Mitochondrial DNA heteroplasmy is shared between human liver lobes

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

[1]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany

[2]Fondation Mérieux, Lyon 69002, France

[3]Institute of Virology, University of Bonn Medical Centre, Sigmund-Freud-Str. 25, 53105 Bonn. Germany

[4]Institut für Rechtsmedizin, Universitätsklinikum Bonn, Stiftsplatz 12, D-53111 Bonn, Germany

* Corresponding authors

E-mail: manja_wachsmuth@eva.mpg.de, stonekg@eva.mpg.de

¶ These authors contributed equally to the manuscript.

Abstract

Background: Mitochondrial DNA (mtDNA) heteroplasmy is the presence of mtDNA molecules with different sequences in the same individual. Previous studies have shown that different tissues have different heteroplasmic patterns in an individual. However, to date patterns of heteroplasmy within a single tissue have not been investigated. We therefore investigated heteroplasmy in blood (Bl) and two liver (Liv) samples (one from each lobe) from 85 humans, sampled at autopsy.

Results: Minor allele frequencies (MAF) at heteroplasmic sites were significantly correlated between liver samples from an individual, with more sharing of heteroplasmic sites in the control region than in the coding region. Neither specific sites nor the presence of 7S DNA can explain this pattern. While age was highly correlated with an increase in the total number of heteroplasmic sites, the correlation of MAFs between liver samples of an individual was independent of age. Although there was a significant excess of non-synonymous heteroplasmies in the coding region, synonymous heteroplasmies were more likely to be shared.

Conclusions: While several mechanisms could potentially explain these results, the most likely is transfer of DNA between cells, with fragments originating from the mtDNA control region more likely to serve as primers for mtDNA replication than fragments from the coding region. The result is an integration of the same mutant alleles into the genomes of new cells. However, the significant association of synonymous mutations with heteroplasmy sharing between tissue regions indicates that the mechanisms of intercellular DNA exchange is more complex than previously suspected, and may reflect different processes.

Keywords

Liver lobes, heteroplasmy distribution, mtDNA, control region, mutation, selection, 7S DNA

Background

The human mitochondrial DNA (mtDNA) genome is ~16.6 kilo base pairs long and consists of a non-coding control region and a coding region, in which 13 protein subunits of the electron transport chain are encoded [1, 2]. Transcription of all genes is initiated at 3 promoters in the control region. This control region often contains a third DNA molecule (7S DNA), which is about 650 nt long and embedded within the so-called Displacement- or D-loop region, a subsection of the mtDNA in the 1.1 kb long control region [3]. Due to elevated exposure to reactive oxygen species in combination with a lack of protecting histones, the mtDNA mutation rate is two orders of magnitude higher than that of the nuclear genome [4].

The mitochondrial genome is present in many copies in a single cell, and inter-individual variation in the mitochondrial genome of an individual is called mtDNA heteroplasmy [2]. In humans, it has been shown that detrimental mtDNA mutations are usually present in a heteroplasmic fashion at low frequencies, with high frequencies of the deleterious allele leading to functional defects and a disease phenotype [2, 5, 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 [5]. The total number of heteroplasmic sites strongly correlates with age and several studies have shown that heteroplasmic sites are tissue specific [7-13]. This means that sites which are often heteroplasmic in one tissue can be homoplasmic in all other tissues of the

same individual. While most age-related heteroplasmies occur in the control region, liver tissue is unusual in showing an excess of heteroplasmies involving non-synonymous mutations of the encoded electron transport chain proteins, more than would be expected with relaxation of purifying selection [9]. Moreover, these mutations are predicted to have functional impacts, suggesting that there is age-related positive selection in liver for somatic mutations that decrease mitochondrial function [9].

To date, the distribution of heteroplasmy throughout an individual's tissue has not been investigated and it is not clear how heteroplasmy differs among different regions of the same tissue. To investigate this issue, we obtained one blood sample and two liver samples (one from each lobe) from 85 European individuals, collected at autopsy. MtDNA heteroplasmy was evaluated by capture-enrichment sequencing, and we analyzed sharing of mtDNA heteroplasmy between the liver lobes for different regions of the mitochondrial genome. We find a surprising correlation in the MAF at heteroplasmic sites in the control region between the two liver samples, but a much weaker correlation in the coding region.

Results

Heteroplasmy sharing within the liver

We investigated mitochondrial DNA heteroplasmy in liver and blood tissue samples of 85 individuals. For liver, two samples taken from different lobes were analyzed in order to compare the heteroplasmic pattern in different parts of the tissue. Applying a threshold of 2 % MAF, we detected 780 heteroplasmic sites at 485 different positions (Table S1). More heteroplasmic sites were observed in the coding region for liver (307 sites) compared to blood (118 sites), but the most abundant heteroplasmic sites

in liver were in the control region (site 72: 62 individuals, site 60: 29 individuals, site 94: 27 individuals), which were only rarely observed in the blood (site 72: 1 individual, site 60: 1 individual, site 94: no individuals). These data are in accordance with results from a previous study ([9], Figure S1), indicating that heteroplasmy is tissue specific, with different individuals exhibiting similar heteroplasmic patterns.

Virological tests revealed that two individuals had active hepatitis B virus infection, one had active hepatitis C virus infection and one individual was HIV positive, with low viral load (Table S2). Those individuals were kept in all downstream analyses, as the number of positive cases was too low to analyze separately. There was no effect of liver fat content on either the total number or the MAFs of heteroplasmic sites (Figure S2, p>0.05).

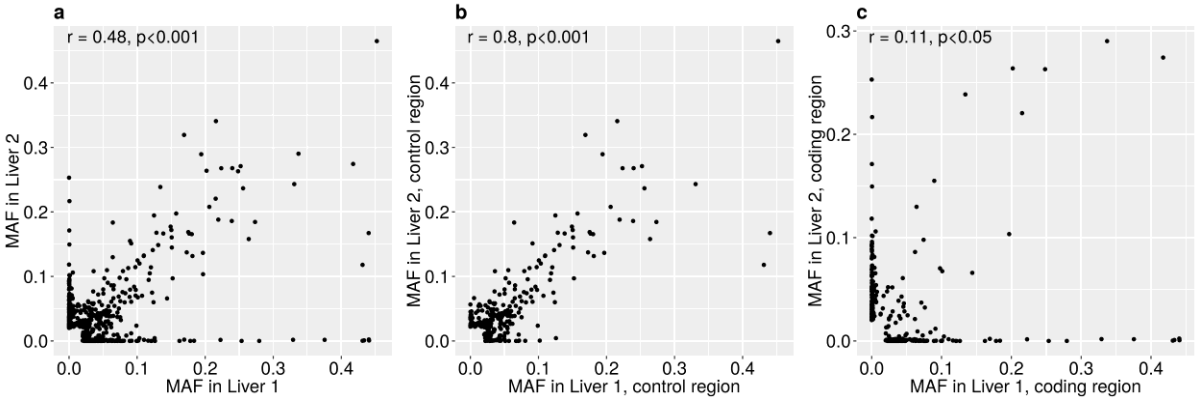

Figure 1: Correlation of MAFs at heteroplasmic sites in liver lobes. Each dot is one heteroplasmic site in one individual. Pearson's correlation coefficient r is given. Heteroplasmic sites are compared in **a** liver sample 1 and 2, **b** the control region of liver sample 1 and 2, **c** the non-control region of liver sample 1 and 2.

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=0.18 and 0.16, p<0.001, Figure S3). However, the correlation between the two liver samples was higher (Figure 1a, r=0.48, p<0.001). While 481 sites were heteroplasmic in only one of the two liver samples of an

individual, 166 sites were heteroplasmic in both liver samples and these exhibited similar MAFs (Figure 1a).

Moreover, there were more shared heteroplasmies from the control region than from the rest of the genome (Table 1, $p<10^{-15}$), and MAFs in the two liver samples were more highly correlated for the control region (Figure 1b, r=0.8, p<0.001) than for the coding region (Figure 1c, r=0.11, p<0.05, difference significant with p<0.001 (Figure S4)).

