteroplasmic segregation unit in m. 3243A> G cybrid cells. *PLoS One* **7**, e52080–e52080 (2012).
18. Wilton, P. R., Zaidi, A., Makova, K. & Nielsen, R. A population phylogenetic view of mitochondrial heteroplasmy. *Genetics* **208**, 1261–1274 (2018).
19. Avital, G. *et al.* Mitochondrial DNA heteroplasmy in diabetes and normal adults: role of acquired and inherited mutational patterns in twins. *Human Molecular Genetics* **21**, 4214–4224, https://doi.org/10.1093/hmg/dds245 (2012).
20. Kowald, A. & Kirkwood, T. B. L. Evolution of the mitochondrial fusion–fission cycle and its role in aging. *Proceedings of the National Academy of Sciences* **108**, 10237–10242 (2011).
21. Stewart, J. B. *et al.* Strong purifying selection in transmission of mammalian mitochondrial DNA. *PLoS biology* **6**, e10–e10 (2008).

22. Li, M. *et al*. Detecting Heteroplasmy from High-Throughput Sequencing of Complete Human Mitochondrial DNA Genomes. *American Journal of Human Genetics* **87**, 237–249, https://doi.org/10.1016/j.ajhg.2010.07.014 (2010).
23. Li, M. & Stoneking, M. A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biology* **13**, R34, https://doi.org/10.1186/gb-2012-13-5-r34 (2012).
24. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *Plos One* **5**, e14004, https://doi.org/10.1371/journal.pone.0014004 (2010).
25. Wachsmuth, M., Hübner, A., Li, M., Madea, B. & Stoneking, M. Age-Related and Heteroplasmy-Related Variation in Human mtDNA Copy Number. *Plos Genetics* **12**, e1005939, https://doi.org/10.1371/journal.pgen.1005939 (2016).
26. Nicholls, T. J. & Minczuk, M. In D-loop: 40 years of mitochondrial 7S DNA. *Experimental Gerontology* **56**, 175–181, https://doi.org/10.1016/j.exger.2014.03.027 (2014).
27. Roberti, M. *et al*. Multiple protein-binding sites in the TAS-region of human and rat mitochondrial DNA. *Biochemical and Biophysical Research Communications* **243**, 36–40, https://doi.org/10.1006/bbrc.1997.8052 (1998).
28. Soares, P. *et al*. Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *American Journal of Human Genetics* **84**, 740–759, https://doi.org/10.1016/j.ajhg.2009.05.001 (2009).
29. Gensler, S. *et al*. Mechanism of mammalian mitochondrial DNA replication: import of mitochondrial transcription factor A into isolated mitochondria stimulates 7S DNA synthesis. *Nucleic Acids Research* **29**, 3657–3663, https://doi.org/10.1093/nar/29.17.3657 (2001).
30. Antes, A. *et al*. Differential regulation of full-length genome and a single-stranded 7S DNA along the cell cycle in human mitochondria. *Nucleic Acids Research* **38**, 6466–6476, https://doi.org/10.1093/nar/gkq493 (2010).
31. Wynne, H. A. *et al*. The effect of age upon liver volume and apparent liver blood flow in healthy man. *Hepatology* **9**, 297–301 (1989).
32. Sender, R., Fuchs, S. & Milo, R. Revised estimates for the number of human and bacteria cells in the body. *PLoS biology* **14**, e1002533 (2016).
33. Greenway, C. & Stark, R. Hepatic vascular bed. *Physiological reviews* **51**, 23–65 (1971).
34. Levin, L., Zhidkov, I., Gurman, Y., Hawlena, H. & Mishmar, D. Functional recurrent mutations in the human mitochondrial phylogeny: dual roles in evolution and disease. *Genome biology and evolution* **5**, 876–890 (2013).
35. Greaves, L. C. *et al*. Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 714–719, https://doi.org/10.1073/pnas.0505903103 (2006).
36. Grompe, M. Liver stem cells, where art thou? *Cell stem cell* **15**, 257–258 (2014).
37. Burgstaller, J. P. *et al*. Large-scale genetic analysis reveals mammalian mtDNA heteroplasmy dynamics and variance increase through lifetimes and generations. *Nature communications* **9**, 2488–2488 (2018).
38. Guo, Y. *et al*. Very Low-Level Heteroplasmy mtDNA Variations Are Inherited in Humans. *Journal of Genetics and Genomics* **40**, 607–615, https://doi.org/10.1016/j.jgg.2013.10.003 (2013).