Table 1: Heteroplasmic mutations in the control region and the non-control region and shared and non-shared heteroplasmies. $p<10^{-15}$ (Fisher's exact test)

heteroplasmies in Liv1 and Liv2	control region	non-control region	total
shared	136	30	166
not shared	126	355	481
total	262	385	

There are three sites that are frequently heteroplasmic in liver (Table S2) 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 analyzed them separately. While these three sites did indeed show a high correlation in MAF between the two liver samples (r = 0.84, p<0.001, Figure 2a), the remaining sites in the control region still showed a significant correlation (Figure 2b, r=0.74, p<0.001) that is higher than the correlation for the coding region (Figure 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 118 occurrences of heteroplasmy at these three sites, we sampled 118

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 (Figure 2c), indicating that the higher correlation between MAF for the control region than for the rest of the genome is not driven solely by these three sites.

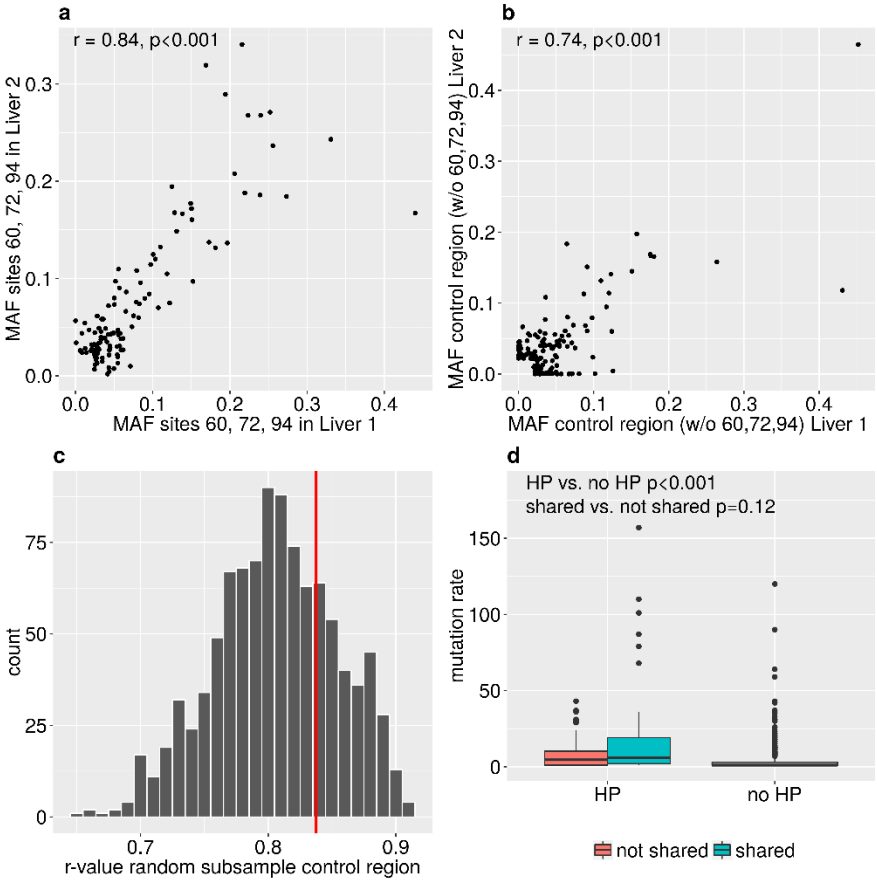

Figure 2: Heteroplasmy sharing in liver samples and correlation with single sites. MAFs per site in liver sample 1 and 2 for: **a** sites 60, 72 and 94; and **b** for all other heteroplasmic sites in the control region. Each dot is one heteroplasmic site in one individual. Pearson's correlation coefficient 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 red bar. **d** Average estimated mutation rate [14] for heteroplasmic sites (HP) that are shared vs. not shared vs. sites that were not heteroplasmic. Mann-Whitney U test was done for shared vs. non-shared heteroplasmies and for heteroplasmies vs. sites that were not heteroplasmic. Only polymorphic sites (i.e. those with reported mutation rates [14]) were included.

We also tested if heteroplasmic sites showed a higher mutation rate than average by comparing the number of heteroplasmies to the inferred mutation rate at each site, based on observed polymorphism data [14]. Heteroplasmic sites were significantly more likely to also exhibit polymorphism between individuals than a randomly chosen subsample of sites of the same size ($p<0.001$, Figure S5). Furthermore, heteroplasmies had significantly higher mutation rates than sites that were not heteroplasmic ($p<0.001$, Figure 2d). However, the mutation rate for sites that were shared did not differ from the mutation rate for sites that were not shared ($p=0.12$). Hence, while heteroplasmic sites are associated with a higher mutation rate than non-heteroplasmic sites, this does not explain the increase in heteroplasmy sharing in the control region, as shared sites have similar mutation rates compared to non-shared sites.

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 [15]. 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 [3, 16]. The relative mtDNA copy number estimates were significantly correlated between the two liver samples of an individual, indicating that the relative estimates from capture-enriched data are reliable (Figure S6a). 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=0.34 and 0.47 for Bl and Liv, respectively, Figure S6b). Nevertheless, the correlation between the MAF for the two liver samples is higher for the D-loop region than for the rest of the control region (r=0.82 vs. r = 0.79, Figure 3a, 3b).

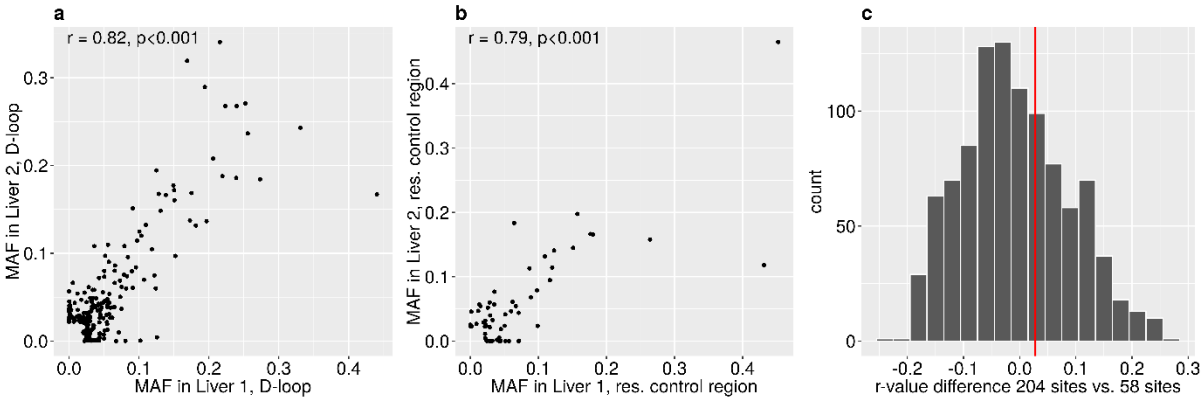

Figure 3: Heteroplasmy sharing in liver samples and correlation with 7S DNA. The MAFs at heteroplasmic sites are compared for: **a** sites in the D-loop region; and **b** sites in the control region but outside the D-loop region. Each dot is one heteroplasmic site in one individual. Pearson's correlation coefficient r is given. **c** Distribution of r-value differences for random partitions of the control region heteroplasmies into 204 sites (matching the number of sites in the D-loop region) and 58 sites (matching the number of sites in the control region outside the D-loop region). The red bar indicates the observed difference between the r-values for the D-loop and the rest of the control region.

In order to test if sites from the D-loop were more likely to have similar MAFs than other sites in the control region, we performed a subsampling test. In the entire control region, 262 sites were heteroplasmic in at least one liver sample of an individual. Out of those, 204 are in the D-loop and 58 are in the rest of the control region. We randomly selected 204 sites from the control region, calculated the difference between r-values for this subset and the remaining 58 sites, and repeated this 1000 times to generate a random distribution of r-value differences. The observed r-value difference between the D-loop and the rest of the control region was not significantly different from the random distribution of r-value differences (Figure 3c). Overall, the significant correlation in MAFs in the control region does not

reflect 7S DNA, nor is it restricted to the D loop region. Hence, the effect likely is a phenomenon of the entire control region.