39. Payne, B. A. I. *et al*. Universal heteroplasmy of human mitochondrial DNA. *Human Molecular Genetics* **22**, 384–390, https://doi.org/10.1093/hmg/dds435 (2013).
40. Berk, A. J. & Clayton, D. A. Mechanism of mitochondrial DNA replication in mouse L-cells: asynchronous replication of strands, segregation of circular daughter molecules, aspects of topology and turnover of an initiation sequence. *Journal of molecular biology* **86**, 801–824 (1974).
41. Bianconi, E. *et al*. An estimation of the number of cells in the human body. *Annals of human biology* **40**, 463–471 (2013).
42. Poovathingal, S. K., Gruber, J., Halliwell, B. & Gunawan, R. Stochastic drift in mitochondrial DNA point mutations: a novel perspective ex silico. *PLoS computational biology* **5**, e1000572 (2009).
43. Miwa, S., Lawless, C. & Von Zglinicki, T. 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).
44. Poovathingal, S. K., Gruber, J., Lakshmanan, L., Halliwell, B. & Gunawan, R. Is mitochondrial DNA turnover slower than commonly assumed? *Biogerontology* **13**, 557–564 (2012).
45. Rogers, R. S. & Bhattacharya, J. When Cells Become Organelle Donors. *Physiology* **28**, 414–422, https://doi.org/10.1152/physiol.00032.2013 (2013).
46. van der Vaart, M. & Pretorius, P. J. In *Cir*culating *Nu*cleic *Acids in Plasma and Serum V* Vol. 1137 *Annals of the New York Academy of Sciences* (eds Gahan, P. B. & Swaminathan, R.) 18–26 (2008).
47. Basak, R., Nair, N. K. & Mittra, I. Evidence for cell-free nucleic acids as continuously arising endogenous DNA mutagens. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **793**, 15–21, https://doi.org/10.1016/j.mrfmmm.2016.10.002 (2016).
48. Yao, Y. G., Kajigaya, S. & Young, N. S. Mitochondrial DNA mutations in single human blood cells. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* **779**, 68–77, https://doi.org/10.1016/j.mrfmmm.2015.06.009 (2015).
49. deGrey, A. A proposed refinement of the mitochondrial free radical theory of aging. *Bioessays* **19**, 161–166, https://doi.org/10.1002/bies.950190211 (1997).
50. Floros, V. I. *et al*. Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos. *Nature cell biology* **20**, 144–144 (2018).
51. Mulisch, M. & Welsch, U. *Romeis - Mikroskopische Technik*. Vol. 19th Edition (Springer Spektrum, 2015).
52. Renaud, G., Kircher, M., Stenzel, U. & Kelso, J. freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics* **29**, 1208–1209, https://doi.org/10.1093/bioinformatics/btt117 (2013).
53. Renaud, G., Stenzel, U. & Kelso, J. leeHom: adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Research* **42**, e141, https://doi.org/10.1093/nar/gku699 (2014).
54. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289–300 (1995).
55. Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, proteomics & bioinformatics* **8**, 77–80, https://doi.org/10.1016/s1672-0229(10)60008-3 (2010).
56. Castellana, S., Ronai, J. & Mazza, T. MitImpact: an exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants. *Hum Mutat* **36**, E2413–2422, https://doi.org/10.1002/humu.22720 (2015).
57. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* **39**, e118, https://doi.org/10.1093/nar/gkr407 (2011).

Acknowledgements

Author Contributions

M.S., B.M., M.L., A.H. and M.W. designed the study. B.M. collected the samples. R.S. performed DNA isolation and sequencing library preparation, A.M.E.H. performed virological assays and B.M. performed histological analyses. M.L. and A.H. wrote pipeline and executed heteroplasmy detection, A.H. and M.W. performed statistical analyses. A.H., M.W., A.M.E.H. and M.S. wrote the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-47570-1.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.