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. The total number of different heteroplasmic sites in an individual increased with age for both the control region and the coding region (Figure 4a-c, $r=0.47$ for the coding region, 0.36 for the control region, $p<0.001$). The MAF at heteroplasmic sites did not increase with age (Figure 4d-f, $r\approx0$, $p>0.05$). Although some heteroplasmies exhibit high MAFs only at ages above 50, many sites remain at low frequencies even at older ages (Figure 4d-f). We also separately tested the three high frequency heteroplasmy sites (60, 72, and 94) for correlations between MAF and age (Figure S7), and found a significant correlation of MAF with age for site 94 ($r=0.36$, $p<0.05$), but not for sites 60 and 72 ($r=0.14$ for both sites, $p>0.05$). Finally, there was a very weak positive correlation with the difference in MAF between the two liver samples and age for all sites combined (Figure 4g, $r=0.08$, $p<0.05$), but this was not present when looking at the control region or the coding region separately (Figure 4h-I, $r=0.04$ or 0.05, respectively, $p>0.05$). This indicates that, if sites are shared between different liver lobes, MAFs change in concert irrespective of age.

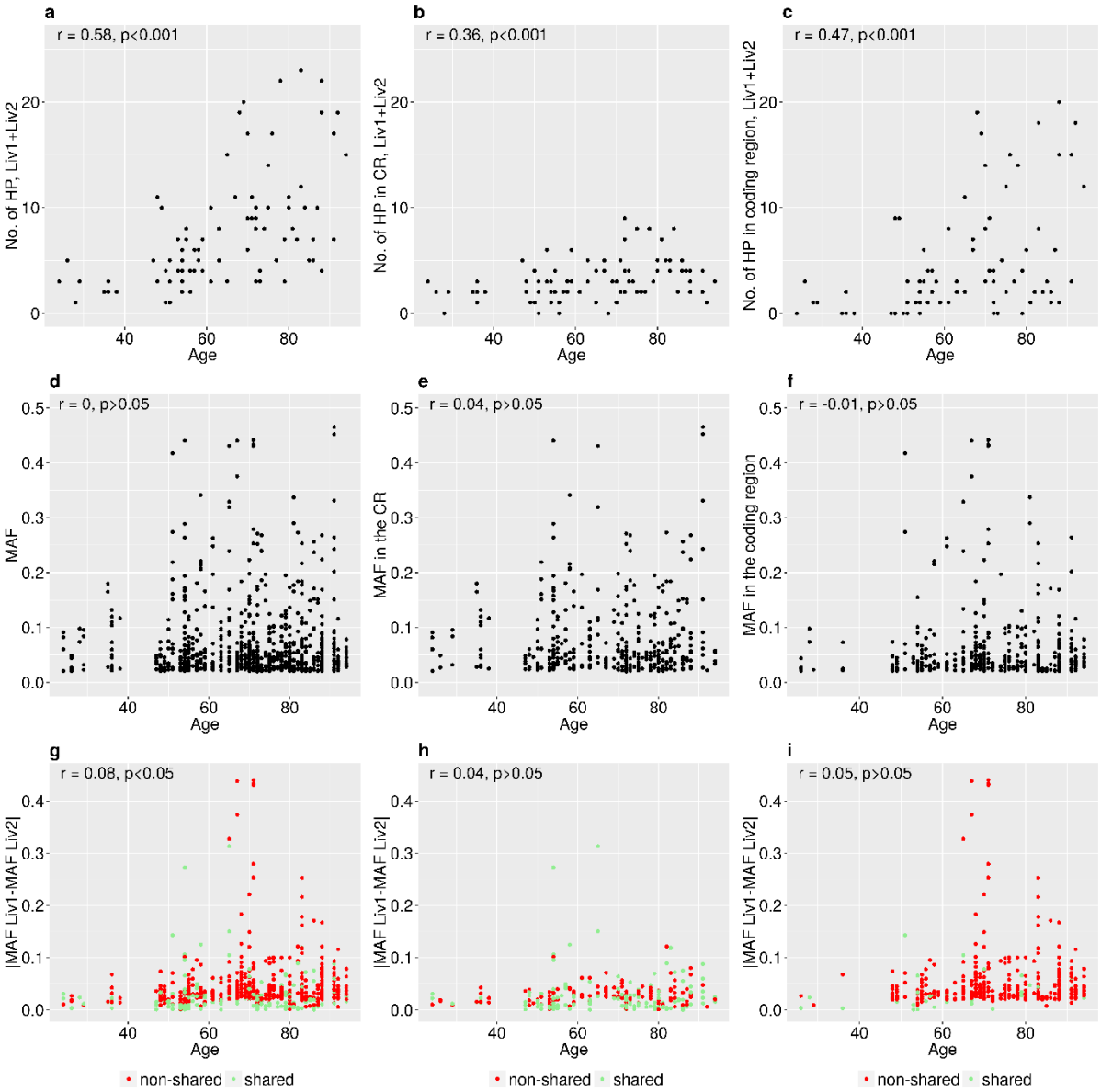

Figure 4: Heteroplasmy and age. Total number of heteroplasmic sites per individual (sum of liver sample 1 and 2) versus age for: **a** the whole mtDNA molecule; **b** the control region; and **c** the coding region. MAF versus age for: **d** the whole mtDNA molecule; **e** the control region and **f** the coding region. Each dot is one heteroplasmic site in one sample. MAF difference between liver sample 1 and 2 versus age for: **g** the whole mtDNA molecule; **h** the control region; and **i** the coding region. Green dots indicate sites that were shared (i.e., heteroplasmic in both liver samples from an individual), red dots indicate sites that were not shared.

Heteroplasmy sharing in the coding region

Previous studies showed that liver has a significant excess of non-synonymous heteroplasmies which are predicted to have an impact on function [9]. This is also

the case for our data set (Figure S8). As hN/hS-ratios (analogous to dN/dS ratios) are significantly above 1, the excess of non-synonymous heteroplasmies most likely reflects positive selection rather than relaxation of functional constraints [9]. 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 (213 vs. 94), there were significantly more synonymous sites shared than non-synonymous sites (13 vs. 11, Table 2).

Table 2: Non-synonymous and synonymous heteroplasmic mutations in the coding region and shared and non-shared heteroplasmies. p=0.019 (Fisher's exact test)

heteroplasmies in Liv1 and Liv2	non-synonymous	synonymous	total
shared	11	13	24
not shared	202	81	283
total	213	94	

Neither differences in MAF at synonymous vs. non-synonymous sites nor changes in MAF with age can explain this dependency, as there were no significant differences between synonymous and non-synonymous heteroplasmies with respect to either (p>0.05, Figure S9).

Discussion

Shared heteroplasmies within liver

In this study, we compared mtDNA heteroplasmy patterns in samples collected from the two liver lobes of 85 individuals sampled at autopsy. We showed that there is a significant correlation in MAF across the entire genome, which is significantly higher

for the control region than for the coding region. The results indicate that this is neither due to the presence of 7S DNA in the sequencing data set nor to selection or mutation acting on just a few sites. If sites are heteroplasmic in both liver samples of an individual, the MAFs at those sites are similar, independent of age.

Heteroplasmy sharing within liver could occur due to several potential mechanisms. First, it could derive from a pre-existing, inherited heteroplasmy [17-20] that remains at similar frequencies across the tissue because drift is limited. Nonetheless, one would expect the difference in MAF between the lobes to increase with increasing age of an individual, because different regions of a tissue diverge with increasing numbers of cellular generations. However, the data presented here show that variation in MAFs between liver lobes does not increase with age, suggesting that shared heteroplasmies are not inherited.

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 in the data set (Figure 2), and hence this explanation is unlikely.

Third, it has been shown that colonic stem cells with non-synonymous mtDNA mutations can spread within a colonic crypt by crypt fission, resulting in clonal expansion from a few cells that spread throughout a tissue and retain the same level of mutant mtDNA on the cellular level [21]. 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. This does not seem very likely, as clonal

expansion was shown for single intestinal crypts only and is supposed to be rather slow [21].

Finally, an equilibrium of heteroplasmy across an entire tissue could also involve 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 [22]. 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 remains of apoptotic cells [23]. While the uptake and integration of cell-free nuclear DNA material has been shown [24], it is unclear whether cells would also accept mitochondrial DNA. However, studies of heteroplasmy at the single cell level (reviewed in [25]) do suggest the possibility of transfer between cells. At the single-cell level, the proportion of mutant alleles differs significantly between different cells, ranging from 0 to almost 100 %. For example, many heteroplasmic sites occur at low levels in single cells from muscle fibers [26], but in buccal epithelium and heart muscle, strong clonal expansion of mutant alleles to almost homoplasmy has been described for single cells [27]. Taken together, many studies show that different heteroplasmic sites are present at the single cell level, supporting an intercellular expansion of mutant DNA molecules. Furthermore, in cell culture mixes of fluorescently labeled cell lines, mtDNA is exchanged between similar cell lines, although the mechanism is unclear [28]. 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.

Implications for mechanism of mtDNA replication

We found higher correlations in MAF for the control region heteroplasmies than for coding region heteroplasmies. Similarly, a previous study gave evidence that heteroplasmic sites in the control region, but not in the cytochrome B gene or in the tRNA proline gene, are shared between different single cells extracted from one region of human substantia nigra [29]. However, this effect was not observed for neuronal cells, indicating that the effect of cellular DNA exchange might differ between different cell types and that single cell analysis could provide more insights into the tissue-wide distribution of heteroplasmic sites.

The results presented here indicate that the sequenced DNA is derived from the mtDNA genome and not a product of 7S or other free DNA. Indeed, the 7S DNA in healthy cells has an estimated half-life of only 45 min [30], minimizing the chance that 7S DNA would be present in tissues sampled at autopsy. Hence, minor alleles detected in the tissue samples likely derive from intact mitochondrial genomes. As the liver lobes of an individual do not share heteroplasmy throughout the whole mtDNA genome, but mainly in the control region, an exchange of entire mtDNA molecules between cells is unlikely. Instead, it may be that short mtDNA fragments enter the cells and mitochondria and serve as primers for the replisome, with mutant alleles being incorporated into the mitochondrial genome of new cells. In fact, mtDNA is abundant in human plasma, albeit in fragments shorter than nuclear DNA fragments, possibly due to the lack of histone protection [31]. Furthermore, significant proportions of cell-free oligonucleotides enter the nucleus of human cells within 30-60 min of co-incubation [32] and readily integrate into the nuclear genome, thereby acting as mobile genetic elements [33].

One potential explanation for the observed heteroplasmy sharing is thus that very short DNA fragments of mitochondrial origin, which are either products of apoptosis or actively secreted by healthy cells [23], pass the cellular membrane [24] and enter the mitochondria, as has been observed for the nucleus [32]. The strong regional bias of sharing for sites in the control region could then derive from the mechanism of replication itself. The mitochondrial RNA Polymerase POLRMT creates short fragments of RNA that are used as primers for DNA replication of the leading strand [1]. The processivity of the mitochondrial replisome depends on the POLG β subunit, which has a much higher binding affinity for the control region than for the rest of the molecule [34]. Short DNA primers complementary to the control region are therefore more likely to be used for replication, as this region is more readily accepted as a template for POLG β binding, and thus end up incorporated in mtDNA genomes in different cells. As a high structural and sequence conservation of the origin of replication (oriL) is essential for the replication of the light strand [35], primers incorporating mutations into the mitochondrial genome might only be accepted for initiation of heavy strand replication in the control region.

Sharing of non-synonymous heteroplasmies in the coding region

There was more sharing of synonymous heteroplasmies than non-synonymous heteroplasmies between liver lobes. This supports the idea that sharing is not due to specific sites which are under selection, because in this case a positive correlation of heteroplasmy sharing with non-synonymous heteroplasmies would be expected. We see indications for overall positive selection on non-synonymous heteroplasmies across the coding region, as described previously [9]. For synonymous sites, mutations are expected to be neutral; the only potential difference is a change in

codon usage for the cell [36]. While this might have a marginal influence on mitochondrial translation, the question remaining is if mutations in the coding region are integrated into the mitochondrial genome through transfer of small DNA fragments between cells, and if so is this more likely for fragments carrying synonymous rather than non-synonymous mutations. A putative explanation is that non-synonymous heteroplasmies reduce the activity of mitochondria and slow down aging of the tissue, in accordance with the “survival of the slowest” theory [37]. This results in a decreased metabolism and consequently in reduced production of reactive oxygen species. However, intercellular organelle transfer, which would include the transfer of whole mitochondrial genomes, is supposed to be induced by cellular stress [22], which would favor transfer of synonymous over non-synonymous heteroplasmies. Alternatively, it may also be that cells exceeding a certain threshold of non-synonymous heteroplasmies enter apoptosis, which includes a rapid ingestion of cell contents by phagocytes [38]. Therefore, non-synonymous heteroplasmies would be removed from circulation before they can enter other cells, again favoring the exchange of synonymous over non-synonymous heteroplasmies.

Hence, several mechanisms could co-regulate the exchange of mtDNA material between cells, and the final pattern of heteroplasmy distribution may reflect the outcome of many different processes.

Conclusion

We compared mtDNA heteroplasmy in liver samples taken from different liver lobes of the same individual. We found that heteroplasmic sites in the control region are shared significantly more often between the liver lobes than sites from the coding region. In addition, the MAFs in different tissue regions are highly correlated,

regardless of the individual's age. Neither the presence of small DNA fragments, such as 7S DNA, nor increased mutation rates at single sites can explain these results. While these results may reflect positive selection acting on the same mutations arising independently in different cells, it seems more likely that there is exchange of DNA between cells. In particular, small fragments of mtDNA could be exchanged between cells, with fragments originating from the mtDNA control region more likely to be used as templates for mtDNA replication, thereby resulting in an integration of mutant alleles into the genomes of new cells. However, the mtDNA region-dependent pattern of heteroplasmy sharing throughout a tissue suggests that more than one process may be influencing this pattern.

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. DNA was extracted as previously described [9]. The collection of samples and the experimental procedures 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).

Virological assays and histological investigation

HIV RNA und HCV RNA concentration in blood was determined by using the Abbott RealTime® HIV-1 and HCV systems and the m2000sp/m2000rt instruments according to the instructions of the manufacturer. For detection of HBV DNA the

Abbott RealTime® HBV system was used. The 95% limit of detection (LOD95) 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 LOD95 of the assay to 200, 400, and 600, respectively. In one sample, inhibition could not be eliminated even after 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 LOD95 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 LOD95 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 [39]. 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 [9]. DNA was sequenced on the Illumina HiSeq

platform in rapid mode with 95 bp paired-end reads. Bases were called with FreeIbis [40] and reads were subsequently trimmed and merged using leeHom [41].

Heteroplasmy detection

Heteroplasmy was detected according to the DREEP pipeline [42, 43]. Initially, heteroplasmies were called if the minor allele frequency (MAF) for the most frequent minor allele was at least 2% on both the forward and reverse strand, and if there were at least 3 reads supporting the minor allele on each strand. The following regions were excluded for heteroplasmy analysis: 302-316, 513-526, 566-573, and 16,181-16,194. All sites with a coverage below 20 % or above 200 % of the average coverage of the sample were removed, resulting in the removal of 7 sites from 4 samples. 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 [42, 43]. Nine individuals were removed in this filter step, retaining 85 individuals for the subsequent analyses. 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; all samples passed this filter.

Correlation analysis

Statistical analysis was performed using R [44], with analyses performed for the entire mtDNA genome and separately for the coding 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 than compared the MAFs of these heteroplasmies to the MAF in the other tissues of an individual, even if the site was not detected 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; 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), Mann-Whitney U tests were used to detect significant differences. Fisher's exact test was used to test for an association of sharing of heteroplasmic sites between liver lobes with control region vs. non-control region and synonymous vs. non-synonymous heteroplasmies.

Coverage across the mitochondrial genome

We used per-site coverage determined by the *filter_and_summary.pl* script of the DREEP pipeline [43] 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 [9] was calculated by calculating the Ka/Ks ratio using *KaKs_Calculator 2.0* [45] between the revised Cambridge Reference Sequence (rCRS; Andrews, 1999 doi:10.1038/13779) and the rCRS after introducing all minor alleles of heteroplasmies in the coding region. 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 analyzed by overlapping the position of the heteroplasmy and its minor allele with the MitImpact database [46] and comparing the *MutationAssessor* [47] score. We determined the ratio of medium/ high risk mutations to the overall number of mutations caused by heteroplasmies and tested for significance by randomly selecting the same number of mutations from the MitImpact database and re-calculating the ratio of medium/ high risk mutations to the overall number of mutations. Tests for significance of an excess of non-synonymous heteroplasmies were performed as previously described [9].

Competing interests

The authors declare that they have no competing interests.

Funding

Research was funded by the Max Planck Society.

Acknowledgements

We thank Anne Haubner, Antje Weihmann, Barbara Höber and Melanie Grabmüller for technical support. We thank Enrico Macholdt and Michael Dannemann for valuable discussion.

References

1. Falkenberg M, Larsson N-G, Gustafsson CM: **DNA replication and transcription in mammalian mitochondria.** In *Annual Review of Biochemistry. Volume* 76; 2007: 679-699: *Annual Review of Biochemistry*].
2. Larsson NG. **Somatic mitochondrial DNA mutations in mammalian aging.** *Annu Rev Biochem* 2010, **79:**683-706.
3. Nicholls TJ, Minczuk M. **In D-loop: 40 years of mitochondrial 7S DNA.** *Experimental Gerontology* 2014, **56:**175-181.
4. Penta JS, Johnson FM, Wachsman JT, Copeland WC. **Mitochondrial DNA in human malignancy.** *Mutation Research-Reviews in Mutation Research* 2001, **488:**119-133.
5. Stewart JB, Chinnery PF. **The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease.** *Nat Rev Genet* 2015, **16:**530-542.
6. Wallace DC, Chalkia D. **Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease.** *Cold Spring Harbor Perspectives in Biology* 2013, **5**.
7. Calloway CD, Reynolds RL, Herrin GL, Anderson WW. **The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age.** *American Journal of Human Genetics* 2000, **66:**1384-1397.
8. He Y, Wu J, Dressman DC, Iacobuzio-Donahue C, Markowitz SD, Velculescu VE, Diaz LA, Jr., Kinzler KW, Vogelstein B, Papadopoulos N. **Heteroplasmic mitochondrial DNA mutations in normal and tumour cells.** *Nature* 2010, **464:**610-U175.

9. 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* 2015, **112:**2491-2496.

10. 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* 1999, **286:**774-779.

11. Naue J, Horer S, Sanger T, Strobl C, Hatzer-Grubwieser P, Parson W, Lutz-Bonengel S. **Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA.** *Mitochondrion* 2015, **20:**82-94.

12. Samuels DC, Li C, Li B, Song Z, Torstenson E, Clay HB, Rokas A, Thornton-Wells TA, Moore JH, Hughes TM, et al. **Recurrent Tissue-Specific mtDNA Mutations Are Common in Humans.** *Plos Genetics* 2013, **9**.

13. Wang Y, Michikawa Y, Mallidis C, Bai Y, Woodhouse L, Yarasheski KE, Miller CA, Askanas V, Engel WK, Bhasin S, Attardi G. **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* 2001, **98:**4022-4027.

14. Soares P, Ermini L, Thomson N, Mormina M, Rito T, Rohl A, Salas A, Oppenheimer S, Macaulay V, Richards MB. **Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock.** *American Journal of Human Genetics* 2009, **84:**740-759.

15. Wachsmuth M, Huebner A, Li M, Madea B, Stoneking M. **Age-Related and Heteroplasmy-Related Variation in Human mtDNA Copy Number.** *Plos Genetics* 2016, **12**.

16. Roberti M, Musicco C, Polosa PL, Milella F, Gadaleta MN, Cantatore P. **Multiple protein-binding sites in the TAS-region of human and rat mitochondrial DNA.** *Biochemical and Biophysical Research Communications* 1998, **243:**36-40.

17. Payne BAI, Wilson IJ, Yu-Wai-Man P, Coxhead J, Deehan D, Horvath R, Taylor RW, Samuels DC, Santibanez-Koref M, Chinnery PF. **Universal heteroplasmy of human mitochondrial DNA.** *Human Molecular Genetics* 2013, **22:**384-390.

18. Guo Y, Li CI, Sheng QH, Winther JF, Cai QY, Boice JD, Shyr Y. **Very Low-Level Heteroplasmy mtDNA Variations Are Inherited in Humans.** *Journal of Genetics and Genomics* 2013, **40:**607-615.

19. Li M, Rothwell R, Vermaat M, Wachsmuth M, Schroeder R, Laros JFJ, van Oven M, de Bakker PIW, Bovenberg JA, van Duijn CM, et al. **Transmission of human mtDNA heteroplasmy in the Genome of the Netherlands families: support for a variable-size bottleneck.** *Genome Research* 2016, **26:**417-426.

20. Rebolledo-Jaramillo B, Su MS-W, Stoler N, McElhoe JA, Dickins B, Blankenberg D, Korneliussen TS, Chiaromonte F, Nielsen R, Holland MM, 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* 2014, **111:**15474-15479.

21. Greaves LC, Preston SL, Tadrous PJ, Taylor RW, Barron MJ, Oukrif D, Leedham SJ, Deheragoda M, Sasieni P, Novelli MR, 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* 2006, **103:**714-719.

22. Rogers RS, Bhattacharya J. **When Cells Become Organelle Donors.** *Physiology* 2013, **28:**414-422.

23. van der Vaart M, Pretorius PJ: **Circulating DNA - Its origin and fluctuation.** In *Circulating Nucleic Acids in Plasma and Serum V. Volume* 1137. Edited by Gahan PB, Swaminathan R; 2008: 18-26: *Annals of the New York Academy of Sciences*].

24. Basak R, Nair NK, Mittra I. **Evidence for cell-free nucleic acids as continuously arising endogenous DNA mutagens.** *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* 2016, **793:**15-21.

25. Yao YG, Kajigaya S, Young NS. **Mitochondrial DNA mutations in single human blood cells.** *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* 2015, **779:**68-77.

26. Del Bo R, Bordoni A, Boneschi FM, Crimi M, Sciacco M, Bresolin N, Scarlato G, Comi GP. **Evidence and age-related distribution of mtDNA D-loop point mutations in skeletal muscle from healthy subjects and**

mitochondrial patients. *Journal of the Neurological Sciences* 2002, **202:**85-91.

27. Nekhaeva E, Bodyak ND, Kraytsberg Y, McGrath SB, Van Orsouw NJ, Pluzhnikov A, Wei JY, Vijg J, Khrapko K. **Clonally expanded mtDNA point mutations are abundant in individual cells of human tissues.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99:**5521-5526.
28. Jayaprakash AD, Benson EK, Gone S, Liang R, Shim J, Lambertini L, Toloue MM, Wigler M, Aaronson SA, Sachidanandam R. **Stable heteroplasmy at the single-cell level is facilitated by intercellular exchange of mtDNA.** *Nucleic Acids Research* 2015, **43:**2177-2187.
29. Cantuti-Castelvetria I, Lin MT, Zheng KN, Keller-McGandy CE, Betensky RA, Johns DR, Beal MF, Standaert DG, Simon DK. **Somatic mitochondrial DNA mutations in single neurons and glia.** *Neurobiology of Aging* 2005, **26:**1343-1355.
30. Gensler S, Weber K, Schmitt WE, Perez-Martos A, Enriquez JA, Montoya J, Wiesner RJ. **Mechanism of mammalian mitochondrial DNA replication: import of mitochondrial transcription factor A into isolated mitochondria stimulates 7S DNA synthesis.** *Nucleic Acids Research* 2001, **29:**3657-3663.
31. Zhang RY, Nakahira K, Guo XX, Choi AMK, Gu ZL. **Very Short Mitochondrial DNA Fragments and Heteroplasmy in Human Plasma.** *Scientific Reports* 2016, **6**.
32. Zamecnik P, Aghajanian J, Zamecnik M, Goodchild J, Witman G. **Electron micrographic studies of transport of oligodeoxynucleotides across eukaryotic cell-membranes.** *Proceedings of the National Academy of Sciences of the United States of America* 1994, **91:**3156-3160.
33. Mittra I, Khare NK, Raghuram GV, Chaubal R, Khambatti F, Gupta D, Gaikwad A, Prasannan P, Singh A, Iyer A, et al. **Circulating nucleic acids damage DNA of healthy cells by integrating into their genomes.** *Journal of Biosciences* 2015, **40:**91-111.
34. Di Re M, Sembongi H, He J, Reyes A, Yasukawa T, Martinsson P, Bailey LJ, Goffart S, Boyd-Kirkup JD, Wong TS, et al. **The accessory subunit of mitochondrial DNA polymerase gamma determines the DNA content of**

mitochondrial nucleoids in human cultured cells. *Nucleic Acids Research* 2009, **37:**5701-5713.

35. Wanrooij S, Fuste JM, Stewart JB, Wanrooij PH, Samuelsson T, Larsson NG, Gustafsson CM, Falkenberg M. **In vivo mutagenesis reveals that OriL is essential for mitochondrial DNA replication.** *Embo Reports* 2012, **13:**1130-1137.
36. Anderson S, Bankier AT, Barrell BG, Debruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, et al. **Sequence and organization of the human mitochondrial genome.** *Nature* 1981, **290:**457-465.
37. deGrey A. **A proposed refinement of the mitochondrial free radical theory of aging.** *Bioessays* 1997, **19:**161-166.
38. van der Vaart M, Pretorius PJ. **The origin of circulating free DNA.** *Clinical Chemistry* 2007, **53:**2215-2215.
39. Mulisch M WU: *Romeis - Mikroskopische Technik.* Berlin Heidelberg: Springer Spektrum; 2015.
40. Renaud G, Kircher M, Stenzel U, Kelso J. **freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers.** *Bioinformatics* 2013, **29:**1208-1209.
41. Renaud G, Stenzel U, Kelso J. **leeHom: adaptor trimming and merging for Illumina sequencing reads.** *Nucleic Acids Research* 2014, **42**.
42. Li M, Schroeder R, Ko A, Stoneking M. **Fidelity of capture-enrichment for mtDNA genome sequencing: influence of NUMTs.** *Nucleic Acids Research* 2012, **40**.
43. Li M, Stoneking M. **A new approach for detecting low-level mutations in next-generation sequence data.** *Genome Biology* 2012, **13**.
44. R Core Team: **R: A Language and Environment for Statistical Computing.** Vienna, Austria: R Foundation for Statistical Computing; 2015.
45. 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* 2010, **8:**77-80.
46. Castellana S, Ronai J, Mazza T. **MitImpact: an exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants.** *Hum Mutat* 2015, **36:**E2413-2422.

47. Reva B, Antipin Y, Sander C. **Predicting the functional impact of protein mutations: application to cancer genomics.** *Nucleic Acids Res* 2011, **39:**e118.

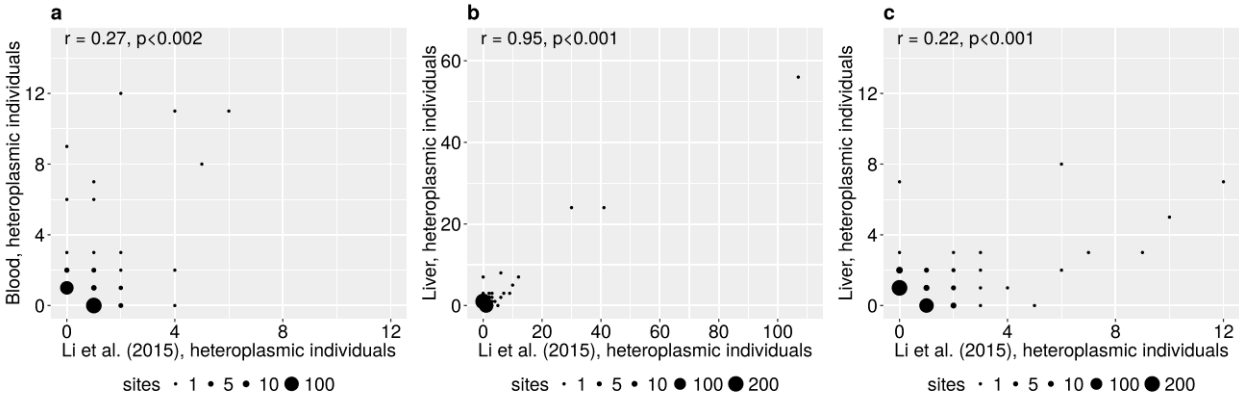

Figure S1: Comparison of heteroplasmic sites found in this study with a previous study [9]. Number of heteroplasmic individuals for each site identified in this study is compared to that from a previous study [9]. The size of a dot indicates how many sites are plotted at the same position in the scatter plot. Comparison for **a** blood samples; **b** liver samples; **c** liver samples, without sites 60, 72, 94.

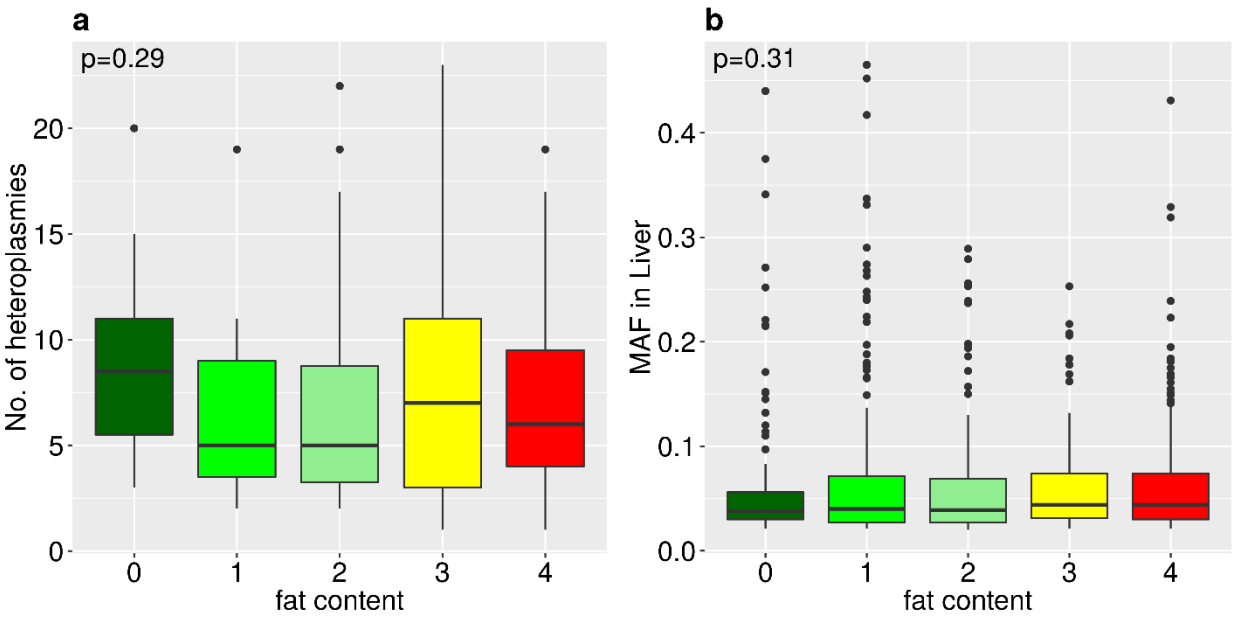

Figure S2: Liver fat content and heteroplasmy. For all fat content groups (no fat (0), low fat (1), medium fat (2), high fat (3) and adipohepatic (4)), the **a** total number of heteroplasmic sites and **b** MAF at single sites are shown. P-values for the Mann-Whitney U test comparing group 0 and group 4 are given.

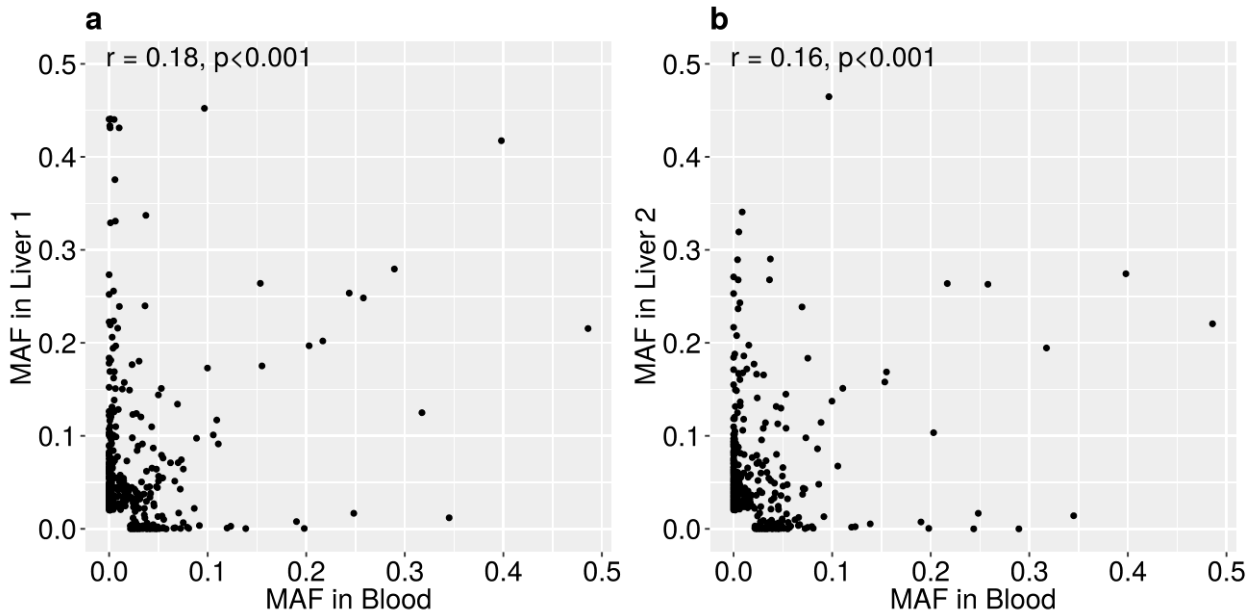

Figure S3: Correlation of MAFs at heteroplasmic sites in blood and liver. Each dot is one heteroplasmic site in one individual. Pearson's correlation coefficient r is given. Heteroplasmic sites are compared in **a** blood and liver sample 1, **b** blood and liver sample 2.

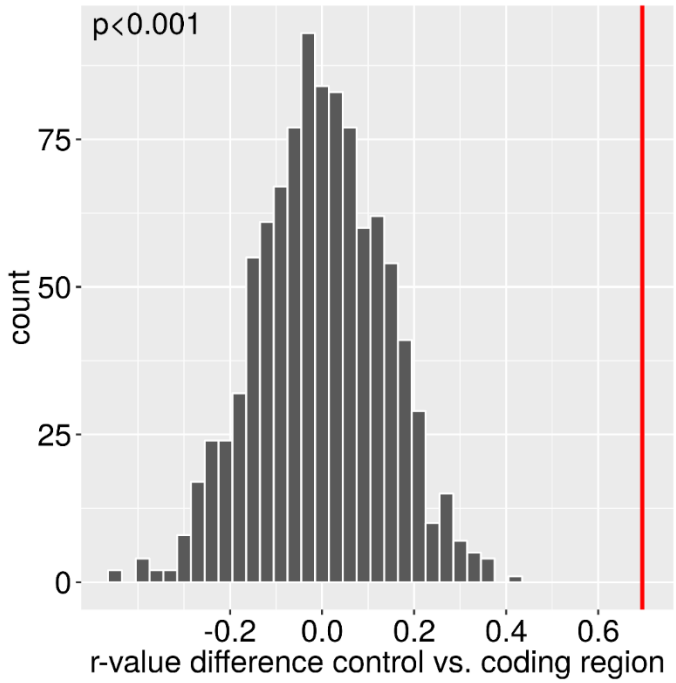

Figure S4: Distribution of r-value differences for random partitions of all sites into either control (262 sites) or non-control region (385 sites). The red bar indicates the observed difference between the r-values for the control region and the non-control region.

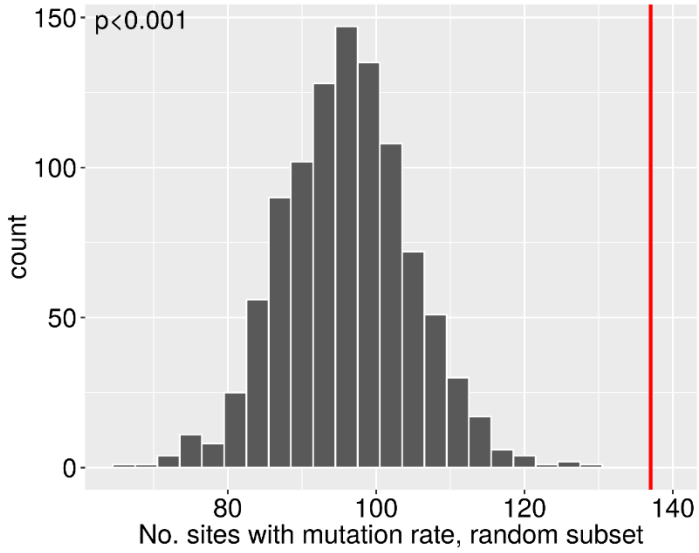

Figure S5: Expected vs. observed number of heteroplasmic sites with reported mutation rates. A random subset of sites from the mtDNA with the same size as the total number of heteroplasmic sites was taken and sites with a reported mutation rate [14] were counted. The distribution of the number of sites with a reported mutation rate is shown. Vertical red line indicates the real number of heteroplasmic sites with a reported mutation rate [14].

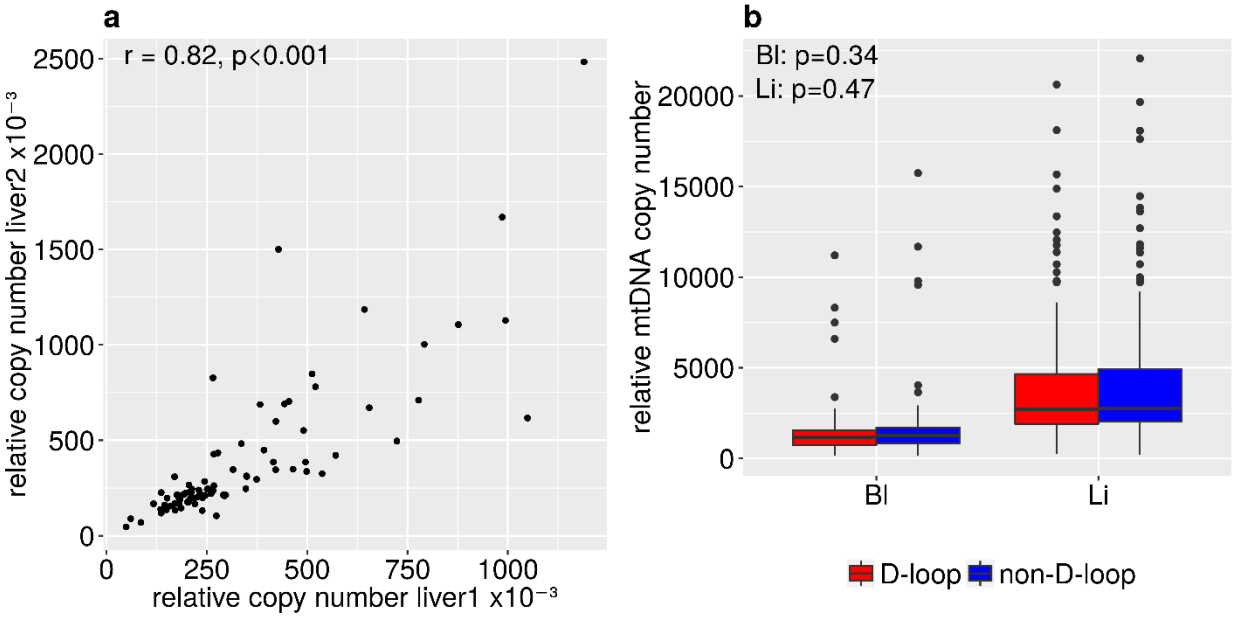

Figure S6: Relative mtDNA copy number. a relative copy number in liver sample 1 versus sample 2. Pearson's correlation coefficient r is given. **b** distribution of relative mtDNA copy numbers per mtDNA region (D-loop (red) vs. non-D-loop (blue)) and tissue (Bl = blood, Li = liver). P-value for the Mann-Whitney U test comparing D-loop and non-D-loop is given.

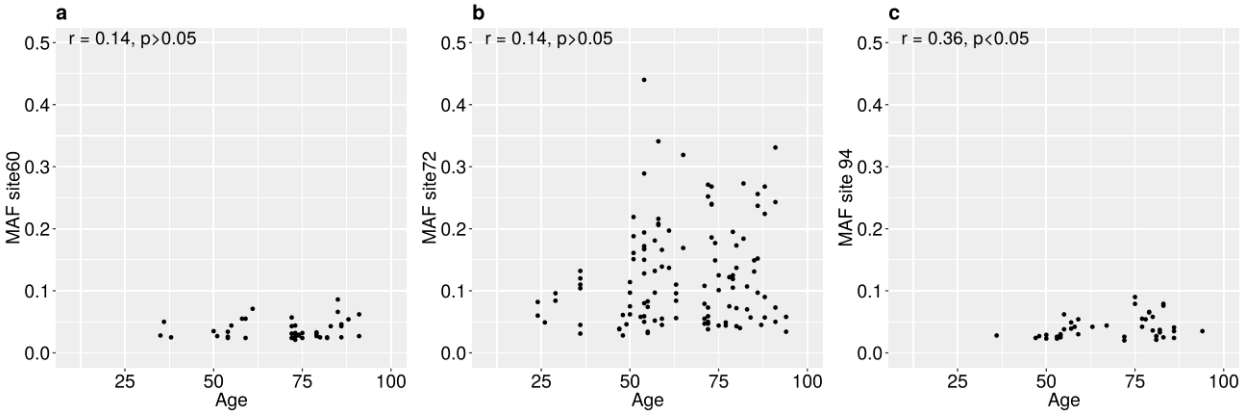

Figure S7: MAF and age correlation for sites 60, 72 and 94. Each dot is one heteroplasmic site in one individual. Pearson's correlation coefficient r is given. **a** site 60. **b** site 72. **c** site 94.

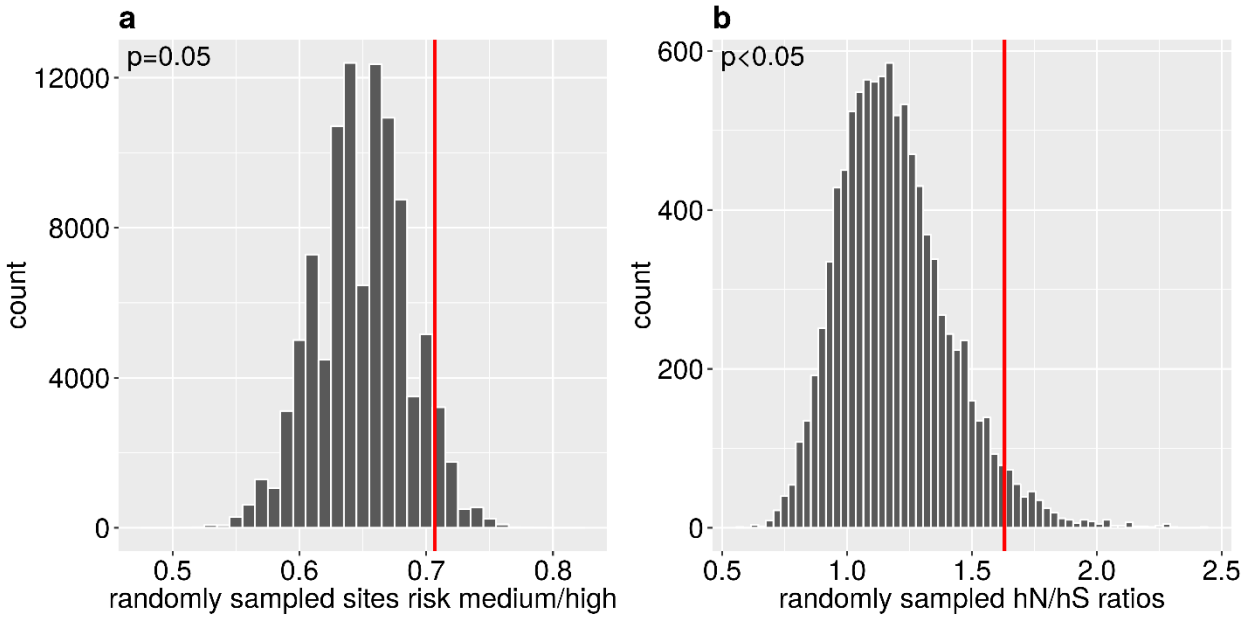

Figure S8: Significance tests for synonymous and non-synonymous sites. a The red vertical bar indicates the observed proportion of high/medium risk liver specific heteroplasmies with *MutationAccessor* scores (n=200). Distribution of the proportion of medium/ high risk mutations based on the *MutationAccessor* score for 100,000 random subsamples of 200 sites with MutationAccessor scores. **b** The red vertical bar indicates the observed hN/hS ratio for liver specific heteroplasmic sites. Distribution of 10,000 hN/hS-ratios calculated by comparing the rCRS to itself after introducing the mutations observed for liver specific heteroplasmies at random sites across the coding region.

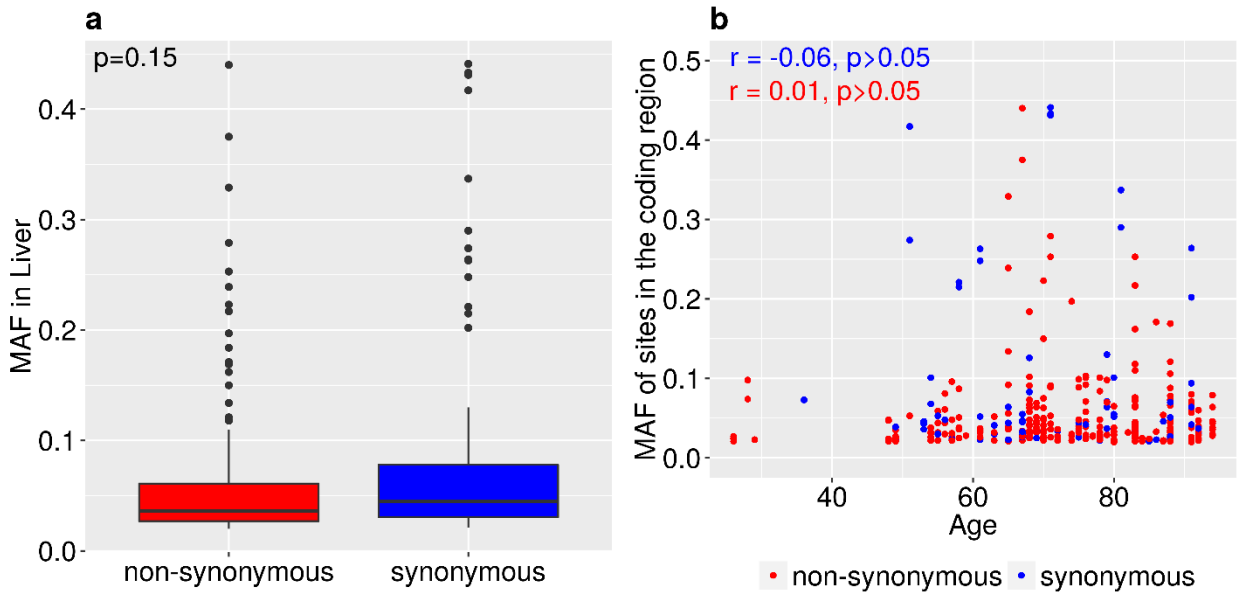

Figure S9: Heteroplasmies in the coding region, MAFs and age. a Distribution of MAFs for non-synonymous (red) and synonymous (blue) heteroplasmies. The p-value for the Mann-Whitney U test comparing non-synonymous and synonymous heteroplasmies is given. **b** correlation of MAFs with age, plotted for non-synonymous (red) and synonymous heteroplasmies (blue). Each dot is one heteroplasmic site in one individual. Pearson's correlation coefficient r is given for both subsets.